ARQUITETURAS DE REDES DE ARMAZENAMENTO DE DADOS

Ariovaldo Veiga de Almeida

Trabalho Final de Mestrado Profissional em Computação

Instituto de Computação
Universidade Estadual de Campinas

ARQUITETURAS DE REDES DE
ARMAZENAMENTO DE DADOS

Ariovaldo Veiga de Almeida

Julho de 2006

Banca Examinadora:

- **Prof. Dr. Nelson Luis Saldanha da Fonseca**
  Orientador - Instituto de Computação - Unicamp
- **Prof. Dr. Omar Branquinho**
  PUC-Campinas
- **Prof. Dra. Islene Calciolari Garcia**
  Instituto de Computação - Unicamp
- **Prof. Dra. Maria Beatriz Felgar de Toledo (Suplente)**
  Instituto de Computação – Unicamp

ARQUITETURAS DE REDES DE ARMAZENAMENTO DE DADOS

Este exemplar corresponde à redação final do Trabalho Final devidamente corrigido e defendido por Ariovaldo Veiga de Almeida e aprovado pela Banca Examinadora.

Campinas, Julho de 2006.

Prof. Dr. Nelson Luis Saldanha da Fonseca
(Orientador – Instituto de Computação - Unicamp)

Trabalho Final apresentado ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Computação na área de Redes de Computadores.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Maria Júlia Milani Rodrigues – CRB8a / 2116

Almeida, Ariovaldo Veiga de
AL64a Arquiteturas de redes de armazenamento de dados / Ariovaldo Veiga de Almeida -- Campinas, [S.P. :s.n.], 2006.

Orientador : Nelson Luis Saldanha da Fonseca
Trabalho final (mestrado profissional) - Universidade Estadual de Campinas, Instituto de Computação.

1. Redes de computação – Protocolos. 2. Redes de informações. 3. Armazenamento de dados. 4. Desempenho. I. Fonseca, Nelson Luis Saldanha da. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Título em inglês: Storage networks architectures.

Palavras-chave em inglês (Keywords): 1. Computer network protocols. 2. Information networks. 3. Information warehousing. 4. Performance.

Área de concentração: Redes de Computadores

Titulação: Mestre em Computação

Banca examinadora: Prof. Dr. Nelson Luis Saldanha da Fonseca (IC-UNICAMP)
Prof. Dr. Omar Branquinho (PUC-Campinas)
Profa. Dra. Islene Calciolari Garcia (IC-UNICAMP)

Data da defesa: 26/07/2006

Programa de Pós-Graduação: Mestrado em Computação

. Agradecimentos

Agradeço em primeiro lugar a Deus pela minha existência e por ter me dado a chance de evoluir academicamente e como pessoa ao enfrentar a tarefa de escrever este trabalho.

Agradeço aos meus pais, Gioconda Veiga de Almeida e Gonzaga de Almeida, pela vida e educação que me deram, amo muito vocês.

Agradeço aos meus irmãos Carlos Gonzaga de Almeida e Gerson Veiga de Almeida pelo carinho e amizade que disfrutamos todos esses anos.

Agradeço muito ao Prof. Nelson, pela orientação, apoio e enorme paciência no decorrer de todo este trabalho.

Agradeço a todo pessoal do Instituto de Computação, em particular a Claudinha, uma pessoa muito especial, atenciosa e colaborativa em todos os momentos.

Por fim agradeço ao meu grande amigo Neto, sem sua obstinação eu não teria retornado ao mundo acadêmico.

Dedicatória

Dedico este trabalho a:

Minhas três queridas Mães: Mãe Mariquinha (minha avó), Mãe Nica (minha tia) e minha Mãe materna. Sempre chamei a todas de “Mãe” , sempre estarão no meu coração.

Meu Pai que sempre lutou para dar uma vida melhor para os filhos.

Meus Filhos: Andre Luis Esteves Almeida e Alexandre Esteves Almeida. Espero que meu exemplo possa de alguma forma ajudá-los no futuro.

Resumo

As Redes de Armazenamento de Dados oferecem aos sistemas computacionais acesso consolidado e compartilhado aos dispositivos de armazenamento de dados, aumentando sua eficiência e disponibilidade. Elas permitem que os dispositivos de armazenamento de dados de diferentes fornecedores, mesmo que usem diferentes protocolos de acesso, possam ser logicamente disponibilizados para acesso. Elas permitem que as funções de gerenciamento de dados, como backup e recuperação, replicação de dados, ambientes de recuperação de desastres, e migração de dados, possam ser realizados de maneira rápida e eficiente, com o mínimo de sobrecarga nos sistemas computacionais.

Na década de 80, observou-se a descentralização dos sistemas computacionais que evoluíram dos ambientes centralizados, como no caso dos sistemas *mainframe*, para plataformas distribuídas, onde os sistemas eram separados em blocos operacionais, com cada um dos blocos realizando uma função específica.

Não foram somente os sistemas computacionais que evoluíram, mas também os sistemas de armazenamento de dados evoluiram para arquiteturas distribuídas. A evolução natural dos dispositivos de armazenamento de dados dos sistemas computacionais foi do uso de conexão direta e dedicada aos computadores para uma forma mais flexível e compartilhada. A forma adotada foi através do uso de infra-estruturas das redes de computadores.

Este trabalho analisa as tecnologias das redes de armazenamento de dados *Storage Area Networks* (SAN) e *Network Attached Storage* (NAS), que são as principais arquiteturas que utilizam as tecnologias de redes para o armazenamento e compartilhamento de dados. Enfoca-se as vantagens decorrentes dessas arquiteturas quando comparadas com a forma tradicional de conexão direta do dispositivo de armazenamento de dados aos computadores, a denominada arquitetura *Direct Attached Storage* (DAS).

Palavras-chave: Redes de armazenamento de dados, Armazenamento de dados, Protocolos de redes, DAS, NAS, SAN, *Direct Attached Storage*, *Network Attached Storage*, *Storage Area Network*.

Abstract

Storage Networks offer shared access to data storage devices, increasing the efficiency and the availability of storage data. They allow data storage devices, from different suppliers, using different access protocols, to be logically available for access. They also allow management of data, backup and recovery, data replication, disaster recovery environments, and data migration can be done in a fast and efficient way, with minimum overhead to the computer systems.

In the 80's, we observed the decentralization of the computational systems. They evolved from a centralized environment to distributed platforms, where systems were separated in operational blocks, with each block executing specific functions.

Both the computational systems and the storage envolved to a distributed architecture. The natural evolution of the storage devices was to move from the direct connection to computational systems to a more flexible and shared approach. This happened by the adoption of infrastructures used by computer networks.

This work analyzes Storage Networks architectures: *Storage Area Network* (SAN) and *Network Attached Storage* (NAS), which are the main architectures that employ computer networks technologies. We will show the advantages of these architectures compared to the traditional form of direct connection of storage devices to computers, the named *Direct Attached Storage* (DAS) architecture.

Keywords: Storage Networks, Storage, Network Protocols, DAS, NAS, SAN, *Direct Attached Storage*, *Network Attached Storage*, *Storage Area Network*.

"As Redes de Armazenamento de Dados podem melhorar de maneira significativa os procedimentos de backup e recuperação, provendo funcionalidade avançada ao mesmo tempo em que diminuim o Custo Total de Propriedade (TCO) e fornecem um Retorno de Investimento (ROI) significativo quando comparadas com ambiente DAS."

Salomon Smith Barney, 2001

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviações

ACL	-	Access Control List
AFS	-	Andrew File System
ANSI	-	American National Standards Institute
ARP	-	Address Resolution Protocol
ARPA	-	Advanced Research Projects Agency
ARPANET	-	ARPA Network
ASCII	-	American Standard Code for Information Interchange
ATA	-	Advanced Technology Attachment
ATM	-	Asynchronous Transfer Mode
CIFS	-	Common Internet File System
CIM	-	Common Information Model
COBOL	-	Common Bussiness Oriented Language
CPU	-	Central Processing Unit
DARPA	-	Defense ARPA
DAS	-	Direct Attached Storage
DIMM	-	Dual Inline Memory Module
EBCDIC	-	Extended Binary Coded Decimal Interchange Code
ESCON		Enterprise Systems Connection
FAS	-	Fabric Attached Storage
FAT	-	File Allocation Table
FC	-	Fibre Channel
FC-AL	-	Fibre Channel Arbitrated Loop
FCIA	-	Fibre Channel Industry Association
FCIP	-	Fibre Channel over TCP/IP
FCP	-	Fibre Channel Protocol
FDDI	-	Fiber Distributed Data Interface
FORTRAN	-	Formula Translator
FS	-	File System
FTP	-	File Transfer Protocol
GGF	-	Global Grid Forum
GSI	-	Grid Security Infrastructure
HBA	-	Host Bus Adapter
HIPPI	-	High Performance Parallel Interface
HPC	-	High Performance Computing
HTTP	-	Hypertext Transfer Protocol
ICMP	-	Internet Control Message Protocol
IDE	-	Integrated Drive Electronics
IEEE	-	Institute of Electrical and Electronics Engineers
IEFT	-	Internet Engineering Task Forc*e*
IFCP	-	Internet Fibre Channel Protocol
IP	-	Internet Protocol
IPX/SPX	-	Internetwork Packet Exchange/Sequenced Packet Exchange
iSCSI	-	Internet Small Computer System Interface

ISO	-	International Standarization Organization
LAN	-	Local Area Network
LVM	-	Logical Volume Manager
MAN	-	Metropolitan Area Network
MTBF	-	Mean Time Between Failure
NAS	-	Network Attached Storage
NetBIOS	-	Network Basic Input/Output System
NFS	-	Network File System
NIC	-	Network Interface Card
NNTP	-	Network News Transfer Protocol
NTFS	-	New Technology File System
OGSA	-	Open Grid Services Architecture
OSI	-	Open Systems Interconnection
PATA	-	Parallel Advanced Technology Attachment
RAID	-	Redundant Array of Independent Disks
RFC	-	Request For Comments
ROI	-	Return Of Investment
RPC	-	Remote Procedure Calls
SAN	-	Storage Area Networks
SASI	-	Shugart Associates Systems Interface
SATA	-	Serial Advanced Technology Attachment
SETI	-	Search for ExtraTerrestrial Intelligence
SCSI	-	Small Computer Systems Interface
SNA	-	System Network Architecture
SNIA	-	Storage Networking Industry Association
SMB	-	Server Message Block
SMTP	-	Simple Mail Transfer Protocol
SSA	-	Serial Storage Architecture
TCP	-	Transmission Control Protocol
TCP/IP	-	Transfer Control Protocol/Internet Protocol
TCO	-	Total Cost of Ownership
TI	-	Technolgy Information
UFS	-	Unix File System
UDP	-	User Datagram Protocol
USENET	-	idem NNTP
VLAN	-	Virtual LAN
VLSI	-	Very large Scale Integration
VO	-	Virtual Organization
VOIP	-	Voice Over IP
XFS	-	XFS File System
XML	-	Extensible Markup Language
WTC	-	World Trade Center
WAN	-	Wide Area Network
WWW	-	World Wide Web

1. Introdução

"Com a sem-precedente conectividade provida pela Internet, muitas novas aplicações são desenvolvidas. Uma imensa quantidade de dados tornou-se disponível para acesso, satisfazendo a demanda dessas novas aplicações"

David H. C. Du, 2003

As informações estão certamente entre os bens mais importantes e críticos das organizações. Podemos comprovar isso quando do ataque terrorista às Torres Gêmeas do WTC[1], em 2001. Segundo Rory Nolan, diretor técnico da ITRM, empresa irlandesa de gerenciamento de risco, com a destruição das torres 43 % da empresas que sofreram grandes perdas de dados não reabriram. Do restante de empresas, 51% delas fecharam nos dois anos seguintes ao desastre e somente 6% das empresas sobreviveram no longo prazo [1].

Assim, devido à importância que os dados representam, é crítica a sua armazenagem. Atualmente as organizações empresariais precisam armazenar, organizar, gerenciar e disponibilizar dados de uma forma global, além de garantir sua integridade durante todos os processos, em alguns casos, devendo mantê-los por vários anos por razões legais.

Os dados nos sistemas computacionais crescem continuamente de forma a ocupar quase todo armazenamento disponível. Um estudo [2] publicado em 2003 pela School of Information Management and Systems da Universidade da California, Berkeley, mostrou que, em 2002, foram armazenadas em mídia magnética aproximadamente 5 EB[2] de informações. Desse volume, aproximadamente 2 EB em disco rígido. Esse mesmo estudo estimou que o volume de informações armazenadas por pessoa, por ano, no mundo, foi de 800 MB[3]. Esse crescimento vertiginoso tem levado a uma busca contínua pelo aumento na capacidade e desempenho dos dispositivos de armazenamento de dados.

[1] O World Trade Center possuia duas torres que foram derrubadas por ataques terroristas em 11/09/2001.
[2] Exabyte equivale a 10^{18} bytes.
[3] Megabyte equivale a 10^{6} bytes.

Desde o aparecimento do primeiro disco magnético, desenvolvido pela IBM em 1956, as tecnologias associadas aos discos tem evoluido continuamente. Elas procuram atender ao crescente volume de conteúdo digital, impulsionado pela também crescente utilização dos computadores e, mais recentemente, da Internet.

Em Dezembro de 2005, o IDC[4] publicou um estudo [3] sobre sistemas externos de armazenamento de dados em disco magnético para sistemas abertos. Nesse estudo, somente no terceiro quartil de 2005, o tamanho do mercado mundial para esse tipo de armazenamento foi de $5.7 bilhões de dólares, equivalente a um volume de 505 PB[5] de dados, correspondendo, respectivamente, a um crescimento de 13.3% e 58% em relação ao ano anterior.

Segundo outra pesquisa do IDC [4], publicada em 2005, a projeção de crescimento do mercado brasileiro de armazenamento de dados deverá ser, em média, de 35% ao ano até o ano de 2009. Em 2004 o volume vendido em equipamentos de armazenamento corporativo em disco foi de aproximandamente 7 PB sendo que a previsão para 2009 deverá ser superior a 30 PB. Isso aumentará, com certeza, a necessidade das empresas brasileiras em administrar grandes volumes de dados.

Em 2001, as empresas Merrill Lynch e McKinsey & Company, publicaram um estudo [5] onde relatam que o gerenciamento tradicional de grandes volumes de dados de forma isolada tem sido bastante ineficiente, tanto que os resultados do estudo mostram que até 50% dessa capacidade de armazenamento pode ser desperdiçada ou subutilizada quando gerenciada dessa forma. Essa baixa utilização motivou ao aparecimento de novas formas de armazenamento buscando otimizar o gerenciamento de dados. Elas surgiram como evolução natural dos dispositivos de armazenamento de dados dos sistemas computacionais se deslocando da conexão direta e dedicada aos computadores para formas mais flexíveis, consolidadas e compartilhadas. A tecnologia adotada foi através da utilização de infra-estruturas baseadas em redes de computadores.

[4] IDC é o principal provedor global de inteligência no mercado e serviços de aconselhamento.
[5] Petabyte equivale a 10^{15} bytes

A conexão do dispositivo de armazenamento de dados através de uma forma direta e dedicada computador é conhecida como DAS (*Direct Attached Storage*). A conexão DAS é a forma pioneira. Embora ofereça uma plataforma sólida, conhecida e dominada pelos usuários, apresenta limitações relacionadas a gerenciabilidade, escalabilidade, disponibilidade, confiabilidade e recuperabilidade.

As limitações da arquitetura DAS levaram a vários desenvolvimentos, tanto na área de armazenamento de dados, quanto na área de tecnologias de redes, que, segundo Preston [6], estão convergindo para infra-estruturas integradas, as chamadas Redes de Armazenamento de Dados (*Storage Networks)*. Existem, atualmente, dois tipos de redes de armazenamento de dados: as redes *Storage Area Networks* (SAN) e as redes *Network Attached Storage* (NAS).

As redes de armazenamento buscam solucionar as limitações da arquitetura de armazenamento de dados dedicado (DAS), oferecendo um alto nível de desempenho com maior escalabilidade e flexibilidade, possibilitando que os departamentos de informática das organizações atinjam altos níveis de serviço na utilização e gerenciamento de dados.

A seguir é apresentada uma breve descrição dessas arquiteturas, que serão mais detalhadas e comparadas nos Capítulos 3 e 4.

- **DAS** - arquitetura que consiste em conectar o dispositivo de armazenamento de dados de forma dedicada e direta ao computador. Exemplos típicos desse tipo de arquitetura podem ser: discos internos dos computadores, dispositivos JBOD[6], etc.
- **SAN** - arquitetura que contém dispositivos de armazenamento que se comunicam através do protocolo serial SCSI[7] na forma dos protocolos: FCP[8] - transporte de comandos e blocos de dados SCSI através de tecnologia Fibre Channel, ou iSCSI[9] - transporte de comandos e blocos de dados SCSI através de protocolo TCP/IP. A arquitetura SAN é caracterizada pela "transferência de blocos de dados" entre os sistemas computacionais e os dispositivos de armazenamento de dados.

[6] Just a Bunch Of Disks, termo usado para equipamento modular com discos, sem funcionalidades adicionais.
[7] Small Computer System Interface – tecnologia e protocolo que será visto no próximo capítulo.
[8] Fibre Channel Protocol – protocolo que será visto no próximo capítulo.
[9] Internet SCSI – protocolo que será visto no próximo capítulo.

- **NAS** - arquitetura que contém dispositivos de armazenamento que se comunicam através de redes baseadas em TCP/IP e usam protocolos de compartilhamento de arquivos, sendo os mais comuns: NFS[10] e CIFS[11], que são protocolos nativos nos sistemas operacionais baseados em UNIX e Microsoft Windows, respectivamente. Na arquitetura NAS, os dispositivos de armazenamento de dados são vistos, e se comportam, como servidores de arquivos, com seus próprios sistemas operacionais processando protocolos de comunicação. Ela é caracterizada pela "transferência de arquivos" entre os sistemas computacionais e os dispositivos de armazenamento de dados.

O presente estudo descreve e compara as tecnologias de redes de armazenamento de dados SAN e NAS, que representam as principais arquiteturas que utilizam as tecnologias de rede para o armazenamento e compartilhamento de dados. Faz-se uma exposição dos dispositivos básicos utilizados na implementação das redes de armazenamento de dados, como os computadores, os discos magnéticos, e os protocolos de transporte e comunicação de dados. Apresenta-se, também, as vantagens dessas tecnologias quando comparadas com a forma tradicional de conexão DAS.

1.1. Objetivos

1.1.1. Objetivo geral

O objetivo principal deste estudo é discutir o estágio atual do armazenamento de dados nos sistemas computacionais, descrevendo e comparando as arquiteturas DAS, SAN e NAS. Apresenta-se neste trabalho a forte ligação entre redes de armazenamento de dados e computação grid, mostrando que as redes de armazenamento de dados oferecem a infra-estrutura básica para o armazenamento de dados em grid.

[10] Network File System – protocolo que será visto no próximo capítulo.
[11] Common Internet File System – protocolo que será visto no próximo capítulo.

1.1.2. Objetivos específicos

Para alcançar o objetivo geral deste estudo é necessário alcançar, também, alguns objetivos específicos sobre as redes de armazenamento de dados, são eles:

- Definir o conceito de armazenamento de dados;
- Definir o conceito de acesso remoto a dados;
- Desenvolver uma comparação entre os sistemas de armazenamento de dados;
- Avaliar a convergência das atuais formas de armazenamento e compartilhamento de dados, com as infra-estruturas de redes de armazenamento de dados;
- Avaliar a conectividade entre os sistemas e o desempenho no gerenciamento e acesso aos dados;
- Analisar a disponibilidade de dados e entre os diversos sistemas da organização;
- Avaliar a questão dos *backups* e do tempo na implantação de novos sistemas;
- Analisar por quais motivos e benefícios do uso das redes de armazenamento de dados, e como isso se traduz em uma diminuição nos custos de armazenamento e gerenciamento de acesso a dados;
- Apresentar as tecnologias e soluções utilizadas na implementação das redes de armazenamento de dados que permitem obter o máximo de seus recursos;
- Apresentar quais arquiteturas de armazenamento oferecem flexibilidade para implementação de novas soluções de armazenamento de dados com facilidade e com uma boa relação custo/benefício.
- Apresentar os conceitos de computação grid e como armazenamento de dados em redes se integra a essa nova tecnologia.

1.2. Justificativa do estudo

A justificativa para o desenvolvimento deste trabalho deve-se ao armazenamento de dados em rede ser a solução atual que permite o gerenciamento eficiente dos grandes volumes de dados que crescem diariamente. A consolidação que essa solução oferece, permitindo conectar quaisquer computadores aos dispositivos de armazenamento de dados, oferece uma economia e melhor uso dos investimentos em armazenamento de dados.

O armazenamento de dados em rede pode diminuir custos de gerenciamento e processamento de informações, pois permite a consolidação de vários ambientes e o compartilhamento de dispositivos e capacidades de armazenamento. Outra vantagem é a criação de um ambiente unificado para os vários ambientes computacionais, permitindo o processamento de informações a qualquer momento, a partir de qualquer sistema computacional que tenha acesso à rede de armazenamento de dados.

1.3. Motivação para o trabalho

Os avanços nas tecnologias de armazenamento de dados, como: grande capacidade de armazenamento, velocidade de acesso aos dados, confiabilidade e redução de custos, têm possibilitado administrar grandes volumes de informações, transformado-as em grandes bases de dados organizadas.

As redes de armazenamento de dados permitem aos sistemas computacionais acesso compartilhado aos dispositivos de armazenamento de dados, aumentando sua eficiência e disponibilidade. Elas permitem a separação entre sistema computacional e dispositivo de armazenamento de dados. Mesmo que o sistema computacional não esteja disponível, independente da causa (por alguma falha, manuetenção, etc), os dados estarão disponíveis e protegidos. Permitem ainda que as funções de gerenciamento de dados, como *backup* e recuperação, replicação de dados, ambientes de recuperação de desastres e migração de dados, possam ser realizadas de maneira rápida e eficiente, com o mínimo de sobrecarga nos sistemas computacionais.

1.4. Limitações do trabalho

A consolidação das redes de armazenamento de dados leva a uma diminuição de custos, melhor gerenciamento, melhor utilização de recursos, maior disponibilidade de dados e grande escalabilidade. No entanto, já que existem várias situações onde qualquer uma das arquiteturas poderia ser utilizada, fica um questionamento: Qual das diferentes arquiteturas deve ser adotada para cada ambiente computacional ?

Assim, uma limitação deste trabalho é a falta de uma comparação que considere os requisitos de desempenho das aplicações a serem usadas nas redes de armazenamento de dados. Não é simples essa comparação, já que as duas arquiteturas de redes de armazenamento de dados tratam de diferentes unidades básicas de informação, de um lado "transferência de blocos de dados" no caso de SAN e, por outro lado, "transferência de arquivos" no caso de NAS. Além disso, em NAS**,** o controle da distribuição de dados fisicamente nos discos é feito em de nível de sistema de arquivos do próprio dispositivo de armazenamento de dados, enquanto, em SAN, esta tarefa fica a cargo do próprio sistema operacional do computador que está utilizando o dispositivo de armazenamento de dados.

1.5. Estrutura do trabalho

O restante deste trabalho está organizado da seguinte forma:

O Capítulo 2 é o referencial teórico sobre as tecnologias que são usadas para se "montar" ou "construir" sistemas de armazenamento de dados em redes. Neste capítulo, são abordados, desde os sistemas computacionais, dispositivos de armazenamento de dados, até as redes de computadores e seus protocolos de comunicação e transporte de dados.

O Capítulo 3 utiliza o formalismo do modelo SNIA[12] de armazenamento de dados compartilhado para descrever as arquiteturas de armazenamento de dados DAS, SAN e NAS.

[12] Storage Nerwork Industry Association – Associação de empresas ligadas a redes de armazenamento de dados.

O Capítulo 4 compara as formas de armazenamento DAS, SAN e NAS, ressaltando, principalmente, a utilização de armazenamento de dados em rede. É formalizada a unificação de SAN e NAS no que tem sido denominada arquitetura *Fabric Attached Storage* (FAS).

O Capítulo 5 complementa o estudo de redes de armazenamento de dados ao mostrar sua integração com a computação grid.

Finalmente, no Cápitulo 6, apresentam-sa as conclusões e recomendações para armazenamento e compartilhamento de dados. Faz-se uma apreciação de como sistemas de armazenamento de dados devem complementar o ambiente computacional no futuro.

2. Componentes de uma Rede de Armazenamento

"Eu acho que há um mercado mundial para talvez cinco computadores"
Thomas Watson, Chairman da IBM, 1943.

"Os computadores no futuro não devem pesar mais que 1.5 toneladas"
Revista americana Popular Mechanics, 1949.

O objetivo deste capítulo é oferecer uma revisão dos principais ambientes relacionados ao armazenamento de dados em rede. Aborda-se computadores, dispositivos de armazenamento de dados, redes de computadores e finalizando com principais protocolos relacionados com armazenamento de dados e redes.

2.1. Computadores

Os computadores fazem parte dos instrumentos que o homem inventou e desenvolveu buscando agilizar suas atividades. As primeiras aplicações dos computadores foram como instrumento para agilizar a execução de operações matemáticas, contudo hoje sua aplicação se encontra em praticamente todos as atividades do ser humano.

Os computadores são os grandes responsáveis pela existência dos dispositivos de armazenamento de dados. Desde a invenção dos computadores viu-se necessidade de armazenar dados, tanto os dados a serem processados quanto aos dados produzidos a partir de processamento.

Para muitos, a história do computador remonta ao surgimento do mais antigo equipamento para cálculo, o ábaco. Surgido da tentativa do homem de se livrar dos trabalhos manuais e repetitivos e da necessidade inata de se fazer contas mais rápida e precisamente, o ábaco provavelmente foi criado por volta de 2500 A.C. Em latim, uma pedrinha do ábaco era chamada de *calculus*, daí a raiz das palavras ligadas a calcular.

O primeiro computador mecânico [8], projetado e desenvolvido como protótipo por Charles Babbage, em 1822, foi "*Difference Engine",* máquina para tabular polinômios. Em 1834, Babbage projeta a "*Analytical Engine*", máquina que usava cartões perfurados para armazenar os programas. Para essa máquina, Ada Byron King, a condessa de Lovelace, matemática talentosa, criou programas, se tornando a primeira programadora de computador do mundo.

Em 1936, Alan Turing publica "*On Computable Numbers, with an application to the Entscheidungsproblem"*, onde introduz uma máquina de computar digital abstrata, agora chamada de "Turing machine", uma concepção dos princípios do computador moderno.

Existem várias formas de se organizar os principais momentos históricos ligados à invenção e desenvolvimento dos computadores modernos. Um forma de se fazer isso é a divisão dos computadores em gerações, iniciando a partir do uso de válvulas Assim usa-se a Tabela 2.1 para descrever as gerações de computadores. Na Tabela 2.2 apresenta-se os principais marcos da evolução dos computadores. Deve-se observar que existem pequenas divergências sobre as datas exatas de alguns eventos que levaram a definir as várias gerações, mas a ideía básica de cada geração é geralmente um consenso.

Tabela 2.1 Gerações de computadores

Legenda – Geração dos computadores	
I	**Válvulas a vácuo (1943 a 1958)** Caracteriza-se pela construção de computadores a partir de válvulas a vácuo, resultando em máquinas grandes que podiam pesar muitas toneladas. Estas máquinas consumiam uma grande quantidade de energia devido ao grande número de válvulas que usavam para funcionar.
II	**Transistores (1959 a 1964)** A invenção do transistor, em 1947, foi feita pelos pesquisadores da Bell Labs, John Bardeen, Walter Brattain e Willian Shockley. O primeiro computador experimental transistorizado foi feito, em 1953, pela Universidade de Manchester, e mostrou a possibilidade de substituir as válvulas dos computadores, garantindo menor consumo, maior poder computacional e confiabilidade a eles. Essa geração usa também compiladores FORTRAN e/ou COBOL facilitando o desenvolvimento de aplicações.

III	**Circuitos integrados (1964 a 1972)** O circuito integrado foi inventado, em 1958, por Jack St Clair Kilby da Texas Instruments e, num trabalho separado, por Robert Noyce da Fairchild Semiconductors Corporation. Com os circuitos integrados surgiram os grandes computadores (*mainframes*) e nessa época aparecem os primeiros Sistemas Operacionais.
IV	**Microprocessadores (1972 a 1993)** Em 1971, a Intel produziu o primeiro microprocessador comercial (4004) que operava com 2.300 transistores e executava 60.000 cálculos por segundo.
V	**Microprocessadores ULSI (1993 até os dias atuais)** Circuito central com tecnologia ULSI. Utiliza-se o Processamento Paralelo (múltiplos processadores executando múltiplas tarefas) e Inteligência Artificial (capacidade de processar o conhecimento). Amplia-se a capacidade de processamento de dados, armazenamento e taxas de transferência.

Tabela 2.2 Principais marcos na evolução dos computadores

Geração	Data	Autor / Instituição	Comentários
I	1943	Dr.Thomas Flower The Post Office Research Labs	Colossus – foi o primeiro computador eletrônico digital programável. Ele continha 2400 válvulas e foi desenvolvido em segredo para decodificação de mensagens alemãs durante 2ª Guerra Mundial.
I	1944	Howard Aiken Universidade de Harvard	Primeiro computador eletrônico, Harvard Mark I. Ocupava 120 m^3, pesava 5 toneladas, possuía 17 m de comprimento por 2,5 m de altura e era composto de milhares de relés e precisava de 3 segundos para operar dois números de 10 dígitos. Foi criado para a Marinha dos EUA criar tabelas balísticas.

Geração	Data	Autor / Instituição	Comentários
I	1946	John W. Mauchly e J. Presper Eckert Universidade da Pensilvânia	Um dos primeiros computadores totalmente eletrônico, o ENIAC pesava 30 toneladas, possuía 5 m de largura por 24 m de comprimento e era composto de 18.000 válvulas, 70.000 transistores, 10.000 capacitores e 800 Km de fios. Consumia 160 KW de potência. Executava até 5000 operações de soma/subtração por segundo. Foi usado pelo Exército dos EUA para cálculo de trajetórias balísticas e testar teorias para o desenvolvimento da bomba de hidrogênio.
I	1951	John W. Mauchly e J. Presper Eckert Remington Rand	O primeiro computador comercial a ter sucesso, o UNIVAC I foi também o primeiro computador de propósito geral. A primeira unidade foi para o U. S. Census Bureau dos Estados Unidos.
I	1951	Jay Forrester MIT	O primeiro computador a trabalhar em tempo-real, o Whrilwind foi o primeiro a permitir computação interativa através do uso de um teclado e um *display* tubo de ráios catódicos. Foi feito para ser usado pelo US Air Defense System.
II	1954	Bell Laboratories Força Aérea Norte-americana	Desenvolvido o primeiro computador usando transistores, o TRADIC. Ocupava menos de 1 m^3, possuía 800 transistores e 10.000 diodos. Dissipava 100 W de potência e executava 1.000.000 de operações por segundo.
II	1960	Digital	Surgimento do primeiro mini computador, o PDP 1, do qual foram vendidas unidades.
II	1961	IBM	Aparecimento do IBM-1401, pequeno computador comercial com grande sucesso.

Geração	Data	Autor / Instituição	Comentários
II	1964	Seymour Cray na empresa Control Data Corporation	Lançamento do CDC 6600, que utilizava múltiplas unidades funcionais. Acredita-se que tenha sido o primeiro computador a ser designado como "supercomputador". Ele possuia a velocidade de *clock* mais rápida de sua época (100 nanossegundos). Foi um dos primeiros computadores a usar o líquido Freon para refrigeração. Foi o primeiro computador comercial a usar um *display* tubo de ráios catódicos.
IV	1974	Jonathan Titus	Desenvolvimento do primeiro computador pessoal chamado de Mark-8.
IV	1975	Steve Wozniac e Steve Jobs	Lançamento do computador pessoal Apple I. O início da popularização de microcomputadores.
IV	1981	IBM	Lançamento dos microcomputadores IBM PC, operando a com processador Intel 8088 com velocidade de 4.7 MHz.
IV	1990	IBM	Lançamento dos microcomputadores IBM PC-AT 386, operando com processador Intel 8086 com velocidade de 20 MHz, usando os microchips VLSI.
V	1993	Intel	Lançamento do primeiro processador Pentium, que posteriormente evoluiu para o Pentium II, III e 4, utilizando memórias DIMM e barramento de 64 bits.

A partir da geração atual de computadores, é possível que não possamos mais empregar o conceito de gerações tecnológicas por causa das rápidas mudanças que vem ocorrendo nas tecnologias de *hardware*, *software* e comunicações, o que dificulta a definição de parâmetros claros de classificação.

2.2. Dispositivos de armazenamento de dados

Desde o surgimento dos sistemas computacionais existe a necessidade de armazenamento de dados. Eles são passíveis de serem armazenados para uso imediato ou futuro, tanto os dados de entrada, que serão processados pelos sistemas computacionais, como os dados intermediários, usados durante o processamento, quanto os dados de saída, resultado final do processamento.

Os dispositivos de armazenamento de dados dos sistemas computacionais podem ser classificados de acordo com o tipo de dados que eles armazenam, podendo ser: dados *on-line*, dados *off-line*, dados transientes ou dados persistentes.

Entre os dispositivos de armazenamento de dados, os dispositivos magnéticos são amplamente usados pelos sistemas computacionais, pois oferecem as seguintes vantagens:

- **Baixo custo**: devido às tecnologias empregadas e ao grande volume de produção. Para se ter uma idéia, hoje o armazenamento em disco magnético é mais barato que o armazenamento em papel.
- **Alta confiabilidade**: Os discos magnéticos estão entre os dispositivos eletromagnéticos mais confiáveis produzidos atualmente. A maioria dos discos magnéticos de mercado possui MTBF[13] maior que um milhão de horas.
- **Universalidade:** Nas últimas décadas, as tecnologias de conexão dos discos magnéticos aos sistemas computacionais têm sido padronizadas. Hoje, a maioria dos discos magnéticos pode ser usada com a maioria dos sistemas computacionais.

As fitas e/ou cartuchos magnéticos são muito usados como dispositivos de armazenamento de dados em situações de *backup* e manutenção de dados *off-line*. São usados para arquivamento, transporte de dados entre sistemas e/ou localidades, armazenamento de dados históricos, replicação de dados em local remoto, etc.

[13] Mean Time Between Failure – tempo médio de falha, ou seja, é um indicador de possibilidade de falha.

Já os discos magnéticos são usados nos processamentos *on-line* e em tempo real . A gravação e leitura de dados podem ser feitas continuamente e de forma imediata. A grande vantagem dos discos em relação às fitas e cartuchos magnéticos é a possibilidade que os discos oferecem para acesso aleatório aos dados armazenados, agilizando muito a manipulação de informações. O disco magnético tem se apresentado como a solução tecnológica mais usada para o armazenamento de dados persistentes e *on-line*.

Existem outras tecnologias de armazenamento de dados, como os discos ópticos, os discos de estado sólido, etc, porém neste trabalho o enfoque principal será dado às tecnologias de discos magnéticos, pois são amplamente utilizados [2] e oferecem, a baixo custo, o tipo de armazenamento permanente e *on-line* necessário aos sistemas computacionais atuais.

2.2.1. Disco magnético

O disco magnético ou disco rígido é o dispositivo de armazenamento de dados persistente mais comum nos sistemas computacionais. Desde os primeiros computadores ficou clara a necessidade de armazenar dados. No início com uso de cartões perfurados e depois fita magnética. Foi somente em 1956 que a IBM lançou o primeiro sistema com discos rígidos magnéticos, o RAMAC 305 [9]. Ele possuia um conjunto de 50 discos ("pratos") e tinha a capacidade de armazenar o equivalente a 5 MB de dados. Ele permitia que informações fossem codificadas (gravadas) nos discos de forma magnética, e, podeiam ser recuperadas posteriormente de forma aleatória. Essa foi considerada uma verdadeira revolução na indústria de computadores. Foi o primeiro passo para aplicações com grandes volumes de informações *on-line* e em tempo real, comuns nos dias de hoje.

2.2.1.1. Princípios de funcionamento

Os discos magnéticos sofreram várias melhorias tecnológicas desde a sua invenção, mas preservam os mesmos princípios de funcionamento, ou seja, com o uso de certos materiais ferromagnéticos é possível magnetizar de forma permanente pequenas regiões através da sua

exposição a um campo magnético [10]. Essas regiões podem ser posteriormente lidas ou regravadas. Deste modo, podemos usar esse meio magnético para armazenamento de dados.

Na Figura 2.1, apresenta-se o modelo geral de um disco magnético, onde são identificados os seus principais componentes. Os discos magnético são formados por um ou, em geral, mais pratos ("discos") circulares. Cada superfície do prato é recoberta de material ferro-magnético. Cada superfície possui uma "cabeça magnética" responsável pela criação do campo magnético para leitura e gravação de dados. Com a finalidade de identificar e localizar dados nos discos magnéticos, esses são organizados em círculos concêntricos chamados trilhas. As cabeças de gravação e leitura são fixadas em atuadores que movem as cabeças de leitura e gravação de trilha em trilha. Cada trilha é dividida em setores, que são blocos de dados de tamanho fixo (geralmente de 512 a 520 bytes). Todas as trilhas num mesmo raio são coletivamente chamadas de cilindro.

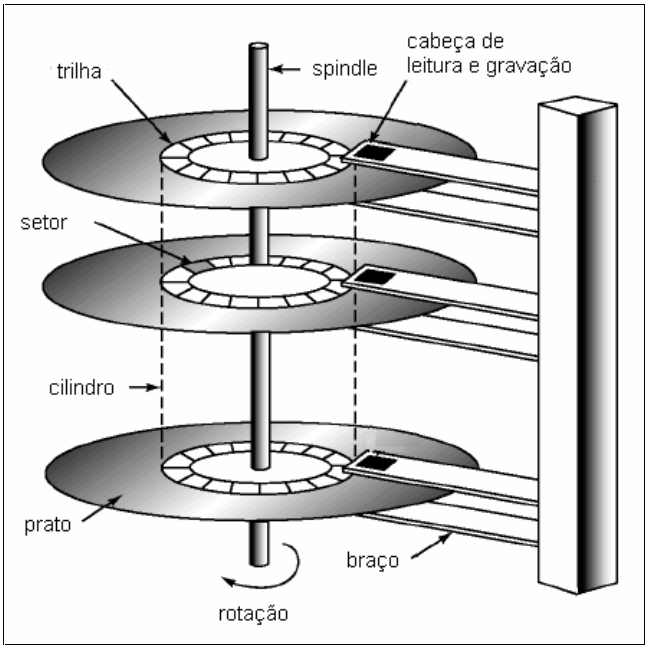


Figura 2.1 Disco magnético com seus principais componentes

Cada setor é inicializado com um padrão de sincronização e identificação seguido dos dados do setor mais um código de correção de erro (ECC[14]), seguido de um padrão indicador de final de setor. Setores adjacentes são separados por um padrão que ajuda a manter a cabeça de

[14] Error Correction Code – sistema de detecção e correção de erros.

leitura e gravação centralizada na trilha. Uma marca no início e final de cada trilha ajuda a lógica de controle do disco magnético a determinar a posição no disco e manter posição rotacional.

Dentre as várias melhorias que os discos magnéticos tiveram e continuam tendo, destacam-se: o aumento da densidade de gravação, aumento no desempenho, diminuição do tamanho, diminuição do consumo e aumento na confiabilidade. Todas essas características e a ampliação do seu uso em computadores pessoais e móveis levou a uma grande diminuição do seu custo e, conseqüentemente, sua popularização.

2.2.1.2. Tecnologias de discos magnéticos

Os discos magnéticos possuem interfaces controladoras que os conectam aos sistemas computacionais. Através dessas interfaces é possível executar e administrar a transferência de dados entre os sistemas computacionais e os discos. Várias tecnologias de discos e interfaces controladoras foram desenvolvidas. A seguir são descritas as principais tecnologias em uso atualmente.

2.2.1.2.1. SSA

A tecnologia SSA (*Serial Storage Architecture*) [11] foi inventada pela IBM em 1990. Apesar de ser vista como uma tecnologia proprietária da IBM, foi definida como padrão ANSI[15] número X3T10.1. Apesar de padrão, nunca foi usada amplamente pela indústria. Ela especifica uma forma de conexão serial, com cabeamento em laço bidirecional, de alto desempenho, que permite a conexão de até 127 discos *hot swappable*[16]. Os discos possuem duas portas de conexão. Cada controladora SSA suporta até 32 conjuntos de discos com proteção RAID[17], podendo estar conectados por fio metálico ou fibra óptica. Para facilitar a portabilidade, o SSA mantém muitas características do protocolo lógico do SCSI-2, que será apresentado a seguir. As implementações atuais operam a uma taxa de transferência de dados de até 80 MB/s.

[15] American National Standards Institute
[16] Hot swappable – característica que permite a adição ou substituiçao de discos com o equipamento em funcionamento.
[17] RAID – mecanismo de virtualização e proteção de discos que será visto ainda neste capítulo.

2.2.1.2.2. ATA

A tecnologia ATA (*Advanced Technology Attachment*) [12] tem um padrão de interface paralela para conexão de periféricos (como discos, CD-ROM etc.) a computadores pessoais. Foi originalmente desenvolvido em 1986 pelas empresas: Imprimis, uma divisão CDC[18], Western Digital, e Compaq Computer. Outro nome usado quase como sinônimo para essa interface é IDE (*Integrated Drive Electronics*). Na realidade IDE é uma tecnologia de unidade de disco e não de conexão como é o ATA. O cabeamento ATA padrão possui 40 vias e pode ter até 45 cm de comprimento, suportando somente dois dispositivos por interface controladora (um chamado mestre e outro escravo).

Com o passar do tempo, várias melhorias foram implementadas no padrão ATA. A Tabela 2.3 apresenta um histórico de melhorias implementadas no padrão ATA desde a sua criação.

Tabela 2.3 Comparação das especificações do padrão ATA

Padrão	Padrão ANSI	Nome pelo qual é conhecido	Taxa de transferência (em MB/s)
ATA-1	1994	ATA, IDE	De 3.3, 5.2, 8.3 MB/s
ATA-2	1996	EIDE,Fast-ATA,Fast-IDE,UltraATA	De 11,1 a 16,6 MB/s
ATA-3	1997	EIDE	16,6 MB/s
ATA-4	1998	ATAPI-4, ATA/ATAPI-4	De 16,7 a 33,3 MB/s (chamado de Ultra-DMA 33)
ATA-5	2000	ATA/ATAPI-5	De 44.4 a 66.7 MB/s (chamado de Ultra DMA 66)
ATA-6	2002	ATA/ATAPI-6	100 MB/s (chamado de Ultra DMA 100)
ATA-7	2005	ATA/ATAPI-7	133 MB/s (chamado de Ultra DMA 133)
ATA-8		ATA/ATAPI-8	Em desenvolvimento

Em 2003, com a introdução do Serial ATA ou SATA (Serial Advanced Technology Attachment) [13], o padrão ATA passou a ser chamado para PATA (Parallell Advanced

[18] Control Data Corporation

Technology Attachment). O padrão Serial ATA é uma evolução proativa da interface ATA, saindo de uma arquitetura de barramento paralelo para uma arquitetura de barramento serial. Na sua introdução a taxa de transferência foi de 150MB/s, porém hoje já possui taxas de 300 MB/s e com planejamento de chegar a 600 MB/s no futuro. O cabeamento SATA possui 7 vias, nele podemos conectar somente um dispositivo (somente conexão ponto a ponto). O cabeamento tem tamanho máximo de 100 cm.

2.2.1.2.3. SCSI

O padrão SCSI (*Small Computer Systems Interface*) [14, 15] foi criado em 1979 quando a empresa Shugart Associates desenvolveu a interface paralela de conexão a discos chamada SASI (*Shugart Associates Systems Interface*). Ela suportava um conjunto pequeno de comandos e funcionava a uma taxa de 1,5 MB/s. Em 1981, a Shugart Associates e outra empresa, a NCR Corporation, unem-se para convencer o comitê de padronização da ANSI a tornar o SASI um padrão. Foi, somente, em 1986, que o primeiro padrão SCSI foi publicado pelo grupo de trabalho X3T9.2 da ANSI.

Uma das principais diferenças da interface SCSI com as outras interfaces na época era que o controle do processo de comunicação estava no próprio periférico. Outras vantagens incluem cabeamento mais comprido, possibilidade de conectar até 7 periféricos (posteriormente foi aumentado para 15) em uma única interface SCSI. Os periféricos podem ser de vários tipos (discos, fitas, *CDs*, *scanners*, etc). Outra vantagem da interface SCSI sobre as demais é a possibilidade de manter e administrar uma fila de comandos e permitir o enfileiramento de requisições de vários periféricos. Isto significa, por exemplo, que a controladora de disco trabalha em multitarefa.

O primeiro padrão SCSI, também conhecido por SCSI-1, definiu um barramento paralelo de 8 bits trabalhando a uma taxa de transferência de 5 MB/s, com possibilidade de conectar até 7 periféricos.

O primeiro aperfeiçoamento veio com SCSI-2, publicado como padrão ANSI em 1994. Ele foi desenvolvido para ser um aperfeiçoamento do SCSI-1, assim, mantêm compatibilidade com este. Entre as melhorias implementadas temos o aumento do número de periféricos no barramento SCSI, que passa de 8 para 16, além disso a taxa transferência de até 10 MB/s para barramento de dados de 8 bits ou de até 20MB/s para o novo barramento de dados de 16 bits.

Um novo aperfeiçoamento veio com SCSI-3, publicado como padrão ANSI em 1996. Uma das principais novidades desse aperfeiçoamento foi a adição de um esquema de interconexão serial além da paralela. O padrão foi dividido em múltiplos níveis, oferecendo mais alternativas para o nível físico, podendo ser: SCSI Serial, Fibre Channel, SSA, IEEE 1394, InfiniBand, entre outras como ilustra a Figura 2.2.

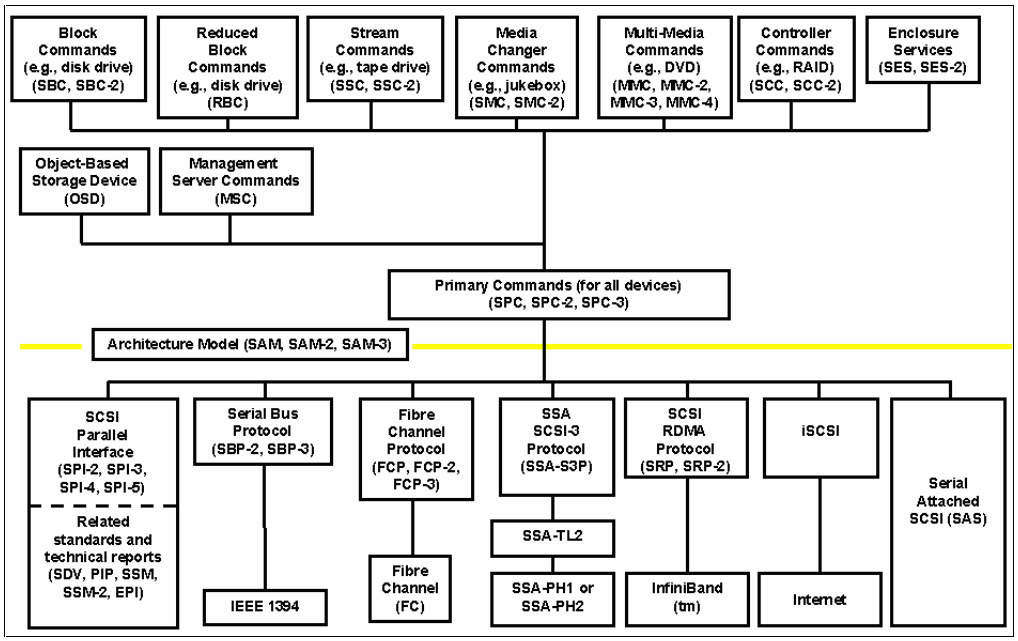


Figura 2.2 Arquitetura SCSI-3 com as várias alternaticas de conexão física.

Na Tabela 2.4, as várias especificações de SCSI são comparadas. Nela pode-se comparar os seguintes aspectos: tamanho do barramento de dados, velocidade de comunicação máxima, número máximo de periféricos, e distância máxima de cabeamento que podemos ter para cada especificação SCSI.

Tabela 2.4 Comparação das especificações do padrão SCSI

Especificações	Tamanho do barramento (bits)	Taxa de transferência (MB/s)	Tamanho máximo do cabeamento (m)	Número máximo de dispositivos
SCSI	8	5	6	8
Fast SCSI (SCSI-2)	8	10	1.5 a 3	8
Fast Wide SCSI (SCSI-2)	16	20	1.5 a 3	16
Ultra SCSI	8	20	1.5 a 3	8
Ultra Wide SCSI (SCSI-3)	16	40	1.5 a 3	16
Ultra2 SCSI	8	40	12	8
Ultra2 Wide SCSI	16	80	12	16
Ultra3 SCSI (Ultra-160)	16	160	12	16
Ultra-320 SCSI	16	320	12	16

Na Tabela 2.4 pode-se constatar que a evolução de SCSI tem buscado atender ao crescimento do armazenamento de dados. Isso pode ser visto pelo aumento da taxa de transferência e do número de dispositivos que podem ser conectados a uma interface controladora SCSI.

2.2.1.2.4. Fibre Channel

O padrão de discos Fibre Channel [16] faz uso da tecnologia Fibre Channel implementada na controladora de discos. Os discos Fibre Channel estão na mesma categoria que os discos SCSI. São considerado, pelos fabricantes, discos para ambientes "Enterprise", pois possuem características de alto MTBF, baixo rúido e alto desempenho. São próprios para os sistemas computacionais chamados servidores e são amplamente usados nos sistemas de armazenamento de dados corporativos. Usam controladoras de discos, chamadas HBA (Host Bus Adapter) que funcionam a velocidades de 100 MB/s, 200 MB/s ou 400 MB/s. A conexão de discos com a controladora é feita através de cabo serial com 4 vias e podem conectar num laço com até 126 dispositivos.

2.2.1.3. Sistema de agregação e proteção de discos

A necessidade de armazenar cada vez mais informações cresceu mais rapidamente do que a capacidade dos discos magnéticos podiam oferecer isoladamente. Uma solução foi desenvolvida para atender a essa necessidade. Ela é baseada num processo de agregação e virtualização de vários discos, criando-se um disco virtual de maior capacidade.

O conceito de agregação de discos magnéticos surgiu para aumentar a capacidade e melhorar o desempenho e a disponibilidade dos dispositivos de armazenamento. As funções básicas na agregação de discos são [10]:

- **Concatenação** – discos concatenados se apresentam como se fosse um grande e único disco virtual de maior capacidade.
- **Distribuição** – também conhecido com "*stripping*" permite aumentar o desempenho ao distribuir informações em vários discos físicos simultaneamente. Além disso, apresenta um grande e único disco virtual de maior capacidade.
- **Espelhamento** – informações idênticas são escritas em dois ou mais discos. Do ponto de vista do sistema computacional, o espelhamento é visto como um único disco.
- **Combinação** – vários discos são agregados usando-se técnicas RAID para distribuír dados entre eles. Grava-se uma informação de redundancia nos discos para garantir integridade de dados. Além disso, apresenta um grande e único disco virtual de maior capacidade.

2.2.1.3.1. RAID

O RAID (*Redundant Array of Independent Disks*) é uma tecnologia na qual os dados são armazenados de forma distribuída entre grupos de disco para conseguir ao mesmo tempo redundância e taxas mais altas de transferência de dados. Ao invés de armazenar os dados em um único disco rígido que pode falhar, o RAID mantém uma forma de redundância de informação baseada nos dados gravados entre diversos discos do grupo de disco.

Em 1987, Patterson et al., publicaram o artigo "*A Case for Redundant Arrays of Inexpensive Disks (RAID)*" [17]. O objetivo inicial era trabalhar com discos mais baratos que os discos de *mainframe* da época, daí o uso da palavra *Inexpensive* (barato), mas logo a palavra foi substituída por *Independente,* pois, pelas características de RAID, é possível agregar discos para aumentar a capacidade de armazenamento, bem como, aumentar o nível de proteção ao gravar informações redundantes em discos. Com essa proteção, mesmo falhando um disco, é possível recuperar as informações a partir dos discos restantes.

Os autores descreveram 5 configurações (RAID-1 a RAID-5), combinando múltiplos discos. Eles podem ser vistos como um único disco com aumento de desempenho e confiabilidade. Eles também descrevem uma configuração chamada RAID-0 que não impelmenta redundância, como pode ser visto a seguir:

- **RAID 0**: Distribuição de dados em vários discos (s*tripping*). Neste caso, as informações são espalhadas em vários discos para se ter um desempenho maior fazendo a gravação em paralelo entre eles. As taxas de transferências são muito altas, mas não há proteção contra falhas nos discos.
- **RAID 1**: Espelhamento *(mirror*). Todos os dados são sempre gravados em dois ou mais discos, o que oferece a mais alta confiabilidade de dados. Para leitura, a taxa de transferência de dados é mais alta do que para um único disco, pois pode ler de qualquer um dos discos simultaneamente.
- **RAID 2**: Distribuição de dados em vários discos, com informação de redundância (paridade) sendo gravada em multíplos discos. A paridade que se usa é o código de detecção de erros *Hamming code*. Na prática, não é um método usado porque as próprias controladoras de discos atuais já possuem mecanismos de detecção e correção de erros.
- **RAID 3**: Distibuição de dados em vários discos com um disco adicional de redundância (paridade). Todos os discos trabalham de forma sincronizada. A gravação é feita de forma simultânea, em "tiras" ("*stripes*") por todos os discos de dados. No disco de paridade, a informação gravada é operação lógica XOR de todos os dados da "tira". A unidade de

informação usada dos discos é um único *byte*. No caso da falha de qualquer disco, é possível continuar entregando dados a partir dos discos restantes.

- **RAID 4**: Similar ao RAID 3, porém não trabalha de maneira sincronizada e a informação básica de informação da "tira", usada para calcular a redundância (paridade) é um blocos de dados (em média de 1 a 8 Kbytes).
- **RAID 5**: Similar ao RAID 4, porém ao invés de usar a paridade em um único disco, todos os discos contém tiras para dados e tiras para armazenar a paridade dos outros discos do grupo RAID.

Mais tarde, outras configurações RAID foram definidas, inclusive pela combinação de mais de um nível. A seguir apresentamos duas deles:

- **RAID 6**: Similar ao RAID 5, mas grava um segundo disco de paridade. Assim, é possível mesmo depois da falha simultânea de dois discos de dados, continuar entregando dados;
- **RAID 0+1**: Também chamado pela indústria de RAID 10, usa de maneira conjunta as duas técnicas: divisão de dados e espelhamento. Obtém-se o melhor desempenho por conta do paralelismo do RAID-0 e a proteção oferecida pelo RAID-1.

2.3. Redes de computadores

As redes de computadores são agregações de nós distribuídos (como computadores pessoais, estações de trabalho, servidores, periféricos etc.), que através de protocolos de comunicação suportam interações entre si. Esses nós são não-estruturados e não-previsíveis. Assim, um número maior de decisões de roteamento de dados devem ser feitas para que haja sucesso da comunicação entre um nó e outro da rede. As redes têm relativamente latência maior que as conexões em canal, já que as decisões de roteamento exigem mais processamento, fazendo com que sejam relativamente mais lentas.

As redes de computadores foram desenvolvidas para conectar computadores, permitindo que uns tivessem acesso aos outros. Dessa maneira, poderiam compartilhar seus recursos disponibilizados na rede [18].

São várias as vantagens que as redes de computadores oferecem, entre elas pode-se citar:

- Permitir o acesso simultâneo a programas e dados importantes;
- Permitir às pessoas compartilhar dispositivos periféricos;
- Facilitar o processo de realização de cópias de segurança (*backups*) em máquinas remotas;
- Agilizar as comunicações pessoais como, por exemplo, o correio eletrônico.

A classificação de redes, em categorias, pode ser realizada segundo diversos critérios. Os critérios mais comuns são: dimensão ou área geográfica ocupada, capacidade de transferência de informação, topologia, meios físicos de suporte ao envio de dados, ambiente em que estão, método de transferência dos dados, tecnologia de transmissão, etc. A seguir apresenta-se as principais características que são importantes para as redes de armazenamento de dados, começando, porém, com o modelamento de redes de computadores.

2.3.1. O Modelo de Referência OSI/ISO

No início na década de 1970, diversos esforços foram realizados para se estabelecer um padrão único para redes de computadores. Vários modelos de referência foram formalmente propostos, porém somente um tem sido considerado de maneira geral, o chamado Modelo de Referência OSI/ISO[19] [19].

Em março de 1977, a Organização Internacional para Padronização (ISO), constituiu um grupo de trabalho para estudar a padronização da interconexão de sistemas de computação. Em

[19] Open Systems Interconnection / International Organization for Standardization

1984, foi definida uma arquitetura geral, denominada Modelo de Referência OSI, para servir de base para a padronização da interconexão de sistemas abertos.

Esse modelo define os processos de comunicação em rede através de camadas. O modelo especifica sete camadas e a interface, escopo funcional, requisitos e serviços de cada camada para que haja troca de mensagens entre as camadas adjacentes. Ele utiliza sucessivos encapsulamentos de protocolos, de modo que um protocolo de uma camada superior seja envolvido pelo protocolo de um camada inferior.

A Tabela 2.5 mostra o nome e descrição das sete camadas do Modelo de Referência OSI.

Tabela 2.5 Descrição das camadas do Modelo de Referência OSI

Número camada	Nome da Camada	Descrição
7	Aplicação	Seleção de serviços apropriados a aplicações
6	Apresentação	Formatação e reformatação de dados
5	Sessão	Interface para o estabelecimento de sessões
4	Transporte	Gerenciamento de conexões
3	Rede	Protocolos de roteamento de dados, interconexão de redes
2	Enlace	Método de acesso
1	Física	Transporte físico. Especifica conector, pinagem, etc.

Quando uma mensagem passa da camada $n+1$ para a camada n são acrescidos outros dados relevantes à camada n (como, por exemplo, tipo da mensagem, endereços, tamanho da mensagem, código de detecção de erro etc.). Estes dados são retirados quando a mensagem chega na camada de mesmo nível na estação de destino.

Embora as camadas estejam interligadas, elas são independentes, pois o modelo permite uma flexibilidade na implementação funcional de cada camada usando a tecnologia que seja mais apropriada (por exemplo, a camada de enlace pode ser implementada com as tecnologias Ethernet, Token Ring, FDDI, etc). Deste modo, as funções de uma camada superior podem ser suportadas por uma grande variedade de implementações das camadas inferiores.

2.3.2. Topologias de rede

A forma com que os nós (dispositivos) são conectados influenciará a rede em diversos pontos considerados críticos como flexibilidade, velocidade e segurança. Da mesma forma que não existe "o melhor" computador, não existe "a melhor" topologia de rede. Tudo depende da necessidade e aplicação.

2.3.2.1. Ponto a ponto

Na topologia ponto a ponto, um nó está ligado diretamente e, de forma única, a outro nó da rede. Na Figura 2.3, pode-se ver uma representação desse tipo de topologia. Neste tipo de topologia, toda banda da rede está totalmente disponível para comunicação entre os nós conectados.

Figura 2.3 Representação gráfica de uma rede ponto a ponto.

2.3.2.2. Linear ou em barramento

Na topologia linear ou em barramento, todos os nós compartilham um mesmo meio de conexão. Neste caso, a banda da rede será compartilhada entre todos os nós da rede. Na Figura 2.4, pode-se ver uma representação dessa topologia.

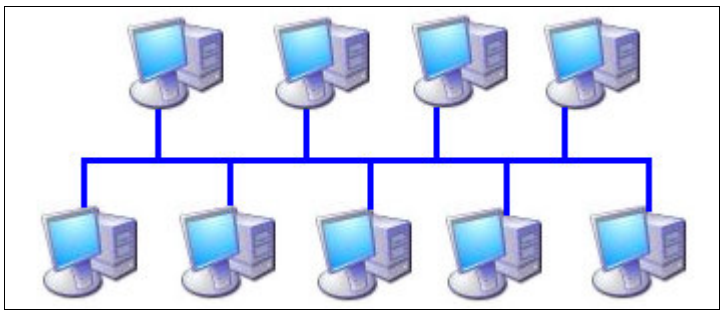


Figura 2.4 Representação gráfica de uma rede linear ou de barramento

2.3.2.3. Estrela

Na topologia estrela, todos os nós são conectadas a um equipamento concentrador, podendo ser um *hub* ou um comutador. Esta topologia é a mais usada atualmente, porque, ao contrário da topologia linear, onde a rede inteira deixa de funcionar quando algum trecho da rede se rompe, na topologia estrela, apenas a estação conectada naquele trecho deixa de utilizar a rede. A Figura 2.5 ilustra este tipo de topologia.

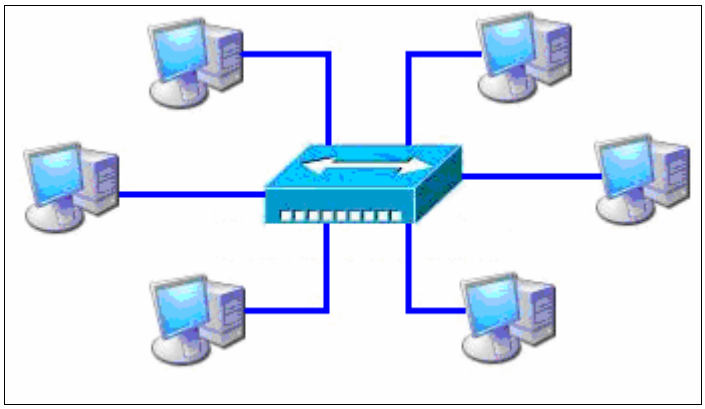


Figura 2.5 Representação gráfica de uma rede em estrela

O *hub* é um periférico que repete para todas as suas portas as informações (pacotes) que chegam em uma porta. Da mesma forma como acontece na topologia linear, a banda da rede é compartilhada entre todos os nós da rede. Já o comutador é um equipamento que tem a capacidade de analisar o endereçamento de um pacote de dados, enviando-o diretamente à porta de destino, sem replicá-lo desnecessariamente para todas as portas. Isso permite que a banda da rede possa ser usada na sua totalidade entre duas portas diferentes. Além disso, duas ou mais transmissões podem ser efetuadas simultaneamente, desde que tenham portas de origem e destino diferentes.

2.3.2.4. Anel

Na topologia em anel, os nós formam um laço fechado. Neste laço, a informação de um nó para outro circula pelos nós intermediários do anel. A informação sai do nó origem circula até chegar no nó destino. Na Figura 2.6, pode-se ver a representação desse tipo de topologia.

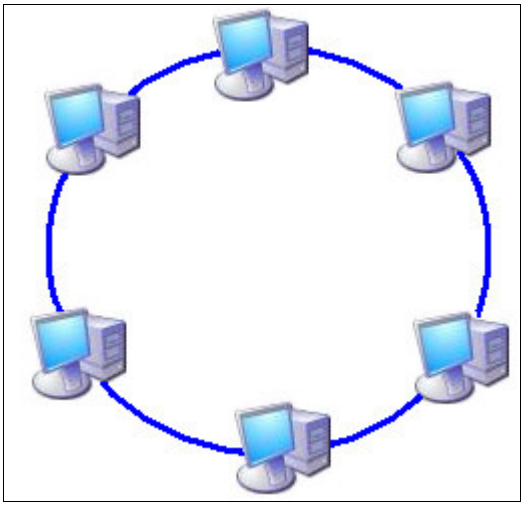


Figura 2.6 Representação gráfica de uma rede em anel

2.3.3. Protocolos de rede

Os protocolos de rede formam um conjunto de regras que definem os procedimentos, as convenções e os métodos utilizados para transmissão dos dados entre dois ou mais dispositivos em rede. A troca de dados entre dois dispositivos (origem e destino) começa na origem, onde o fluxo de dados para o destino é dividido em pequenos blocos, chamados "pacotes", que devem ser transmitidos pela rede até chegar no destino. No destino, esses pacotes são remontados e passam como um fluxo de dados para o sistema operacional do dispositivo no destino entregar para a aplicação apropriada. Tudo isso é especificado e controlado por vários protocolos.

Dos vários protocolos existentes que são usados nos sistemas computacionais atuais, destacamos:

- TCP/IP (*Transfer Control Protocol/Internet Protocol*) – protocolo padrão usado na maioria das redes locais, é o protocolo padrão da Internet.
- IPX/SPX (*Internetwork Packet Exchange/Sequenced Packet Exchange*) – protocolo padrão das primeiras redes Netware/Novell, foi muito usado na década de 1990 quando do surgimento das primeiras redes locais de computadores pessoais.
- NetBIOS (*Network Basic Input/Output System*) – protocolo padrão das redes locais baseadas no Microsoft Windows.

- SNA (*System Network Architecture*) – protocolo desenvolvido pela IBM em 1974. É muito usado nas redes de para comunicação com os *mainframes*.
- AppleTalk – protocolo padrão das redes locais de computadores pessoais da empresa Apple.

A seguir, detalha-se-á o conjunto de protocolos TCP/IP por sua grande difusão e por ser a pilha de protocolos adotados na Internet.

2.3.3.1. TCP/IP

TCP/IP é o nome geral de um conjunto de protocolos de comunicação, comumente chamado de "conjunto ou *suite* de protocolos TCP/IP". O nome refere-se principalmente a dois protocolos TCP (*Transmission Control Protocol)* e IP *(Internet Protocol).*

O TCP/IP tem sua origem em 1969 através de um projeto de pesquisa que havia se iniciado no início da década de 60 para a agência ARPA (*Advanced Research Projects Agency)* para o Departamento de Defesa dos Estado Unidos [20]. Como resultado, surgiu a rede ARPANET, uma rede experimental, que foi convertida em uma rede operacional em 1975, após ter demostrado seu sucesso.

Em 1983, o novo conjunto de protocolos TCP/IP foi adotado como um padrão, e todos os computadores da rede ARPANET passaram a utilizá-lo. Quando a ARPANET finalmente cresceu e se tornou a Internet, em 1990, o uso do TCP/IP espalhou-se principalmente após o lançamneto da versão UNIX de Berkeley que, além de incluir esses protocolos, colocava-os em domínio público para serem usados por qualquer organização.

Os protocolos do conjunto TCP/IP são muito conhecidos atualmente, pois fornecem transporte de dados para todos os serviços disponíveis na Internet. Alguns desses serviços incluem:

- Navegação e acesso a WWW (*World Wide Web*)

- Troca de correio eletrônico;
- Transferência de arquivoS;
- Entrega de notícis a grupo de usuários;
- Comunicação instantânea;
- Jogos interativos
- Comércio eletrônico.

O TCP/IP opera através do uso de uma pilha de protocolos. Essa pilha é a soma total de todos os protocolos necessários para comunicação entre dispositivos na rede. Na Tabela 2.6, pode-se ver esta pilha dividida em quatro camadas.

Tabela 2.6 Camadas do protocolo TCP/IP

Camada	Descrição
Aplicativo	Quando o usuário inicia uma transferência de dados, esta camada passa as solicitações para a camada de transporte. Como por exemplo, Telnet, FTP, e-mail, etc.
Transporte	TCP e UDP
Rede	Aqui são adicionados os endereços de IP de origem e destino para propósitos de roteamento.Protocolos IP, ICMP, IGMP.
Enlace	Efetua as verificações de erros sobre o fluxo de dados entre os protocolos acima e a camada física.

Todo esse processo emprega um sistema complexo de verificação de erros, tanto na dispositivo de origem como no destino. Cada camada da pilha pode se comunicar com a camada adjacente enviando e recebendo dados.

A seguir detalha-se um pouco mais or três protocolos básicos do TCP/IP, que são TCP, UDP e IP.

2.3.3.1.1. TCP

O TCP (*Transmission Control Protocol*) é um dos principais protocolos empregados na Internet. Ele facilita tarefas de missão crítica, como transferências de arquivo e sessões remotas através de um método chamado de transferência de dados assegurando que eles cheguem na

mesma seqüência e estado em que foram enviados. O TCP é um protocolo orientado à conexão, a conexão estabelecida entre o dispositivo solicitante e seu destinatário é feita através de um processo dividido em fases, freqüentemente referido como *handshake* de três partes.

O TCP fornece capacidades de verificação de erro através de um valor numérico gerado para cada bloco de dados transmitido. Se uma transferência não for bem-sucedida, e um erro for recebido, os dados são retransmitidos, a não ser que o erro seja fatal, quando a transmissão é normalmente interrompida. Da mesma forma, se nenhuma confirmação for recebida durante um período de tempo especificado, as informações também deverão ser retransmitidas.

O TCP provê um mecanismo que permite ao transmissor distinguir entre múltiplos receptores num mesmo equipamento destinatário.

2.3.3.1.2. UDP

O UDP (*User Datagram Protocol*) é um protocolo da camada de transporte. Ele é um protocolo muito mais simples que o TCP, pois oferece um serviço sem conexão e não confiável, pois não garantia de entrega de mensagem ao destinatário, nem que os dados chegarão em perfeito estado. Da mesma forma que o protocolo TCP, provê um mecanismo que permite ao transmissão distinguir entre múltiplos receptores num mesmo equipamento destinatário.

2.3.3.1.3. IP

O protocolo IP (*Internet Protocol*) é o protocolo básico usado pelo TCP/IP. Através dele todos os dados dos protocolos TCP, UDP, ICMP e IGMP são transmitidos como datagramas IP. O IP não oferece garantia de entrega de dados ao destinatário. Ele oferece um serviço que não é orientado à conexão. Ele pertence à camada de rede, fornece uma forma de transporte de datagramas da origem ao destino, independentemente das máquinas estarem na mesma rede ou não.

Como mostra a Figura 2.7, um datagrama IP é composto de um cabeçalho e uma área de dados. O cabeçalho contém uma área de dados variados e os endereços IP de origem e de destino. Esses elementos juntos formam um cabeçalho completo. A parte restante do datagrama contém os dados que estão sendo enviados.

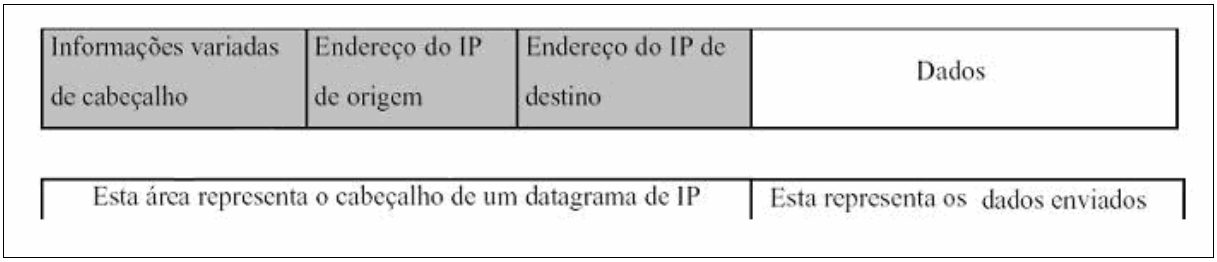


Figura 2.7 Datagrama IP

As outras informações contidas em um datagrama IP incluem: a identificação do protocolo utilizado, uma soma de verificação de cabeçalho (*checksum*), uma especificação de tempo de vida, tamanho total do datagrama e o nível de segurança da informação.

2.4. Tecnologias de rede

As tecnologias de redes de computadores aqui abordados serão aquelas relacionadas aos ambientes de armazenamento de dados. As principais tecnologias de transporte usadas nas redes de armazenamento de dados atualmente são Ethernet e Fibre Channel. A seguir essas duas tecnologias são descritas.

2.4.1. Ethernet

O padrão Ethernet [25], como conhecemos hoje, começou em julho de 1972, quando Bob Metcalfe foi trabalhar no Laboratório de Ciência da Computação no Centro de Pesquisa da Xerox em Palo Alto, EUA. Lá ele entra em contato com o trabalho do professor Norman Abramson e seus colegas da Universidade do Havai sobre uma rede de computadores denominada ALOHA [22]. Baseado nessa rede, no final de 1972, Metcalfe e seu colega David Boggs desenvolveram uma rede própria para conectar vários computadores da Xerox. Em 22 de Maio de 1973 a rede de

Metcalfe funcionou. Neste dia, ele escreve um memorando anunciando a rede e batizá-a com o nome da rede Ethernet, em referência à palavra "ether", meio pelo qual se imaginava, no passado, que as ondas eletromagnéticas se propagavam.

Em Junho de 1979, as empresas DEC, Intel e Xerox (DIX) fazem reuniões trilaterias a respeito da rede Ethernet. No ano seguinte, publicam a primeira especificação de Ethernet no livro "Ethernet Blue Book" ou "DIX Ethernet Versão 1.0". Dois anos depois, melhoram o padrão e publicaram a especificação "Ethernet Versão 2.0". Em Junho de 1981, o projeto 802 do IEEE[20] decidiu formar um subcomitê 802.3 para produzir um padrão de rede, baseado no trabalho da DIX, que pudesse ser aceito internacionalmente. Isso ocorre em 1983 com a primeira especificação de Ethernet como padrão IEEE 10BASE5. Esse nome foi escolhido, pois o padrão especificava uma velocidade de transmissão de 10 Mbps usando sinalização banda base e permitia distância entre nós de 500 metros. Os padrões IEEE para Ethernet são 10BASE5, 10BASE2 e 10BASE-F.

Desde o primeiro padrão IEEE, Ethernet tem evoluido continuamente. Várias novas especificações já foram publicadas e estão em uso. A seguir apresentamos as mais relevantes para os ambientes de redes de armazenamento de dados.

2.4.1.1. Fast Ethernet

O padrão Fast Ethernet manteve o padrão Ethernet no que se refere ao endereçamento, formato do pacote, tamanho e mecanismo de detecção de erro. As mudanças mais significativas em relação ao padrão Ethernet são o aumento de velocidade que foi para 100Mbps e o modo de transmissão que pode ser *half-duplex* ou *full-duplex*. Os padrões IEEE para Fast Ethernet são 100BASE-TX, 100BASE-T4 e 100BASE-FX.

[20] Institute of Electrical and Electronics Engineers

2.4.1.2. Gigabit Ethernet

O padrão Gigabit Ethernet foi ratificado IEEE em 1998. Ele foi desenvolvido para suportar o quadro padrão Ethernet, o que significa manter a compatibilidade com a base instalada de dispositivos Ethernet e Fast Ethernet, sem modificações. Possui taxa de transmissão de 1000 Mbps e, na sua essência, segue o padrão Ethernet com detecção de colisão e regras de repetidores. Aceita os modos de transmissão *half-duplex e full-duplex*. Os padrões IEEE para Gigabit Ethernet são 1000BASE-LX, 1000BASE-SX, 1000BASE-CX e 1000BASE-T.

2.4.1.3. 10-Gigabit Ethernet

O padrão 10-Gigabit Ethernet foi ratificado pelo IEEE em 2002. A idéia do novo padrão foi complementar os padrões Ethernet anteriores de 10, 100 e 1.000 Mbps, oferecendo uma solução capaz de construir redes de velocidade 10.000 Mbps, interligar redes distantes com uma velocidade comparável a dos *backbones* DWDM[21].

O padrão 10-Gigabit Ethernet segue o padrão Gigabit Ethernet, porém seu modo de transmissão é única e exclusivamente *full-duplex*. Originalmente o meio físico foi fibra óptica, podendo atingir até 40 Km em fibra monomodo e 300 metros em fibra óptica multimodo. Em 2004, foi estabelecido o padrão chamado 10GBASE-CX4, que possibilita a operação a velocidade de até 10000 Mbps em fio de cobre com distâncias até 15 metros entre dispositivos. Os padrões IEEE para 10-Gigabit Ethernet são 10GBASE-SR, 10GBASE-LX4, 10GBASE-LR, 10GBASE-ER, 10GBASE-SW, 10GBASE-LW w 10GBASE-EW.

Tabela 2.7 Comparativo de taxa de transmissão e distâncias das topologias Ethernet.

	Ethernet	Fast Ethernet	Gigabit Ethernet	10-Gigabit Ethernet
Taxa de transmissão	10 Mbps	100 Mbps	1.000 Mbps	10.000 Mbps
Fibra multímodo	2 Km	412 m (half-duplex) 2 Km (full-duplex)	500 m	300 m

[21] *Dense Wavelength Division Multiplexing* – tecnologia óptica usada para aumentar a banda sobre backbones existentes de fibra óptica.

Fibra monomodo	25 Km	2 Km	3 Km	40 Km
STP/Coaxial	500 m	100 m	25 m	na
UTP Categoria 5	100 m	100 m	100 m	5 m

2.4.2 Fibre Channel

Fibre Channel (FC) é o nome geral de um conjunto de padrões de comunicação [23] desenvolvido pelo ANSI e corresponde à uma tecnologia de comunicação de propósito geral, desenvolvida para atender às exigências relacionadas à demanda crescente por transferência de dados em alta velocidade, mas que vem sendo usada quase que exclusivamente em ambientes de armazenamento de dados. Na realidade Fibre Channel é a tecnologia de redes usada na implementação da redes de armazenamento de dados chamadas SAN.

Não se deve confundir FC com com o protocolo FCP, que é o protocolo da camada de aplicação do FC, o qual transporta comandos SCSI-3 para a transmissão, recepção e controle de blocos de dados entre sistemas computacionais e dispositivos de armazenamento de dados.

O uso da palavra *Fibre* e não *Fiber* no nome dessa arquitetura aconteceu porque, originalmente, essa arquitetura presumia o uso de fibra óptica (*fiber*) como meio físico de transporte dos dados, porém, no seu desenvolvimento, com a possibilidade de se usar fios metálicos para esse fim, o grupo de trabalho responsável pelo desenvolvimento resolveu utilizar uma nomenclatura que mostrasse que a arquitetura não indicava necessariamente cabos de fibra óptica, daí, usar a mesma palavra na forma européia, ou seja, *fibre*.

Fibre Channel refere-se ao meio físico ou cabeamento, métodos de conexão, topologias de rede, uma metodologia de acesso ao barramento, protocolos de controle de fluxo, reconhecimento e enquadramento, mecanismos de sinalização de baixo nível e esquema de codificação de *bits*. Ele define um transporte serial de dados *full-duplex* em velocidades de 1, 2 ou 4 Gbps.

Embora essa arquitetura seja chamada de Fibre Channel, ela não representa um canal (*channel*), tão pouco é uma topologia real de rede. Ela permite um esquema inteligente de interconexão, baseado em um comutador Fibre Channel, chamado *Fabric*, para conectar dispositivos. Tudo que uma porta Fibre Channel faz é gerenciar uma conexão simples ponto a ponto entre ela mesma e outra porta Fibre Channel ou a porta de um comutador Fibre Channel. Em geral, Fibre Channel tenta combinar o melhor desses dois métodos de comunicação em uma nova interface de entrada/saída que atenda às necessidades dos usuários, tanto de canal quanto de redes [3].

O desenvolvimento do Fibre Channel começou em 1988, quando o grupo de trabalho X3 do ANSI começou a trabalhar nos padrões dessa tecnologia. Na época, foram gerados cerca de 20 documentos, cada um definindo um aspecto do Fibre Channel. Foi somente em 1994 que o padrão foi aprovado pela ANSI .

Fibre Channel, usando a topologia laço arbitrado (que veremos mais à frente), está sendo usado para substituir as conexões tradicionais SCSI. Muitas empresas já implementam adaptadores SCSI para as várias plataformas e sistemas operacionais, e assim como unidades de discos, unidades de fita e dispositivos de armazenamento com interfaces Fibre Channel.

É um padrão aberto que suporta múltiplos protocolos, incluindo alguns de alto nível como o SCSI, FDDI, HIPPI e IPI. O Fibre Channel é capaz de gerenciar a transferência de dados para esses protocolos. Embora possa operar com velocidade desde 133 Mbps a 4 Gbps, hoje é mais utilizado com velocidades de 2 Gbps e caminhando rapidamente para 4 Gbps.

2.4.2.1 Topologias Fibre Channel

A topologia de rede do padrão Fibre Channel evoluiu da tradicional conexão ponto a ponto SCSI dos sistemas computacionais para um modelo em “laço arbitrado” (*arbitrated loop*) com banda compartilhada e, depois, para um modelo baseado em comutador que permite múltiplas conexões ponto a ponto. Todas as três topologias de rede são transparentes aos equipamentos conectados. Assim, no padrão Fiber Channel as três topologias possíveis são:

Ponto a ponto (*Point to Point*), Laço Arbitrado (*Arbitraded Loop*) e Comutador Fibre Channel (*Switch Fabric ou Cross Point*).

2.4.2.1.1. Topologia ponto a ponto

A topologia ponto a ponto é a mais simples em Fibre Channel. A Figura 2.8 mostra dois (e somente dois) dispositivos Fibre Channel conectados entre si diretamente. Um cabo conecta a porta de transmissão de um dispositivo à porta de recepção do outro dispositivo e vise-versa. É necessário que haja uma seqüência de inicialização das portas para que possa ocorrer qualquer transferência de dados entre elas.

Esse tipo de topologia oferece toda a banda da porta Fibre Channel para os dispositivos conectados, embora seja improvável que essa banda toda seja usada por um período longo de tempo.

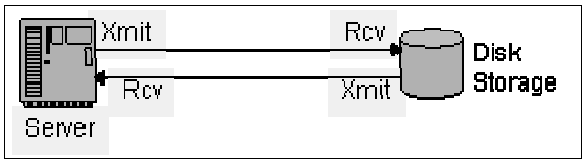


Figura 2.8 Topologia ponto a ponto

2.4.2.1.2. Topologia laço arbitrado

A topologia FC-AL (*Fibre Channel Arbitrated Loop*) é amplamente utilizada e, também, a mais complexa. Tornou-se muito utilizada porque é um modo muito econômico de se conectar até 127 portas Fibre Channel em uma única rede sem a necessidade de se usar um comutador. Ao contrário das outras topologias, o meio de comunicação é compratilhado entre os dispositivos conectados, limitando o acesso de cada dispositivo.

O uso do meio compartilhado por todos os participantes significa que qualquer dispositivo desejando transferir dados deve antes ganhar controle do meio. Este passo é possível com uma seqüência de arbitração. Durante a seqüência de arbitração, a prioridade de cada dispositivo

requisitando acesso ao meio é considerada, sendo que o mais baixo endereço tem a mais alta prioridade. Quando a seqüência de arbitração termina os dois dispositivos que desejam se comunicar, um em cada extremidade da transição, estabelecem uma conexão e controlam o meio. Quando a transição é completada, ambos liberam o controle do meio.

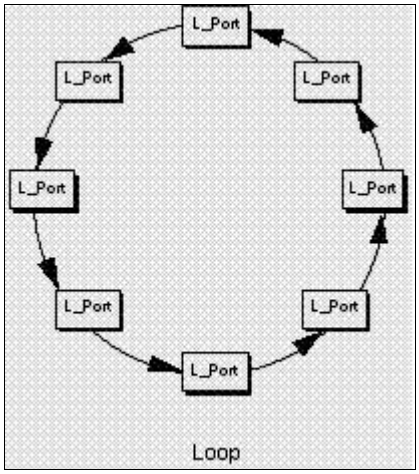


Figura 2.9 Topologia de laço arbitrado

Os dispositivos Fibre Channel estão ligados um ao outro numa configuração em anel usando um laço arbitrado, como mostra a Figura 2.9. Neste caso a porta de transmissão de um dispositivo conecta-se à porta de recepção do próximo dispositivo e assim por diante. O último dispositivo então se conecta ao primeiro dispositivo do mesmo modo. Note que a figura mostra as conexões lógicas somente, fisicamente não há cabo do último dispositivo ao primeiro.

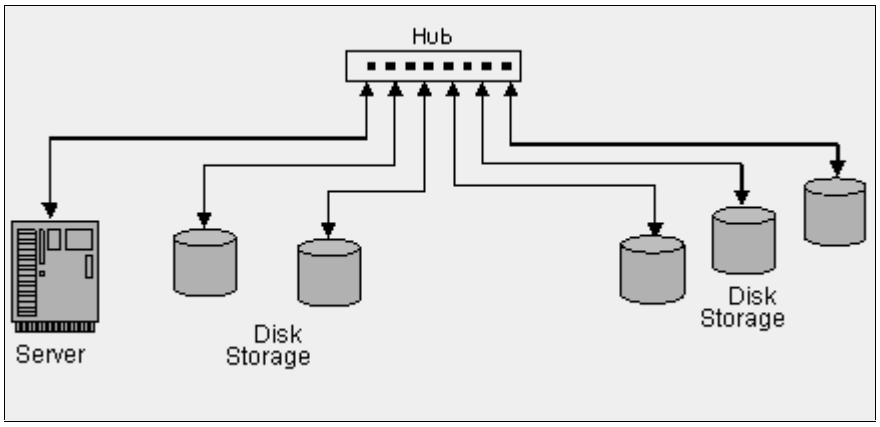


Figura 2.10 Topologia da conexão laço arbitrado da SAN usando *hub*

O cabeamento simples da topologia laço arbitrado na Figura 2.9 é de baixo custo embora a falha em um cabo, conexão ou componente de *hardware* em um único ponto pode fazer o laço inteiro falhar. Para minimizar esse tipo de falha, é que, frequentemente, se usa um *hub* Fibre Channel para configuração de laço arbitrado, como mostra a Figura 2.10.

2.4.2.1.3. Topologia comutador Fibre Channel ou *Fabric*

Um comutador Fibre Channel, chamado *Fabric*, conecta dispositivos Fibre Channel a um ou mais comutadores Fibre Channel. Nesta topologia, pode-se ter, teoricamente, até 2^{24} portas para conexão. Cada dispositivo faz uso de uma porta de conexão ponto a ponto e tem toda a banda disponível para ele, assim toda a banda agregada aumenta na medida que novos dispositivos são adicionados. A Figura 2.11 apresenta um exemplo da topologia *fabric* de uma SAN.

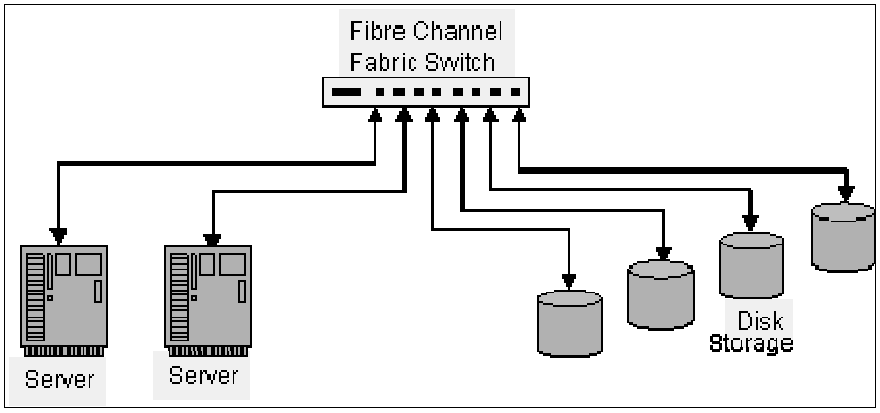


Figura 2.11 Topologia com conexão *fabric* de uma SAN

Do mesmo modo que na topologia ponto a ponto, os dispositivos capazes de se conectar a um comutador *fabric* devem estabelecer uma sessão usando uma seqüência de inicialização antes de transferir dados. Contudo, a seqüência de inicialização e os quadros de dados usados são diferentes nos dois casos.

Os comutadores *fabric* podem e freqüentemente fazem conexões laço arbitrado para outros laços e/ou dispositivos comutadores *fabric*.

Os termos "*switch director*", "*diretor-class switch*" e "*storage director*", comuns nos ambientes de grande porte para armazenamento de dados, referem-se a um comutador *fabric* com um grande número de portas. Enquanto um comutador padrão possui 8, 16 ou 32 portas, um comutador *director-class* possui 64, 128 e até 256 portas Fibre Channel.

2.5. Protocolos de redes de armazenamento

O objetivo dessa seção é apresentar os principais protocolos que, atualmente, estão sendo usados nas redes de armazenamento de dados. São protocolos de movimentação de blocos de dados ou arquivos pela rede.

2.5.1. NFS

O protocolo NFS (*Network File System*) é protocolo padrão para o compartilhamento e gerenciamento de arquivos através de redes baseadas em TCP/IP. Ele é padrão para nos sistemas operacionais baseados em UNIX, e com implementações em outros sistemas operacionais. Ele implementa um sistema de arquivos remoto, implementado na forma cliente/servidor, onde o servidor oferece uma área de armazenamento local que pode ser utilizada por um cliente como se fosse uma área local a ele. Foi projetado e inicialmente implementado pela Sun Microsystems [24].

O protocolo NFS tem um conjunto de procedimentos que permitem que um cliente tenha acesso transparente a arquivos armazenados num servidor remoto. Ele é independente da arquitetura do servidor, do sistema operacional, da rede, e do protocolo de transporte. Foi concebido pela Sun Microsystems, em 1984, e colocado no domínio público. Foi em 1989 que sua primeira especificação se tornou a RFC 1094. Em 1995, foi publicada a versão 3.0 do NFS através da RFC 1813. Já, em 2003, foi publicada a versão 4.0 do NFS através da RFC 3530, sendo atualmente a versão mais recente deste protocolo.

Algumas características gerais do protocolo NFS são:

- O protocolo foi projetado para ser *stateless*, ou seja, não é necessário manter ou guardar o estado da conexão entre cliente e servidor, tornando-o muito robusto para ambientes distribuídos;
- O protocolo foi projetado para suportar a semântica de sistema de arquivos dos ambientes Unix, porém é ficando limitado a esse sistema operacional;
- Os controles de acesso e proteção seguem a semântica de segurança do Unix usando a identificação do usuário (*userid*) e do grupo que ele pertence (*groupid*) para verificação da segurança, podendo usar ACLs[22] na versão 4.0 do NFS;
- O protocolo NFS é independente da camada de transporte. Amplamente usado em TCP/IP, embora fosse construído originalmente usando o protocolo de datagrama UDP, pode também ser implementado com TCP.

2.5.2. CIFS

O protocolo CIFS (*Common Internet File System*) é o protocolo padrão para o compartilhamento de arquivos através da rede para os computadores usando sistemas operacionais baseados em Windows [26]. É baseado no protocolo SMB (*Server Message Block*). Além das funções de compartilhamento de arquivos, ele também pode ser usado para serviços de impressão em rede, serviços de localização de recursos, gerenciamento e administração remota, autenticação na rede estabelecendo serviços seguros e RPC (*Remote Procedure Calls*).

Em 1984, a IBM escreveu o NetBIOS API, um conjunto de chamadas que permitia funções básicas de comunicação entre computadores numa pequena sub-rede. Essas chamadas exigiam um protocolo na camada de transporte para poder enviar e receber dados. A partir daí, a IBM em conjunto com a Microsoft e apoio da Intel e 3Com, continuaram no aprimoramento do que resultou no protocolo SMB.

[22] Access Control List

Em 1992, o SMB foi publicado pelo X/Open como protocolo padrão para interconexão de computadores pessoais através do documento "Protocols for X/OPEN PC Internetworking, SMB Version 2". No final de 1997, o IETF[23] publicou a versão final atualizada do protocolo, agora renomeado para CIFS/1.0 [27].

O protocolo CIFS roda sobre TCP/IP e incorpora uma semântica de alto desempenho, operações de leitura e escrita por múltiplos usuários, bloqueio (*locking*) e compartilhamento de arquivos. Ele permite que múltiplos clientes acessem um mesmo arquivo, ao mesmo tempo em que previne conflitos de acesso ao implementar uma semântica de compartilhamento de arquivos e bloqueio. Estes mecanismos permitem também um cacheamento consistente de dados.

Com o protocolo CIFS, os usuários não precisam montar o sistema de arquivos remoto, bastando referenciá-lo diretamente através do nome global. Para um cliente acessar o sistema de arquivos de um servidor basta usar um mecanismo de resolução de nome para identificar o servidor na rede. A especificação define dois modos para isso: mecanismo de resolução de nomes por DNS ou NetBIOS.

O protocolo requer que os usuários façam uma requisição de autenticação antes de terem acesso o arquivo, e cada servidor autentica seus próprios usuários. Ele define dois métodos de segurança de acesso aos arquivos que pode ser em nível de compartilhamento e em nível de usuário através de ACLs.

2.5.3. FCP

O protocolo FCP (*Fiber Channel Protocol*), que reside logicamente na camada mais alta do Fibre Channel, define endereçamento de dispositivos, recuperação de erros, seqüenciamento de comandos e conteúdo de dados nos quadros Fibre Channel individuais. O protocolo FCP transfere comandos e blocos de dados SCSI através de infra-estrutura Fibre Channel, sendo uma conexão serial de alta velocidade [28].

[23] *Internet Engineering Task Force*

2.5.4. iSCSI

O protocolo iSCSI (*Internet Small Computer System Interface*) é um padrão proposto pela IEFT, através do grupo de trabalho IPS (*IP Storage*), que inclusive é reponsável pelos protocolos iFCP e FCIP que veremos a seguir. O protocolo iSCSI define as regras e processos para transmitir e receber bloco de dados de armazenamento sobre redes baseadas em TCP/IP. iSCSI substitui o esquema de cabeamento paralelo SCSI por tráfego em rede. Servidores e dispositivos de armazenamento iSCSI podem ser conectados diretamente a uma infra-estrutura de rede baseada em roteadores e comutadores existentes [28]. Na Figura 2.12, pode-se ver um exemplo de rede com protocolo iSCSI.

O trabalho do IETF começou em 2000 com um esforço conjunto de várias empresas, incluindo Cisco Systems, IBM e HP. Sendo o protocolo iSCSI ratificado em 2003, posteriormente tornou-se a RFC 3720.

Um primeiro impulsionador de mercado para o desenvolvimento do protocolo iSCSI é permitir o acesso mais amplo à imensa base instalada de DAS sobre infra-estruturas IP. Ao permitir esse acesso, esses recursos de armazenamento podem ser maximizados por qualquer quantidade de usuários ou utilizados por uma variedade de aplicativos como *backup* remoto, recuperação de desastres e virtualização de armazenamento.

Um segundo fator de motivação para o protocolo iSCSI é permitir outras arquiteturas SAN, como Fibre Channel, sejam acessadas a partir de uma ampla variedade de servidores sobre redes baseadas em TCP/IP. O protocolo iSCSI permite o armazenamento em nível de blocos usando roteadores ou comutadores IP, ampliando sua aplicabilidade, como um protocolo de armazenamento baseado em IP.

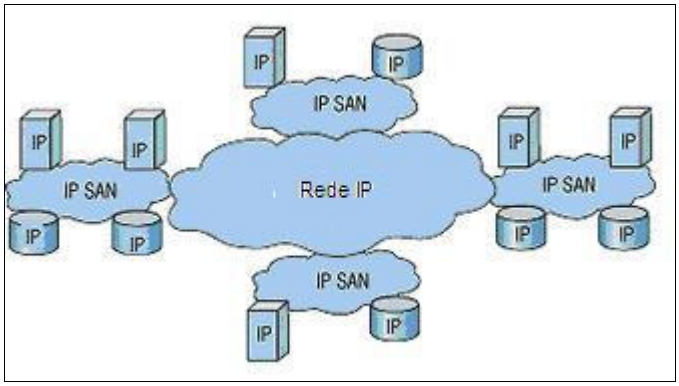


Figura 2.12 Implementação de SANs iSCSI

Na Figura 2.12, pode-se ver uma rede IP conectando redes de dados IP (IP SAN). Na implementação das IP SAN o protocolo usado é o iSCSI. A infra-estrutura que suporta essas redes são os comutadores e/ou roteadores Gigabit Ethernet padrão. Assim, é possível usar uma rede existente para uma implementação de redes iSCSI, reduzindo o custo total de propriedade (TCO) para esse tipo de ambiente.

2.5.5. FCIP

O protocolo FCIP (*Fibre Channel over TCP/IP*) descreve mecanismos de interconexão redes SANs baseadas em Fibre Channel através de redes baseadas em TCP/IP, formando uma rede unificada SAN [28]. O protocolo FCIP foi publicado em Julho de 2004 como RFC 3821. Na Figura 2.13, pode-se ver uma rede IP sendo usada para interconectar redes Fibre Channel (FC SAN). Neste caso o protocolo FCIP é o protocolo usado para interconectar redes FC SANs.

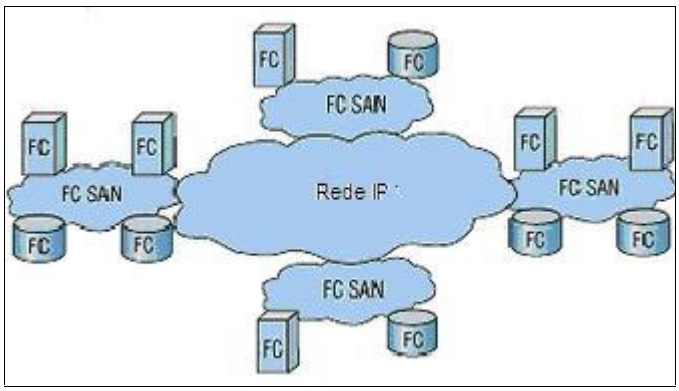


Figura 2.13 Implementação de SANs Fiber-Channel

O FCIP é um protocolo de tunelamento que encapsula pacotes Fibre Channel e os transportam através de TCP/IP, o que permite que aplicativos desenvolvidos para SANs Fibre Channel sejam suportados pelo FCIP sem alterações, pois mantém os serviços Fibre Channel intactos. O FCIP confia nos serviços de rede baseados em TCP/IP para controle e gerenciamento de congestionamento.

2.5.6. iFCP

O iFCP (*Internet Fibre Channel Protocol*) é um protocolo do IETF para conexão de dispositivos Fibre Channel em uma rede baseada em TCP/IP [28]. O padrão iFCP foi publicado na RFC 4172. O iFCP usa o protocolo FCP, do padrão Fibre Channel, mas implementado em redes TCP/IP. Ele usa a mesma técnica de encapsulamento de quadro que o FCIP e depende do TCP para controlar o congestionamento e manipulação de detecção e recuperação de erro. Quando necessário, o iFCP também intercepta e emula os serviços *fabric* exigidos por um dispositivo Fibre Channel. Na Figura 2.14, pode-se ver uma rede IP sendo usada para interconectar redes Fibre Channel (FC SAN). Neste caso o protocolo iFCP é o protocolo usado para interconectar redes FC SANs.

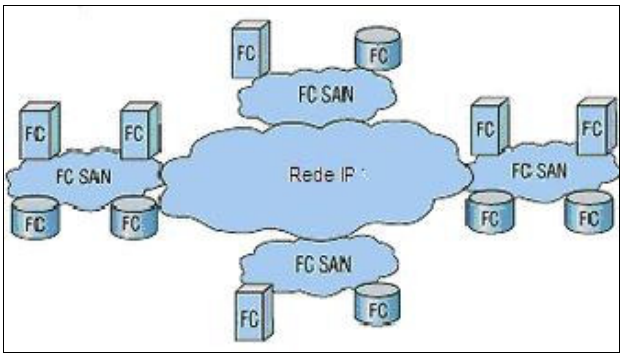


Figura 2.14 Exemplo de uso do protocolo iFCP

Como no caso do protocolo FCIP, o principal impulsionador de mercado para iFCP é a grande base instalada de dispositivos Fibre Channel, em combinação com a tendência na direção das redes de armazenamento IP. O padrão iFCP aproveita a estrutura e a interoperabilidade do protocolo Fibre Channel, ao mesmo tempo que aproveita as vantagens das redes baseadas em TCP/IP.

3. Arquiteturas de Armazenamento de Dados

"A rede é realmente o computador, e o movimento para redes de armazenamento de dados é um passo evolucionário na computação empresarial"

Daniel J. Wordem, 2004.

Chamamos de arquiteturas de armazenamento de dados as formas pelas quais pode-se conectar os dispositivos de armazenamento de dados aos sistemas computacionais. A conexão entre sistema computacional e dispositivo de armazenamento requer um meio de ligação ponto a ponto entre eles. No dispositivo de armazenamento de dados, estabelece-se uma área de armazenamento que é disponibilizada, em geral, para uso exclusivo do sistema computacional.

As arquiteturas de armazenamento de dados são nomeadas de acordo com as formas pelas quais são conectados os dispositivos de armazenamento de dados aos sistemas computacionais e pelo tipo de informação que é trocada entre eles. Este capítulo apresenta as arquiteturas de armazenamento de dados: DAS, SAN e NAS que são, atualmente, as formas amplamente utilizadas para disponibilização de armazenamento de dados aos sistemas computacionais.

3.1. O modelo SNIA de armazenamento de dados compartilhado

Para apresentar as arquiteturas de armazenamento de dados usar-se-á um modelo desenvolvido pelo conselho técnico da SNIA [29]. A SNIA (*Storage Network Industry Association*) é uma entidade sem fins lucrativos, criada em 1997, composta por mais de 100 entidades internacionais que colaboram no desenvolvimento da indústria de armazenamento de dados em rede. O objetivo da SNIA é desenvolver soluções confiáveis e completas de armazenamento e gerenciamento de dados para a comunidade de projetistas e administradores dos sistemas computacionais. A SNIA busca o desenvolvimento de soluções baseadas em padrões de armazenamento de dados que tenham potencial de serem amplamente utilizadas e o foco seja o usuário final.

O modelo criado pela SNIA, semelhante ao modelo OSI de sete camadas para redes de computadores, é chamado de Modelo SNIA de Armazenamento Compartilhado de Dados, e pode

ser visto na Figura 3.1. Este modelo apresenta uma estrutura genérica para arquitetura de armazenamento de dados, incluindo sistemas de armazenamento de dados distribuídos. Ele estabelece um relacionamento geral entre as aplicações dos usuários, que rodam nos sistemas computacionais, e os sistemas de armazenamento de dados. Através desse modelo é possível comparar as arquiteturas de armazenamento de dados a partir de um vocabulário comum.

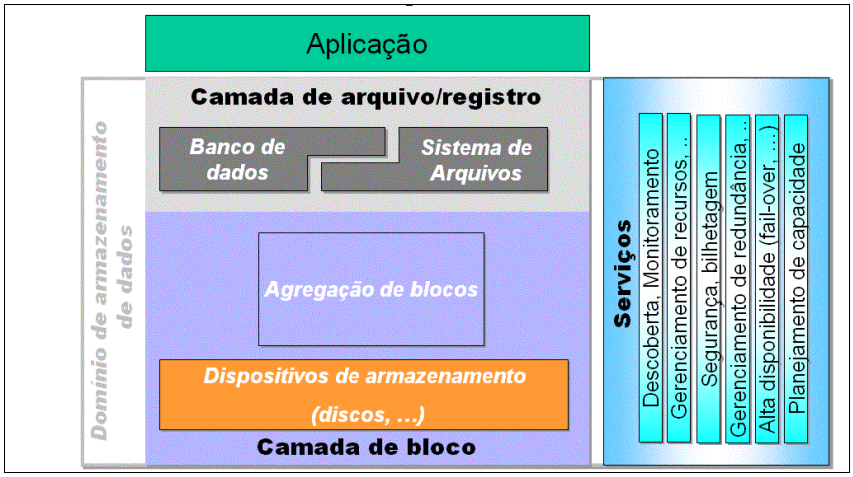


Figura 3.1 Modelo SNIA de armazenamento de dados compartilhado

A área de TI das organizações pode ter um grande desafio para entender, de forma coerente, o papel das aplicações e das arquiteturas armazenamento e gerenciamento de dados. Assim, o modelo oferece uma forma prática e útil para apresentar as relações entre aplicações de alto nível e as respectivas infra-estruturas de armazenamento de dados. A habilidade de mapear as atuais implementações de armazenamento de dados para soluções propostas ajuda a esclarecer o que as arquiteturas estão endereçando e também cria uma base de conhecimento para se projetar futuras necessidades e soluções.

Como se pode ver, o modelo SNIA estabelece um relacionamento entre a aplicação do usuário, que é executada no sistema computacional, e o domínio de armazenamento de dados sob ela. A aplicação pode suportar várias atividades dos usuários, tais como processamento de transações de banco de dados, serviços de arquivos etc.

O Domínio de armazenamento de dados é subdivido em duas camadas:

- **A Camada Arquivo/Registro –** é a interface entre o nível mais alto de aplicações e os recursos de armazenamento (seqüência de *bytes* de informação organizados como registros ou arquivos). Em geral, as unidades de dados manipuladas pelos bancos de dados é o registro, enquanto a maioria das outras aplicações manipula arquivos. Essa camada é controlada somente pelo servidor em armazenamento de dados tradicional, ou por um ambiente cliente/servidor em armazenamento de arquivos em rede. Atualmente as duas implementações mais comuns para esse ambiente cliente-servidor são os ambientes NFS e CIFS.
- **A Camada de Blocos –** é a camada de baixo nível do armazenamento onde os blocos de dados (registros ou arquivos) são gravados ou lidos no dispositivo físico de armazenamento de dados. Esta camada é que realiza as várias as funções de agregação sobre os dispositivos, como é o caso das implementações de proteção RAID.

O subsistema serviços agrupa as aplicações específicas do armazenamento de dados como monitoramento e gerenciamento do sistema de armazenamento, gerenciamento de recursos, configuração, segurança, utilização, gerenciamento de redundância, alta disponibilidade e planejamento de capacidade. Assim, o modelo distingue entre aplicação do usuário final, que é de nível mais alto, e aplicações secundárias, que são usadas para gerenciamento e suporte do Domínio de Armazenamento de Dados no nível mais baixo.

Usando a arquitetura estruturada em níveis do modelo SNIA de armazenamento de dados compartilhado, é possível inserir componentes de servidor ou armazenamento para claramente diferenciar as configurações DAS, SAN e NAS, como será mostrado a seguir.

3.2. Arquitetura DAS

DAS (*Direct Attached Storage*) é o termo usado para descrever dispositivos de armazenamento de dados que são conectados diretamente a um servidor. A forma mais simples de DAS é o disco interno de um computador, embora os dispositivos de armazenamento de dados num gabinete de discos externo também possam ser classificados nessa categoria.

Como se pode ser ver, a arquitetura DAS tradicional conecta, a partir do servidor, o dispositivo alvo do armazenamento através de uma conexão. A tecnologia SCSI paralelo é uma forma muito comum de uso dessa arquitetura.

A arquitetura DAS, sendo a primeira arquitetura de armazenamento de dados inventada, ainda possui um grande número de sistemas instalados e usuários, porém com o desenvolvimento de novas arquiteturas, ela vem sendo substituída constantemente ao longo dos últimos anos, como será apresentado no Capítulo 4.

Como se pode observar na Figura 3.2, tem-se do lado esquerdo um servidor com volumes lógicos (LVM – *Logival Volumes*) e proteção de discos RAID executada por *software*. O LVM é responsável pela abstração da camada de discos mostrando uma imagem coerente dos dados à camada de aplicação no nível superior na forma de volume, com diretórios e subdiretórios. O servidor recebe informações da aplicação que devem ser escritas nos discos, o *software* RAID distribui blocos de dados pelos os discos da camada de blocos.

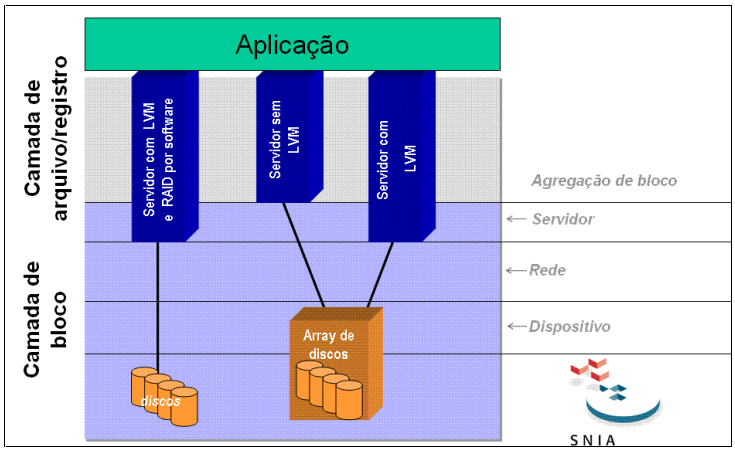


Figura 3.2 Modelo SNIA da arquitetura DAS

O servidor do lado direito da figura possui conexão direta aos discos, contém uma controladora de discos integrada com capacidade de implementar proteção e incremento de desempenho de discos baseado em tecnologia RAID. O LVM é implementado como uma função do sistema operacional do servidor. Por outro lado, na solução do lado esquerdo, o *array* de discos por si só executa a função de RAID pela sua própria controladora de discos. A

implementação de LVM sobre conjunto de discos, dependendo da necessidade, pode ou não ser feita no servidor.

3.3. Arquitetura SAN

SAN (*Storage Area Network*) é o termo usado para a arquitetura de armazenamento de dados onde existe uma infra-estrutura de rede ligando os servidores aos dispositivos de armazenamento de dados como se pode-se observar na Figura 3.3. O tipo de informação que trafega nesta rede é o bloco de dados. Na definição da SNIA sobre SAN, qualquer tipo de rede pode ser usada, no entanto, atualmente, as redes baseadas em Fibre Channel e Gigabit Ethernet com TCP/IP são as mais comuns. As SANs possibilitam conexões de alta velocidade entre os servidores e os dispositivos de armazenamento de dados, suporte a longas distâncias, conexões ponto a ponto e habilidade de implementação de consolidação de armazenamento, compartilhamento de recursos e alta disponibilidade.

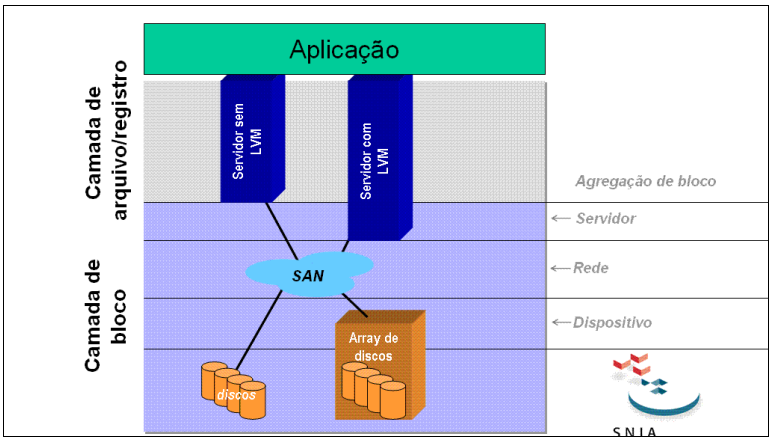


Figura 3.3 Modelo SNIA da arquitetura SAN

3.4. Arquitetura NAS

NAS (*Network Attached Storage*) é o termo usado para a arquitetura de armazenamento de dados na qual existe uma infra-estrutura de rede ligando os servidores aos dispositivos de armazenamento de dados, como se pode ver na Figura 3.4. O tipo de informação que trafega nesta rede é o arquivo. De acordo com a definição da SNIA de NAS, os tipos de redes usadas

nesta arquitetura são as redes que permitem trocas de arquivos entre seus nós, como é o caso das locais (LANs). Atualmente, as redes baseadas em Gigabit Ethernet e TCP/IP são as mais comuns, o que permite a conexão dos servidores aos dispositivos de armazenamento de dados com o benefício de altas velocidades, suporte a longas distâncias, conexões ponto a ponto e habilidade de implementação de consolidação de armazenamento, compartilhamento de recursos e alta disponibilidade.

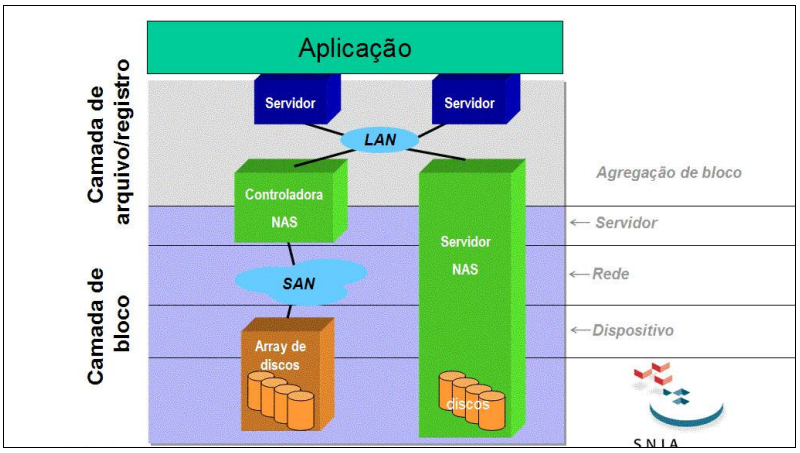


Figura 3.4 Modelo SNIA da arquitetura NAS

A arquitetura de armazenamento de dados NAS baseia-se em dispositivos de armazenamento de dados que se comunicam com os servidores através dos protocolos de compartilhamento de arquivos na rede. O controle do sistema de arquivos é responsabilidade do próprio dispositivo de armazenamento.

Um equipamento NAS pode ser implementado de duas maneiras distintas:

- **Servidor NAS** – é composta de processador integrado a um conjunto de discos próprios. O servidor NAS atua desde a camada de arquivo/registro até a camada de bloco, na qual se faz a agregação dos dispositivos de blocos (os discos). Toda conexão interna ao dispositivo NAS, entre o processador e os discos, é direta e dedicada.
- **Controladora NAS** – também chamada de "*NAS-head*", é composta de um processador conectado à uma rede SAN, que, por sua vez, se conecta a um subsistema de discos

independente. O armazenamento é efetuado em discos que fazem parte de uma SAN interna, chamada de SAN "*back-end*".

Independente da forma de conexão do processador do dispositivo NAS com os discos, a sua conexão aos computadores é sempre a mesma e pode ser feita a partir de rede local (LAN) ou mesmo rede de longa distância (WAN), dedicada ou não, mas que suporte o conjunto de protocolos TCP/IP. Os computadores acessam os dados do dispositivo de armazenamento NAS através de protocolos especializados de acesso e compartilhamento de arquivos. As requisições de arquivos recebidas pelo NAS são traduzidas pelo processador interno em requisições aos discos.

Atualmente, os dois principais protocolos de compartilhamento de arquivos na rede são CIFS e NFS, para os clientes Microsoft Windows e Unix, respectivamente.

3.5. Combinação de arquiteturas

No mundo real das organizações, as arquiteturas utilizadas são variadas, por isso o modelo SNIA permite, também, a modelagem de ambientes mistos com as arquiteturas DAS, SAN e NAS. Na Figura 3.5, usa-se o modelo SNIA para apresentar um ambiente que possui simultaneamente as três arquiteturas de armazenamento de dados.

Pode-se ver que todas as arquiteturas podem ser facilmente mapeadas e entendidas a partir do modelo desenvolvido pela SNIA. Como exemplo, pode-se ver que aplicações atuais que rodam em dispositivos DAS podem ser facilmente redesenhadas com componentes SAN ou NAS. O valor prático dessa modelagem é que se pode facilmente mapear as atuais aplicações em novas infra-estruturas mais flexíveis, escalonáveis e robustas.

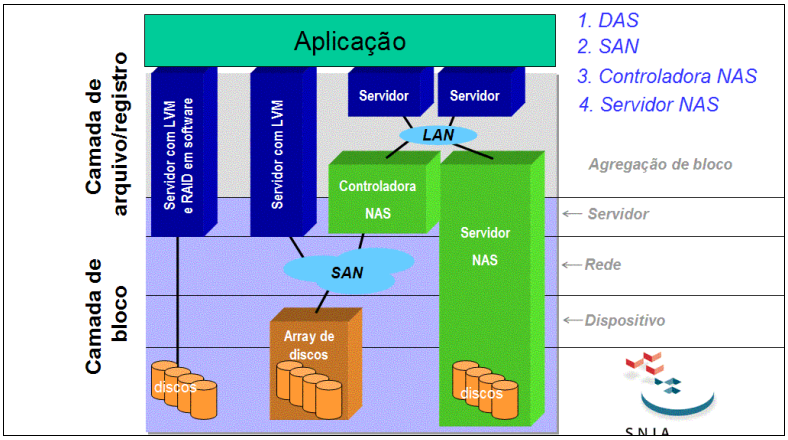


Figura 3.5 Modelo SNIA da arquitetura mista DAS, SAN e NAS

É possível mapear e comparar diretamente as arquiteturas de armazenamento de dados sob o mesmo esquema. Isto oferece uma visão geral dos requerimentos e das opções que podem ser usadas no desenvolvimento de ambientes de armazenamento de dados.

4. Comparação de Arquiteturas de Armazenamento de Dados

"A convergência das arquiteturas SAN e NAS será uma realidade no curto prazo."
Salomon Smith Barney, 2001.

No capítulo anterior, apresentou-se as arquiteturas de armazenamento de dados de um ponto de vista formal utilizando o modelo de referência elaborado pela SNIA. O objetivo deste capítulo é apresentar essas arquiteturas de uma óptica mais prática, fazendo uma comparação entre elas.

Entre as possíveis arquiteturas de armazenamento de dados existe um divisor bem claro do que é considerado o modelo do passado, ou seja, a arquitetura DAS versus o modelo, cada vez mais adotado, baseado em armazenamento centralizado e compartilhado que utiliza infra-estruturas de rede para conectar servidores aos dispositivos de armazenamento de dados, ou seja, as arquiteturas SAN e NAS.

Um novo termo, FAS (*Fabric Attached Storage*), foi usado pelo Gartner Dataquest [30] para mostrar a integração e convergência das arquiteturas SAN e NAS. Esse termo representa hoje uma conceito de unificação destas arquiteturas de armazenamento de dados em rede.

4.1. Arquitetura DAS

A arquitetura DAS representa um armazenamento de dados de acesso restrito e isolado a um único computador, em raras oportunidades é utilizado com dois computadores, como é o caso do chamado disco quorum nos sistemas clusterizados. Nela, todo acesso aos dados no dispositivo de armazenamento depende do servidor. O tráfego de dados entre servidor e dispositivo de armazenamento passa por um barramento de entrada/saída, em geral SCSI ou IDE/ATA. Assim, um dispositivo DAS está intimamente ligado ao sistema operacional do servidor e limitado às distâncias que os dispositivos de armazenamento podem ficar do servidor. Normalmente esta arquitetura está entre as mais baratas em termos de aquisição, e por ser a forma tradicional de armazenamento é amplamente conhecida e utilizada.

A forma mais simples de DAS são os discos internos de um computador, embora discos num gabinete externo conectado a um computador também possam ser classificados nessa maneira.

Uma evolução do armazenamento de dados DAS consiste na consolidação do armazenamento de vários servidores num único dispositivo. Esse processo de consolidação é visto quando se migra do armazenamento local (Figura 4.1) para um armazenamento consolidado (Figura 4.2).

Na Figura 4.1, pode-se observar a arquitetura DAS com armazenamento de dados local a cada servidor, enquanto que na Figura 4.2, tem-se a situação com um dispositivo de armazenamento consolidado que é compartilhado/dividido entre todos os servidores. Nas duas figuras, a sigla "FS" significa *file system* (sistema de arquivos), ou seja, sistema de arquivos, que envolve em regras de distribuição de informações fisicamente nos discos. São exemplos de sistemas de arquivos: FAT, NTFS, UFS, XFS etc. Na arquitetura DAS, isso significa que o servidor é responsável pela distribuição de arquivos e blocos de dados fisicamente nos discos.

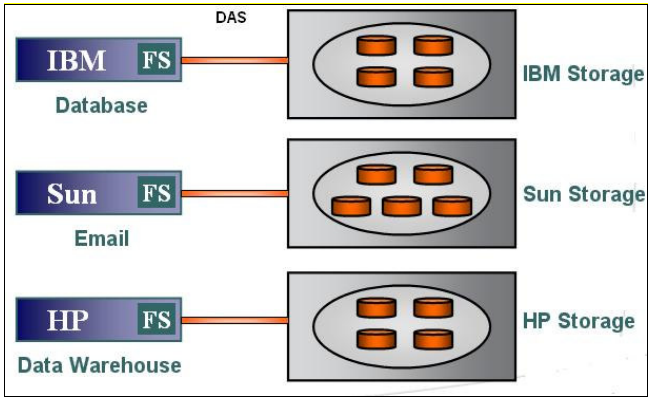


Figura 4.1 Arquitetura DAS com armazenamento local

A arquitetura DAS é, ainda hoje, o método mais conhecido de armazenamento de dados. Entre as décadas de 1980 e 1990, os departamentos de informática das organizações implantaram diversos servidores com armazenamento DAS para atender os requisitos específicos do negócio das organizações. Criaram-se "ilhas de servidores" e "ilhas de dispositivos de armazenamento" para atender aos requisitos específicos do negócio, com muito pouca antecipação de como os dados cresceriam ou seriam compartilhados no futuro.

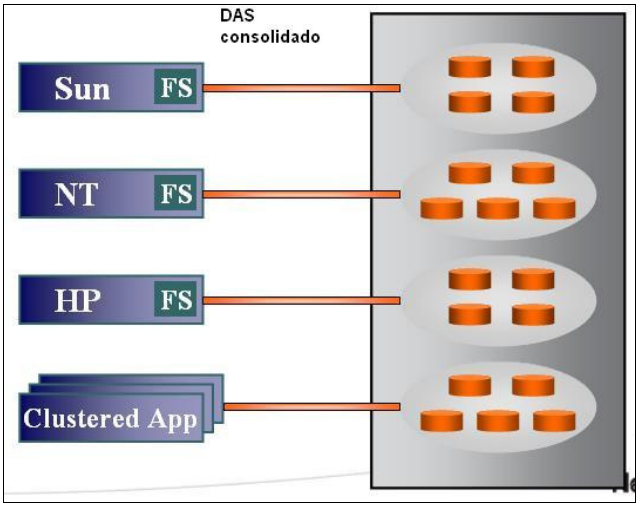


Figura 4.2 Arquitetura DAS com armazenamento consolidado

É muito oneroso gerenciar as ilhas de armazenamento de dados por toda a organização empresarial. Os administradores dos sistemas computacionais devem gerenciar um dispositivo de armazenamento de cada vez. Acesso aos dados armazenados é possível somente através do servidor conectado ao dispositivo de armazenamento, e os servidores são, geralmente, os limitantes do desempenho dessa topologia. Nessa arquitetura, não é possível agregar de forma simplificada a capacidade de armazenamento para otimizar a sua utilização. Por exemplo, se a capacidade utilizada de um servidor é alta e de outro servidor baixa, não há como realocar a capacidade extra de armazenamento do servidor sobre-utilizado para o servidor subutilizado.

Com o passar do tempo essa solução começou a mostrar limitações, uma das principais é ser um "ponto de falha" para o ambiente, ou seja, quando ocorre algum problema no servidor ou no dispositivo de armazenamento, todos os dados armazenados se tornam indisponíveis, e pode levar horas, dias, ou mesmo semanas para resolver o problema e recuperar o ambiente para que os usuários possam acessar novamente os dados.

Algumas das principais características do armazenamento DAS são:

- Armazenamento tradicional e dedicado, conectado gerenciado por um único servidor;
- Configuração simples;
- Baixo custo com desempenho razoável;
- Relação "1 para 1" de servidor para equipamento de armazenamento;

- Conectividade – ligação direta; e
- Outros computadores só podem acessar os dados do servidor através rede.
- Modo de implementação tradicional de armazenamento.

Numa organização com múltiplos computadores, a arquitetura DAS pode parecer inicialmente de baixo custo considerando apenas as conexões individuais. Contudo, numa perspectiva mais ampla da organização de forma global, ao se calcular o custo total de propriedade (TCO[24]) da solução DAS, veremos que este é maior quando comparado às outras soluções, devido à: dificuldade no compartilhamento da capacidade ociosa com outros computadores, falta de escalabilidade e a falta de um ponto central de gerenciamento de múltiplos sistemas de armazenamento.

Na Figura 4.3, pode-se ver um dos resultados do estudo "Storage Report", feito pelas empresas Merrill Lynch & Co. E McKinsey & Co., em 2001 [5]. Para um TCO de três anos, o custo da intervenção manual necessária para compartilhar informações num ambiente DAS é de 84 centavos de dólar por megabyte, muito maior que o custo típico de 35 a 38 centavos de dólar para uma solução de armazenamento em rede NAS e SAN, respectivamente.

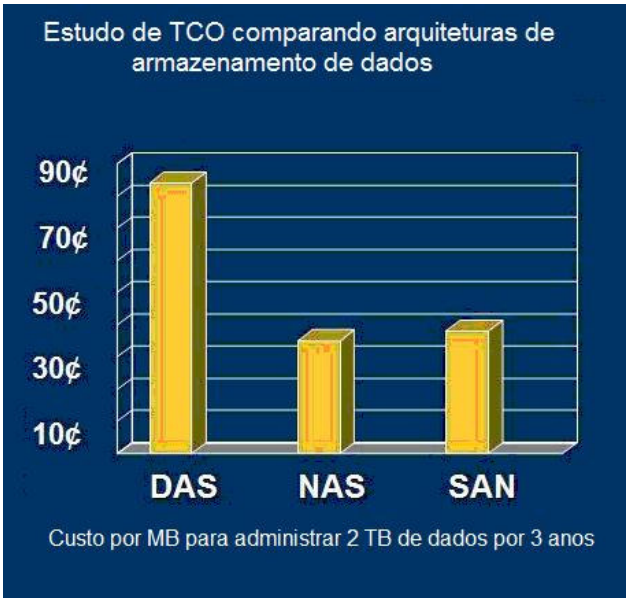


Figura 4.3 Comparativo de custo das arquiteturas de armazenamento. Fonte estudo de Merrill Lynch e McKinsey.

[24] Total Cost of Ownership

Apesar de ainda ser uma tecnologia bastante utilizada nas organizações empresariais, a arquitetura DAS apresenta dificuldades no armazenamento e compartilhamento de dados. Apesar da familiaridade e baixo custo de aquisição, os sistemas computacionais atuais tem demando soluções de armazenamento e compartilhamento de dados que incluam habilidade a para suportar vários sistemas operacionais e diferentes plataformas, acesso universal aos dados, alta escalabilidade, alto desempenho e administração centralização.

A arquitetura DAS é limitada à capacidade de crescimento de armazenamento do servidor, tanto pelo baixo número de portas de conexão, quanto pela limitação de espaço físico do servidor para acrescentar mais discos. Assim, é preciso adicionar mais servidores para atender uma necessidade de crescimento de armazenamento de dados, tornando a escalabilidade muito limitada para essa arquitetura.

Nesta arquitetura, quando um servidor deixa de funcionar ou é desligado para expansão ou manutenção, todos os dados armazenados nele se tornam indisponíveis. A redundância de dados entre servidores requer uma cópia adicional de dados para cada servidor. Existem, também, limitações associadas ao comprimento do cabeamento de conexão e número de discos. Dependendo do tipo da interface de conexão aos discos, se for SCSI, a limitação é de 25 metros de distância entre o servidor e o dispositivo de armazenamento. Neste caso, o número de discos suportados é de 15 por interface, o que limita a expansão do armazenamento quando necessária. O fato dos discos estarem conectados em cascata traz problemas adicionais no caso de manutenção, adição ou remoção de discos do sistema, pois pode tornar indisponíveis todos os discos em cascata, o que implica na indisponibilidade dos dados ali residentes. Um outro problema é que em um barramento único o dispositivo mais lento determina o desempenho do barramento como um todo.

Todas estas dificuldades tem ainda pouco significado ao comparar com o uso exclusivo dos recursos de armazenamento pelo computador a ele conectado. Por exemplo, um servidor de arquivos não pode ter sua capacidade processamento aproveitada para outra aplicação sem comprometer o desempenho das solicitações de acesso, recuperação e atualização dos usuários.

Apesar dessas dificuldades, o armazenamento em conexão direta representa uma tecnologia bastante utilizada nas organizações em geral, pois é familiar, fácil de comprar, não tem custo de aquisição muito alto e, até o aparecimento dos ambientes de alta disponibilidade, tem servido para atender às necessidades de armazenamento de dados.

4.2. Arquitetura FAS

A arquitetura FAS é o termo usado para representar a convergência das arquiteturas SAN e NAS. Essas duas arquiteturas têm como característica comum a existência de uma infra-estrutura de rede ligando os servidores aos dispositivos de armazenamento de dados.

Na Figura 4.4, pode-se ver a representação da topologia da arquitetura FAS. De um lado tem-se os clientes do armazenamento, ou seja, os servidores nos quais rodam as aplicações computacionais. Do outro, tem-se o dispositivo de armazenamento de dados no qual são criadas áreas de armazenamento, de uso exclusivo ou compartilhado, para cada um dos servidores.

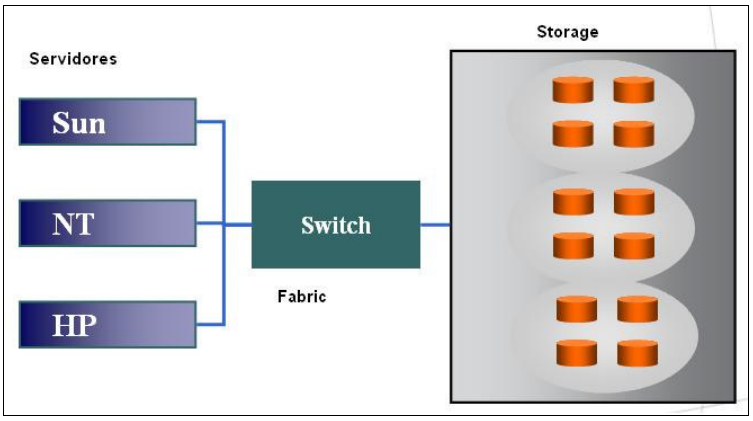


Figura 4.4 Modelo genérico de FAS.

O comutador ligando os servidores ao dispositivo de armazenamento tem a função de disponibilizar as áreas de dados criadas no dispositivo de armazenamento de dados para os respectivos servidores. A denominação *Fabric* vem do comutador Fibre Channel usado nos primeiros ambientes SAN, porém, também estão nessa categoria os comutadores Gigabit Ethernet, que hoje são muito usados tanto nas redes de computadores, quanto nas redes de armazenamento de dados.

Na Figura 4.5, pode-se ver uma implementação do padrão de FAS usando um comutador Fibre Channel. O protocolo de compartilhamento de dados entre servidor e dispositivo de armazenamento é o FCP. Esta implementação é chamada de arquitetura SAN, os tipos de dados que são trocados entre servidor e dispositivo de armazenamento são “bloco de dados”, em geral, dados e comandos SCSI. Por esse motivo SAN é considerado um ambiente de entrada/saída de dados em baseados em blocos.

Na arquitetura SAN, como no caso da arquitetura DAS, a responsabilidade de cuidar dos sistemas de arquivos (FS) fica a cargo do próprio servidor, o dispositivo de armazenamento se encarrega somente de fornecer uma área de armazenamento de bloco de dados.

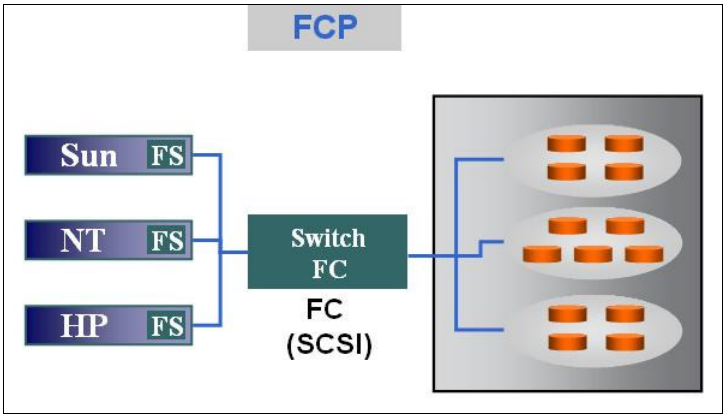


Figura 4.5 Modelo de uma SAN com comutador Fibre Channel

Na Figura 4.6, pode-se observar a implantação de uma SAN com um comutador Gigabit Ethernet / IP. Neste caso, o protocolo usado entre servidores e dispositivo de armazenamento é o iSCSI. Os comentários feitos no parágrafo anterior para SAN com comutador Fibre Channel são os mesmos para este caso.

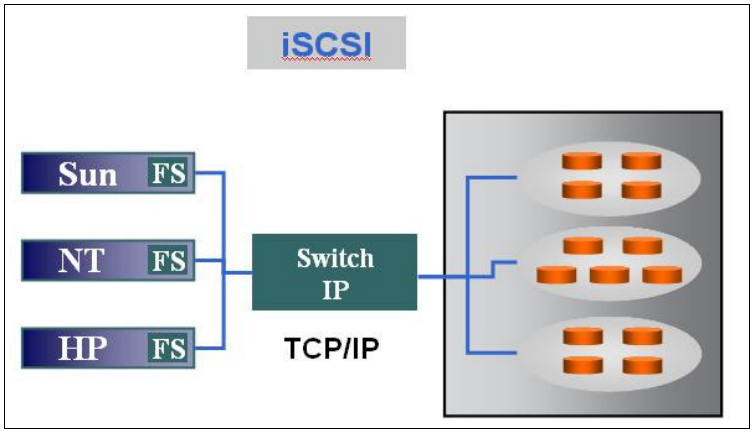


Figura 4.6 Modelo de uma SAN com comutador Gigabit Ethernet

Na Figura 4.7, pode-se ver a implementação do padrão FAS usando um comutador Gigabit Ethernet. Neste caso, os protocolos de compartilhamento de arquivos são SMB/CIFS e/ou NFS. Esta arquitetura é chamada de NAS. Nela, os tipos de dados sendo trocados entre o servidor e dispositivos de armazenamento de dados são baseados em "arquivos". Por esse motivo NAS é considerado um ambiente de entrada/saída de dados em baseados em arquivos.

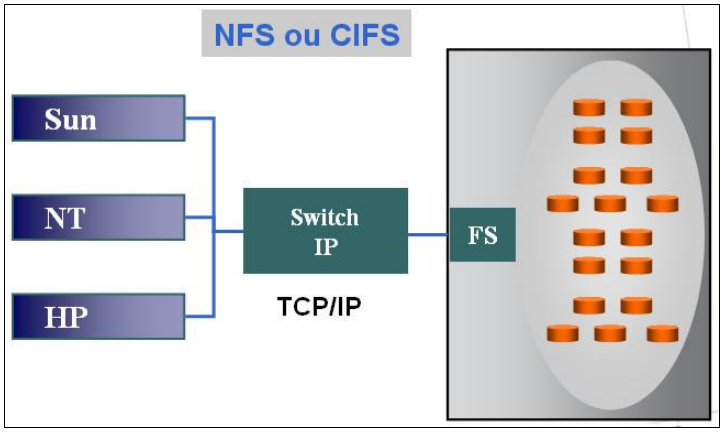


Figura 4.7 Modelo de uma NAS com comutador Gigabit Ethernet

Agora entraremos em mais detalhes nas arquiteturas SAN e NAS.

4.3. Arquitetura SAN

As redes SANs foram descritas pela primeira vez no final da década de 1990 pelo projeto StoreX da Sun Microsystems e pelo livro "*Enterprise Network Storage Architecture*" da Compaq.

Ambas empresas definiram SANs como redes de armazenamento de dados que permitiriam escalabilidade sem limites, dimensionamento de volumes, conectividade heterogênea (tanto para dispositivos de armazenamento de dados quanto para plataformas de servidor), gerenciamento centralizado, ou seja, uma solução inteligente para sanar às necessidades de armazenamento de dados das aplicações existentes [31].

Um evento de grande importância para o surgimento da arquitetura SAN foi o desenvolvimento da tecnologia *Fibre Channel*, que permitia uma transferência de dados em alta velocidade entre dispositivos. Com a popularização das SANs, os fornecedores começaram a definir suas próprias implementações de padrões FC, fato que dificultou o desenvolvimento de um padrão único.

Com o objetivo de desenvolver produtos padronizados, baseados em Fibre Channel, foram criadas associações de empresas e entidades para a definição das regras de utilização e desenvolvimento do novo padrão. Assim, surgiram a FCIA[25] e a SNIA como as principais associações que se destacam neste cenário.

A FCIA, criada em agosto de 1999, surgiu a partir da fusão da Fibre Channel Association e da Fibre Channel Community. O objetivo da FCIA foi criar as bases tecnológicas de infra-estrutura FC, de modo que várias aplicações pudessem ser suportadas pelos mercados de armazenamento de dados e TI.

A arquitetura SAN oferece comunicação entre os servidores e dispositivos de armazenamento através de uma rede dedicada, permitindo múltiplos acessos ao mesmo dispositivo de armazenamento. De acordo com Pollack [32], a grande vantagem das SANs deve-se às altas taxas de conexão (múltiplos *Gigabits* por segundo) entre os dispositivos de armazenamento, e também por permitir acesso simultâneo aos subsistemas de armazenamento através de um grande número de usuários.

[25] Fibre Channel Industry Association - organização sem fins lucrativos, formada por fabricantes, integradores de sistemas, desenvolvedores, vendedores, profissionais da indústria e usuários da tecnologia Fibre Channel.

Numa SAN é possível acrescentar dispositivos de armazenamento aos servidores com mínimo impacto e grande desempenho. As SANs oferecem um ambiente centralizado facilitando a operação e administração do armazenamento de dados.

Uma SAN pode ser dedicada a um servidor local ou remoto, ou compartilhada entre servidores. Existe uma série de possibilidades de interface de conexão de uma SAN, podendo ser:

- ESCON (*Enterprise Systems Connection*, Conexão de sistemas corporativos);
- SCSI (*Small Computer Systems Interface*, Interface de sistemas computacionais de pequeno porte);
- SSA (*Serial Storage Architecture*, Arquitetiura de armazenamento serial);
- HIPPI (*High Performance Parallel Interface*, Interface paralela de alto desempenho);
- FC (*Fibre Channel*);
- Qualquer outra nova forma de interface emergente.

Atualmente, a forma mais comumente encontrada é através da tecnologia Fibre Channel.

Para as SANs criaram-se novos métodos de conexão de armazenamento aos servidores. Elas são usadas para conectar dispositivos de armazenamento compartilhados a vários servidores isolados ou em ambiente de cluster. Elas podem conectar discos ou fitas a variados sistemas computacionais, como *mainframe*, estações de trabalho gráficas, servidores ou clientes da rede. Podendo ainda, criar caminhos redundantes entre dispositivos de armazenamento de dados e sistemas computacionais. A seguir, as possíveis conexões SAN entre servidores e/ou dispositivos de armazenamento de dados são apresentadas.

- **Servidor x armazenamento**: é o modelo tradicional de interação com os dispositivos de armazenamento e cuja vantagem é que um mesmo dispositivo de armazenamento pode ser acessado de forma concorrente por múltiplos servidores;
- **Servidor x servidor*:*** é o modelo no qual é possível obter alta velocidade de comunicação e volume de dados entre servidores; e

- **Armazenamento x armazenamento*:*** é o modelo que permite a transferência de dados entre dispositivos sem intervenção do servidor, liberando o processador para outras atividades. Por exemplo: backup direto de áreas dos discos para fita magnética conectada a SAN ou um espelhamento de dados entre um dispositivo de dados em outro dispositivo remoto através da SAN.

A seguir, estão listados alguns benefícios proporcionados pelo uso de uma SAN:

- Maior disponibilidade e alto desempenho;
- Armazenamento e gerenciamento centralizado e consolidado;
- Qualquer servidor pode acessar qualquer área de dados do dispositivo de armazenamento de dados;
- Os *backups* podem ser realizados sem necessidade de servidor para transportar os dados e sem competir com o uso da rede local de computadores, pois pode ser realizado diretamente do dispositivo de armazenamento para o dispositivo de backup como, por exemplo, uma unidade de fita magnética.

Hoje, as redes SAN podem ser implementadas de duas maneiras, dependendo do tipo de infra-estrutura de transporte de dados entre servidores e os dispositivos de armazenamento. As redes SANs baseadas em Fibre Channel ou SANs baseadas em TCP/IP.

4.3.1. Fibre Channel SANs

Uma rede Fibre Channel SAN utiliza infra-estrutura de rede baseada na tecnologia Fibre Channel. Foi amplamente adotada pelas grandes organizações por ter sido a primeira rede de alto desempenho comercial. O protocolo usado é o protocolo FCP, também conhecido como SCSI sobre Fibre Channel.

Apesar de estarem no mercado há algum tempo, os equipamentos para SAN baseados em Fibre Channel ainda não convergiram totalmente para um padrão comum e aberto entre os

diversos fornecedores. Existem ainda, questões de conectividade e recursos avançados que não são possíveis entre os equipamentos de diferentes fabricantes. Isso pode trazer um grande desafio, ou até inviabilizar, o gerenciamento de ambientes com diferentes fornecedores.

Embora o limite de transmissão do Fibre Channel seja de 10 quilômetros, na prática, a maioria das redes SAN usam conexões de fibra óptica multimodo, limitadas de 250 a 500 metros entre dispositivos. Isso torna a replicação e recuperação de dados entre localidades remotas extremamente difícil e custosa.

Outra característica associada aos equipamentos de uma SAN baseada em Fibre Channel é o custo da infra-estrutura do ambiente Fibre Channel. Para ter-se uma idéia, a preços de 2006, uma porta de comutador Fibre Channel custa cerca de 1.000 dólares e uma interface HBA Fibre Channel para conectar no servidor custa 1.000 dólares.

4.3.2. IP SANs

Um rede SAN baseada em IP pode usar qualquer infra-estrutura de rede que permita o trafego de protocolo TCP/IP. O custo acessível e a permeabilidade da tecnologia Ethernet oferece a oportunidade para uma nova solução de armazenamento de dados, a SAN baseada em IP, também chamada IP SAN. Atualmente, por questões de custos, essas redes tem sido implementadas com comutadores de tecnologia Gigabit Ethernet.

Uma SAN baseada em IP tem todos os benefícios que uma SAN baseada em Fibre Channel. Em termos de funcionalidade apresenta ainda outra vantagem, pois teoricamente uma rede baseada em IP é insensível à distância, podendo transportar dados a qualquer máquina conectada ao IP. Ao comparar com Internet, que também utiliza IP como protocolo padrão, uma IP SAN poderia ter, também, uma abrangência global.

O custo dos equipamentos para montar a infra-estrutura de SAN baseada em IP é bem menor que o correspondente a Fibre Channel, pois os equipamentos não servem exclusivamente para redes de armazenamento de dados, servem para qualquer rede de computadores existente.

Para se ter uma idéia de custo, a preço de Junho de 2006, uma placa de rede Gigabit Ethernet custa no mercado brasileiro por volta de 100 dólares, já uma porta de comutador Gigabit Ethernet sai, também, por cerca de 100 dólares.

Uma desvantagem da SAN IP em relação a SAN FC está no fato que a tecnologia Fibre Channel hoje pode funcionar a 1 Gigabit/s, 2 Gigabit/s e 4 Gigabit/s, enquanto Ethernet funciona a 1 Gigabit/s. Entretanto existem implementações de Ethernet com 10 Gigabit/s, porém seus custos ainda altos para utilização em dispositivos de armazenamento de dados.

Outra desvantagem de SAN IP em relação a SAN FC é quanto ao processamento de pacotes de dados na rede, que no caso de SAN FC é feita pela própria interface HBA Fibre Channel e no SAN IP, o processamento de pacotes TCP/IP e iSCSI é realizado pelo processador do servidor. Para resolução desse problema, vários fabricantes estão desenvolvendo interfaces chamadas TOE e iSCSI HBA [16] para processamento sobre TCP/IP e iSCSI na própria interface de comunicação.

4.4. Arquitetura NAS

A arquitetura de armazenamento de dados NAS baseia-se no protocolo de comunicação TCP/IP e em protocolos de compartilhamento de arquivos, sendo que os principais são: SMB/CIFS e NFS, protocolos padrões dos computadores com sistema operacional Microsoft Windows e Unix, respectivamente.

Um dispositivo de armazenamento de dados NAS é composto de processador integrado a um subsistema próprio de discos. O dispositivo pode ser ligado a uma rede local, ou remota, dedicada ao tráfego de dados de armazenamento, ou pode ser ligado a uma rede local, ou remota, compartilhada com a rede de computadores.

O tipo de dado que é usado nessa rede é o arquivo. As requisições de arquivos recebidas pelo NAS são traduzidas pelo processador interno em solicitações ao subsistema de discos do NAS. A rede local típica usada pelos dispositivos NAS é a rede Ethernet, mais precisamente

Gigabit Ethernet. Mesmo funcionando em Ethernet padrão quanto em Fast Ethernet, o desempenho dessas duas formas de Ethernet não atendem aos requisitos básicos para tráfego de dados. Existem soluções sendo implementadas em 10 Gigabit Ethernet, porém o custo alto ainda é alto para a maioria das organizações.

Na arquitetura NAS, o servidor de aplicação não tem controle, nem conhecimento, de como é a estrutura do subsistema de discos, volume, partição, cilindro, trilha ou setores de disco, como ocorre no caso de DAS e SAN. Dentro do equipamento de armazenamento do tipo NAS, o sistema operacional controla todo o sistema de arquivos. Um fato positivo dessa implementação é que o servidor de aplicação não é onerado desta atividade.

Um equipamento NAS geralmente suporta o armazenamento em disco e, em algumas vezes, dispositivos de fita. A fita conectada diretamente ao dispositivo NAS favorece a execução de *backup* dos discos de forma direta.

Algumas características dos equipamentos NAS são:

- Armazena e recupera dados na forma de arquivos através da rede IP.
- Pode ser centralizado ou distribuído, não possui limitação de distância.
- Consolida o armazenamento melhorando recuperação em caso de desastre e possibilitando a continuidade da atividade.

NAS é geralmente mais fácil de instalar e gerenciar do que SAN, pois é baseado em tecnologias já dominadas pelos departamentos de TI das organizações.

O NAS permite total uso ou compartilhamento da capacidade de armazenamento do dispositivo. Isto deve-se ao fato dos dados poderem ser configurados com um ou mais sistemas de arquivos e serem disponibizados para um ou mais clientes simultaneamente. Além disso, os dispositivos NAS atuais permitem o compartilhamento de arquivos de maneira nativa, e multiprotocolo, ou seja, clientes dos ambientes Microsoft Windows ou UNIX podem acessar de maneira imediata e direta um mesmo arquivo sem qualquer tipo de conversão ou duplicação.

Na medida que os usuários começam a experimentar dificuldades associadas ao gerenciamento de dados em um ambiente DAS, o ambiente NAS freqüentemente representa um próximo passo fácil na direção das redes de armazenamento de dados. A instalação e o gerenciamento são geralmente simples e rápidos.

Uma vantagem da arquitetura NAS é que em um ambiente com muitos servidores rodando sistemas operacionais diferentes, o armazenamento de dados pode ser centralizado com segurança, facilidades de gerenciamento e backup de dados.

Uma outra grande vantagem do NAS é que a maioria das funcionalidades desenvolvidas para os ambientes de redes de computadores, como melhorar desempenho, segurança e proteção contra falhas, estão automaticamente disponíveis aos ambientes NAS.

Alguns exemplos de desenvolvimentos que já ocorreram e são amplamente usados são:

- **IPSec**[26] – é mecanismo usado para criar um canal de comunicação que garanta a confidencialidade e a integridade das comunicações entre dispositivos que tem por base o protocolo TCP/IP. Esse recurso é usado nas redes de armazenamento de dados quando existe trafego de dados entre servidores e dispositivos de armazenamento de dados por *links* públicos ou inseguros.
- **Agregação de enlace** – também conhecido como *trunking*, é o mecanismo de unir uma ou mais portas Ethernet em uma única interface virtual. A porta virtual aparece como um único endereço IP com a banda igual à soma das bandas das portas individuais. A carga da conexão TCP é balanceada entre as portas. Além disso, provê redundância e tolerância a falha das portas individuais de forma transparente para as aplicações. Esse recurso é usado nas redes de armazenamento de dados quando de deseja ter portas de comunicação com maior banda ou quando se deseja implementar tolerância à falha das portas de comunicação do dispositivo de armazenamento de dados.

[26] IP Security – conjunto de protocolos desenvolvido pelo IETF para troca secura de pacotes em redes IP.

- **VLAN**[27] – ou rede local virtual é o mecanismo pelo qual redes logicamente independentes podem co-existir num mesmo comutador físico. Ela serve para reduzir o domínio de broadcast na camada MAC[28] e serve para restringir acesso de recursos de rede entre grupos separados de portas ou por endereço MAC. Duas aplicações desse mecanismo são: melhorar o desempenho do dispositivo de armazenamento ao reduzir o domínio de broadcast e restringir seletivamente acesso de clientes ao dispositivo de armazenamento.

No ambiente DAS, um servidor parado significa que os dados que esse servidor mantém não estão mais disponíveis. Já com a arquitetura NAS, os dados continuam disponíveis e podem ser acessados por outros servidores.

Geralmente o custo de um dispositivo NAS é maior do que um dispositivo DAS, pois possui capacidade de processamento local para gerenciar e compartilhar arquivos. Um dispositivo NAS possui ainda as seguintes vantagens:

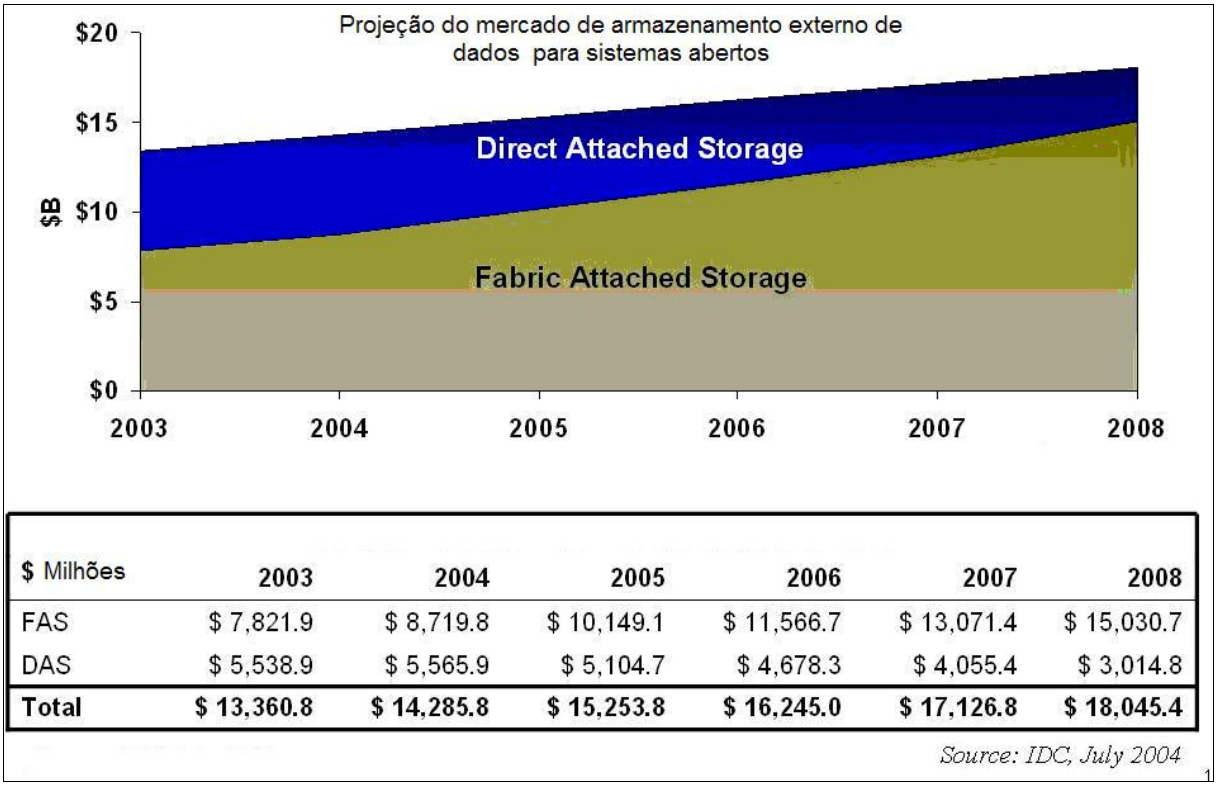
- Maior distância entre dispositivo e servidor, por estar ligado em rede;
- Permitir um grande número de usuários simultâneos acessando o mesmo dispositivo de armazenamento;
- Capacidade compartilhar toda a capacidade de armazenamento entre os servidores ligados ao dispositivo;
- Compartilhamento de arquivos entre servidores, mesmo com diferentes sistemas operacionais;
- Facilidade de gerenciamento;
- Alto desempenho, maior disponibilidade e escalabilidade;
- Suporte heterogêneo à plataforma;
- Infra-estrutura de conexão existente;
- Ferramentas de administração baseada em padrões Web.

[27] Virtual LAN
[28] Media Access Control

Os dispositivos NAS suportam protocolos para compartilhamento de arquivos que são padrões de mercado, tais como: NFS, CIFS, e algumas vezes podendo suportar HTTP e FTP.

4.5. Comparação de DAS, SAN e NAS

Em Julho de 2004, o IDC publicou os resultados de um estudo [2] dimensionando o tamanho dos mercados de armazenamento de dados. Na Figura 4.8, pode-se ver os resultados do estudo que prevê o tamanho do mercado de sistemas externos de armazenamento de dados de sistemas abertos (ou seja, excluindo *mainframes*) até o ano de 2008.



$ Milhões	2003	2004	2005	2006	2007	2008
FAS	$ 7,821.9	$ 8,719.8	$ 10,149.1	$ 11,566.7	$ 13,071.4	$ 15,030.7
DAS	$ 5,538.9	$ 5,565.9	$ 5,104.7	$ 4,678.3	$ 4,055.4	$ 3,014.8
Total	**$ 13,360.8**	**$ 14,285.8**	**$ 15,253.8**	**$ 16,245.0**	**$ 17,126.8**	**$ 18,045.4**

Source: IDC, July 2004

Figura 4.8 Previsão do mercado de discos externos para sistemas abertos. Fonte, estudo do IDC feito em Julho de 2004.

Os resultados mostram o declínio do armazenamento DAS ao longo dos anos. Isso tem ocorrido por vários motivos, pois existem duas áreas onde FAS oferece vantagem significativa em relação à DAS: utilização e gerenciamento.

A aquisição de sistemas de armazenamento DAS é feita de forma a atender as necessidades específicas de cada servidor e esse armazenamento não pode ser compartilhado facilmente com outros servidores mesmo numa situação de armazenamento consolidado. Na Figura 4.9, pode-se ver outro resultado do estudo feito pelo IDC, nele vemos que a utilização de espaço de armazenamento na solução DAS é de 40% a 50%, enquanto nas soluções FAS vai de 85% a 90% de utilização, isso acontece porque, como os recursos de armazenamento numa solução FAS podem ser mais facilmente dimensionados, alocados e compartilhados em vários servidores.

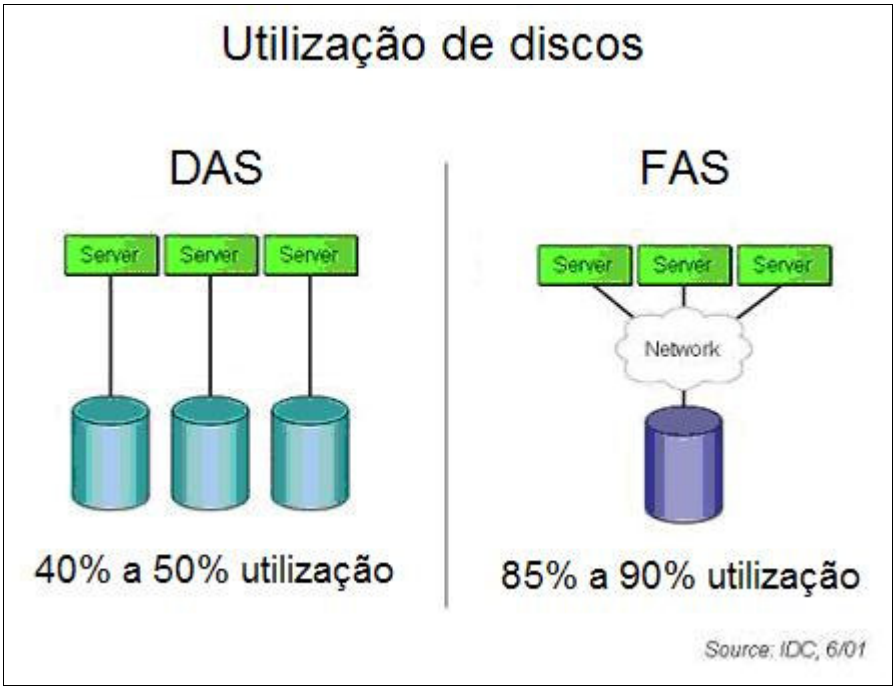


Figura 4.9 Comparativo da utilização de espaço em disco entre DAS e FAS. Fonte estudo feito pelo IDC em Junho de 2001.

O gerenciamento de armazenamento de dados em ambientes DAS é diretamente proporcional ao número de servidores e não ao volume de armazenamento como ocorre nas soluções FAS. Na Figura 4.10, pode-se ver os resultados do mesmo estudo do IDC, com informações referentes ao gerenciamento de dados. Pode-se ver que o número de pessoas técnicas para gerenciar o crescimento de armazenamento em DAS é bem maior que o de FAS.

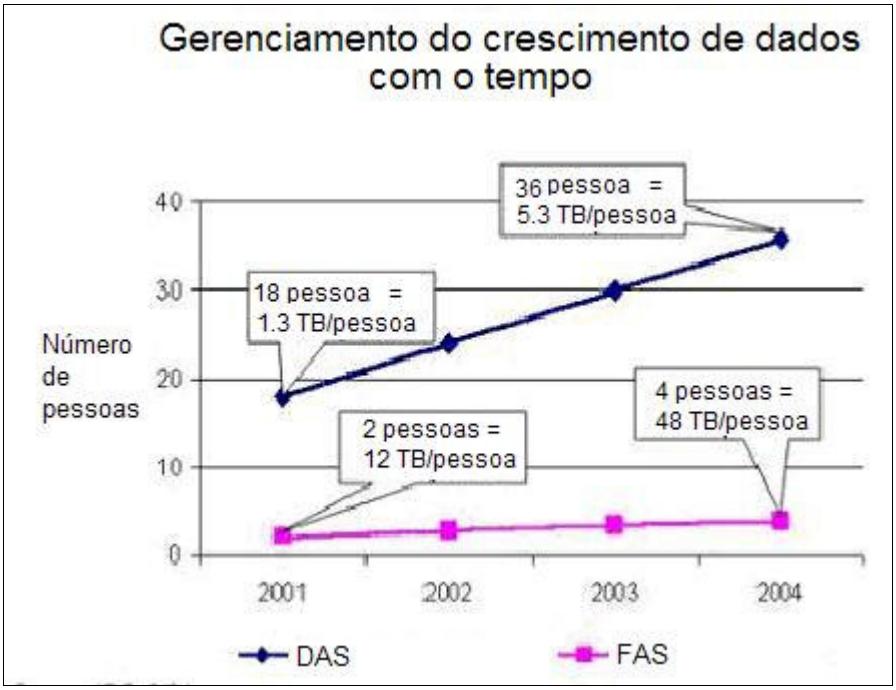


Figura 4.10 Gerenciamento de crescimento de dados em DAS e FAS. Fonte, estudo feito pelo IDC em Julho de 2001.

Para concluir a comparação entre DAS e FAS, a Figura 4.11 apresenta-se os mesmos resultados vistos na Figura 4.8, porém destaca-se os tamanhos dos mercados DAS e FAS para os anos de 2003 e previsão em 2008. Mantida esta previsão, é o claro declínio da arquitetura DAS comparada com a arquitetura SAN.

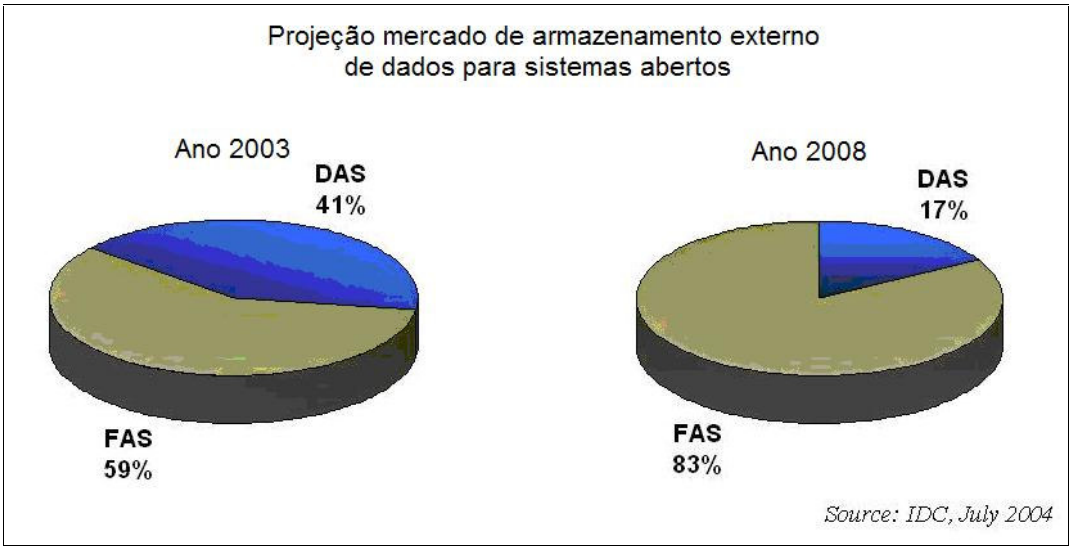
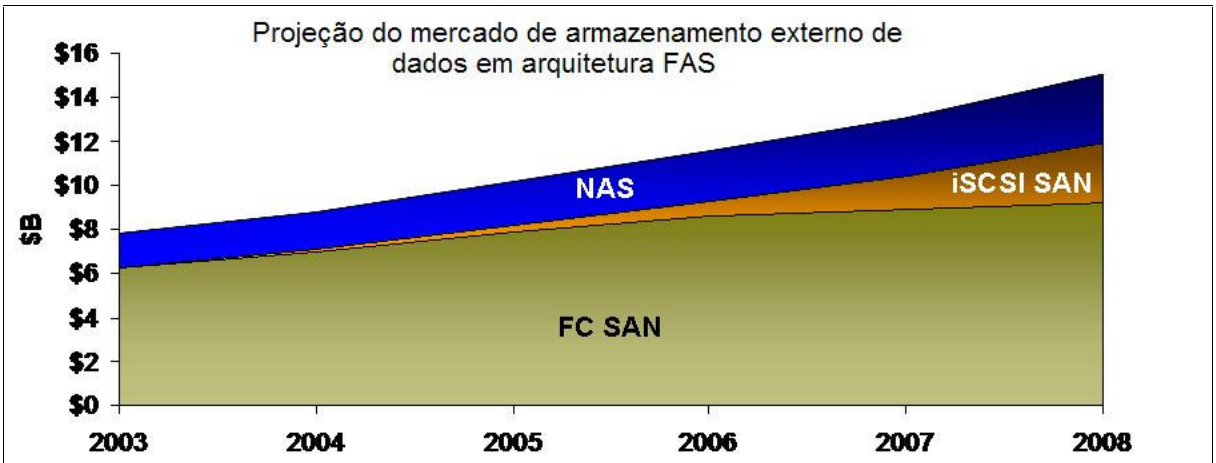


Figura 4.11 Mercado DAS x FAS de 2003 e 2008. Fonte estudo do IDC em Julho de 2004.

4.6. Comparação de SAN com NAS

Uma das principais diferenças entre a arquitetura SAN e NAS está no tipo de dado que trafega na rede. Na arquitetura SAN, os dados que trafegam são os blocos de dados, já na arquitetura NAS, os dados que trafegam são os arquivos.

Na Figura 4.12, pode-se ver a partir de outro estudo do IDC, este realizado em Julho de 2004, que a previsão de evolução do mercado de arquitetura FAS para sistemas abertos mostra uma estabilização do mercado de SAN em Fibre Channel e um grande crescimento de SAN baseada em iSCSI. Já o mercado para arquitetura NAS tem previsão de ser um mercado levemente crescente ao longo dos próximos anos.



Open Systems FAS Market Forecast						
$ Milhões	2003	2004	2005	2006	2007	2008
NAS	$ 1,558.8	$ 1,696.2	$ 1,957.5	$ 2,309.8	$ 2,684.1	$ 3,113.8
iSCSI SAN	$ 18.6	$ 84.8	$ 296.1	$ 682.5	$ 1,470.8	$ 2,690.9
FC SAN	$ 6,244.5	$ 6,983.9	$ 7,895.6	$ 8,574.4	$ 8,916.5	$ 9,226.0
Total	**$ 7,821.9**	**$ 8,719.8**	**$ 10,149.1**	**$ 11,566.7**	**$ 13,071.4**	**$ 15,030.7**

Source: IDC, July 2004

Figura 4.12 Previsão da evolução do mercado FAS para sistemas abertos. Fonte estudo do IDC em Julho de 2004.

Do ponto de vista das aplicações que estão sendo executadas em um servidor, o fato de ele fazer acesso a bloco ou acesso a arquivo pode fazer uma diferença significativa.

No caso de acesso a bloco, o servidor recebe um disco virtual do dispositivo de armazenamento e nesse disco ele executa todas as tarefas básicas que ele executaria em um disco real, assim, tanto para o sistema operacional do servidor quanto para a aplicação nada muda entre dispositivo de armazenamento via rede ou local ao servidor. Já quando se faz acesso a arquivos existe a dependência de saber se aplicação aceita trabalhar com dados em sistema de arquivos e se aceita, qual sistema de arquivos ele reconhece. Isso pode ser um grande obstáculo caso se tenha migrado de uma arquitetura DAS para uma arquitetura NAS.

Uma vantagem da arquitetura NAS é sua independência de plataforma ou sistema operacional. Por exemplo, uma vez que um sistema suporta o protocolo NFS, não importa se está rodando em um processador Intel com sistema operacional Linux, ou em um *mainframe* com suporte a TCP/IP e NFS.

Comparativos entre NAS e SAN.

- O desempenho do NAS em relação ao da SAN, é subjetivo. Isso depende de cada configuração em particular mas, em geral, a SAN é considerada mais rápida. Isso principalmente pelo fato da SAN usar uma rede dedicada com protocolo com pouca sobrecarga e baixa latência comparado com NAS.
- As velocidades da rede SAN hoje são 1 Gbps, 2 Gbps ou 4 Gbps em Fibre Channel. Para Ethernet, temos as seguintes velocidades de rede: 1 Gbps (Gigabit Ethernet) e 10 Gbps (10 Gigabit Ethernet).
- A manipulação do protocolo Fibre Channel é feita na própria placa controladora, chamada HBA, enquanto o tratamento do protocolo TCP/IP é feito pelo próprio processador do servidor, adicionando uma considerável carga adicional no processador do servidor. Existem placas controladoras que fazem tratamento dos protocolos (TCP/IP, iSCSI, etc,.) na própria placa, com isso diminuem carga de uso do processador no servidor.

Tabela 4.1 Resumo comparativo das características de SAN e NAS

	SAN	NAS
Protocolo usado na rede (Tipo de rede)	• FCP (Fibre Channel) • iSCSI (Ethernet /TCP-IP)	• CIFS (Ethernet/TCP-IP) • NFS (Ethernet/TCP-IP)
Tipo de informação que trafega na rede entre servidores e storage	Blocos de dados	Arquivos
Controle de distribuição dos dados nos discos do storage	A responsabilidade de formatar, particionar e distribuir dados nos discos lógicos criados no storage é do sistema operacional do servidor	A responsabilidade de formatar, particionar e distribuir informações nos discos do storage é do próprio storage, independe do sistema operacional do servidor
Latência da rede	• Para FCP a latência é muito baixa, em geral implementada numa rede dedicada. • Para iSCSI a latência depende da implementação da rede – padrão IP	Depende da implementação da rede – padrão IP
Segurança dos dados na rede	Alta, implementada numa rede não compartilhada	Depende como for implementada. Ideal ser uma rede separada.
Hardware no servidor	• Placa de comunicação HBA (Host Bus Adapter) • Placa de Comunicação iSCSI • Placa de Rede NIC (Network Interface Card)	Placa de rede NIC (Network Interface Card)
Hardware na rede	• Comutador Fibre Channel (para FCP) • Comutador Ethernet (para iSCSI)	Comutador Ethernet
Velocidades de transporte de dados no meio físico atualmente	• FCP – 1, 2 ou 4 Gbs • iSCSI – 1ou 10 Gbs	100 Mbs, 1 ou 10Gbs

Tabela 4.2 Resumo de vantagens, desvantagens e aplicações de SAN versus NAS

	SAN	NAS
Distâncias entre servidores e dispositivos de armazenamento de dados	• FCP – Limitado a dezenas de Kms • iSCSI – Praticamente ilimitado, pode ir até onde IP chegar, porém estará limitado em latência	Praticamente ilimitado, pode ir até onde IP chegar, porém estará limitado em latência
Desempenho	Alto desempenho, protocolos voltados a movimentação de blocos de dados	Médio desempenho, pois os protocolos são voltados a movimentação de arquivos, o que traz um trabalho adicional no tratamento dos dados
Escalabilidade das áreas de armazenamento do storage para os servidores	Crescimento das áreas de armazenamento deve ser coordenado com ações no sistema operacional do servidor. Uma vez aumentada a área não é possível sua diminuição	Tanto creascimento quanto diminuição das áreas de armazenamento podem ser realizados sem o ações no sistema operacional do servidor
Adequação das aplicações ao armazenamento de dados em rede	Como uma SAN é uma extensão da conexão DAS, e, em princípio, todas aplicações rodam em DAS, consequentemente todas aplicações devem rodar em SAN	Nem todas aplicações estão preparadas para rodar em NAS, por exemplo, existem aplicações que querem ter o controle da distribuição de dados nos discos, como é o caso de bancos de dados em dispositivos RAW DEVICE.
Aplicações típicas	• Sistemas transacionais ou de missão crítica • Apliacações de banco de dados • Backup centralizado no servidor ligado ao storage • Sistemas com grande volume de dados e alta performance (BI, DW, etc)	• Compartilhar arquivos em ambientes Unix (NFS) e Windows (CIFS) • Backup centralizado, porém compartilhado pelo storage • Sistemas onde exista uma movimentação média de volume ou transações de dados

5. Armazenamento de Dados e Computação Grid

"Grid tem o pontencial de resolver problemas reais de negócio, pois simplifica o acesso global aos serviços computacionais corporativos"

Shane Robinson, HP, 2003.

A Computação Grid representa uma arquitetura que permite que usuários e aplicações utilizem, de forma econômica, segura, eficiente e ampla, um conjunto de recursos computacionais e de dados geograficamente dispersos. Ela pretende compartilhar recursos de forma colaborativa dentro das organizações, bem como entre organizações. A computação grid, como se propõe a usar todos os recursos instalados, possibilita que as organizações tenham grandes ganhos de produtividade, acelerando os processos de computação intensiva através do assinalamento dinâmico de atividades, facilitando a execução de tarefas de forma mais efetiva.

Com a computação grid será possível realizar a convergência entre as seguintes áreas:

- Computação,
- Comunicação e
- Informação (dados).

Hoje, é comum perceber que dispositivos computacionais comunicam, os dispositivos de comunicação computam e todos os recursos de computação, comunicação e informação coexistem dentro de um quadro unificado orientado a serviços.

O início da computação grid ocorreu no ambiente acadêmico através do agregação de recursos que podiam ser compartilhados. Com isso, foi possível a resolução de problemas científicos que não poderiam ser resolvidos, em tempo hábil, de forma tradicional. Além disso, as idéias associadas a computação grid podem também ser usadas para a resolução de problemas comerciais nos ambientes computacionais das empresas.

Algumas pesquisas norte-americanas mostram que certos bancos e empresas automotivas já conseguiram economizar de 50% a 60% de seus custos em *hardware* ao adotarem tecnologias baseadas em computação grid [33]. Segundo pesquisa do IDC, o mercado mundial de computação grid deve chegar a 12 bilhões de dólares por volta de 2007 [34].

5.1 . Início da Computação Grid

Em 1969, o Dr. Leonard (Len) Kleinrock, professor da Universidade da California, Los Angeles, um pioneiro da Internet ao ter hospedado em seu servidor o primeiro nó da rede ARPANET (que evoluiria para a Internet atual) foi, também, um dos primeiros pesquisadores a escrever sobre acesso sob-demanda à computação, dados e serviços. Ele previu "Nós provavelmente veremos a disseminação de '*computer utilities*', que, como os serviços de energia elétrica e telefonia ('*electric and telephone utilities*'), servirão as casas das pessoas e escritórios pelo país todo". A idéia dele pode ser explicada numa analogia simples: hoje ao comprarmos um novo eletrodoméstico para casa, não nos preocupamos como será feito para que ele funcione, simplemente conectamos o aparelho à tomada elétrica de casa e usamos. Ao final do mês pagamos, para a companhia que nos fornece eletricidade, a conta pelo tempo que o equipamento foi utilizado.

Mas foi somente, em 1995, que os conceitos de computação grid começaram a se tornar realidade. Durante o congresso americado "*SuperComputing '95*", em San Diego, Califôrnia, EUA, foram apresentados os resultados do projeto chamado I-WAY[29] [35]. Esse projeto, utilizando diversos supercomputadores e avançados sistemas de visualização distribuídos em várias localidades diferentes, objetivava fazer com esses supercomputadores funcionassem como um único e poderoso computador apto a executar aplicações de realidade virtual distribuída. O projeto explorava os problemas relativos ao gerenciamento e escalonamento de recursos distribuídos. Durante o evento, foram executadas mais de 70 aplicações de 19 diferentes disciplinas científicas e de engenharia. A conexão das localidades foi feita através de uma rede ATM[30] de alta velocidade (155Mbps) que ligava 17 centros de supercomputadores dos EUA. A

[29] Information Wide Area Year – o nome complete do rojeto foi I-WAY: Wide Area Visual Supercomputing
[30] Asynchronous Transfer Mode – tecnologia de rede baseada na transferência de dados em células.

demonstração do projeto foi feita pelo professor Dr. Ian Foster da Universidade de Chicago e do Laboratório Nacional Argonne, EUA. O sucesso do projeto levou o governo americado, através de sua agência DARPA[31], a investir em projetos para a criação de ferramentas que permitissem compartilhar recursos computacionais distribuídos.

A primeira citação do termo Grid ligado à computação foi feita, em 1998, no livro "*The Grid: Blueprint for a New Computing Infrastructure*" [36], de Ian Foster e Carl Kesselman da Universidade Southern California. No livro os autores definem computação grid como sendo: "Computação Grid é uma infra-estrutura de *hardware* e *software* que provê acesso dependente, consistente, pervasivo, e barato a recursos computacionais de alto desempenho".

Já em 2001, Ian Foster e Carl Kesselman, e outro pesquisador do Laboratório Nacional Argonne, Steve Tuecke, escrevem o artigo : "*The Anatomy of the Grid: Enabling Scalable Virtual Organizations*" [38] onde refinam a definição de computação grid dizendo que ela é o "compartilhamento de recursos e resolução de problemas de forma coordenada em organizações virtuais multi-institucionais e dinâmicas. O tipo de compartilhamento com os quais estamos preocupados não é primariamente a troca de arquivos, mas sim acesso direto a computadores, *software*, dados, e outros recursos, como é o requisito para uma ampla estratégia de solução de problemas de distribuição de recursos emergindo na indústria, ciência e engenharia. Esse compartilhamento é, necessariamente, altamente controlado, com provedores de recurso e usuários definindo clara e exatamente o que é compartilhado, quem pode acessar o compartilhamento, e as condições sob as quais ele ocorre. O conjunto de indivíduos e/ou organizações definidos por essas regras de compartilhamento formam o que se chama de uma organização virtual."

Apesar das várias outras definições de Grid existentes, algumas características comuns são encontradas nas soluções de computação grid. Essas características incluem um conjunto compartilhado de recursos disponibilizados em rede que possuem um custo razoável, e que são modulares, flexíveis, balanceados, escaláveis, distribuídos, e devem seguir padrões amplamente

[31] Defense Advanced Research Projects Agency

adotados. Eles são virtualizados, provisionados e agregados de tal forma que os recursos individuais podem ser organizados em uma entidade cooperativa integrada.

Como resultado, os usuários finais podem ter acesso completo a um poder computacional heterogêneo, e os dados podem estar em múltiplos bancos de dados, em diferentes formatos ou distribuídos em localidades geograficamente dispersas. Os sistemas computacionais, na formação de grids, funcionam com um sistema virtual unificado, oferecendo as vantagens na administração centralizada, porém, sem os custos associados aos grandes sistemas monolíticos.

O termo computação grid ainda é visto, por alguns, como sinônimo de computação científica, pois muitas aplicações científicas demandam alto desempenho computacional. No mundo científico o modelamento matemático dos problemas oferece meio de abstrair e simular as situações que se tenta resolver. As modelagens e simulações tornam-se cada vez mais complexas, demandando cada vez mais recursos computacionais que exigiriam custosos supercomputadores, inviáveis para a resolução de um grande número de problemas computacionais atuais.

O compartilhamento de recursos computacionais entre entidades científicas e acadêmicas faz parte da própria cultura científica. Como exemplo, houve a criação da Web, ou WWW (*World Wide Web*) que nasceu para o compartilharmento de documentos no meio científico. As primeiras experiências em computação grid nasceram nos meios científicos e acadêmicos, foram projetos de pesquisa desenvolvidos de forma colaborativa. Esses projetos pretendiam resolver problemas que exigiam grande poder computacional. A idéia era usar computação grid para agregar o poder computacional existente nas universidades e centros de pesquisa para criar um sistema computacional muito mais poderoso.

Alguns importantes projetos de computação grid originados nas universidades e centros de pesquisas, entre vários outros, são:

Projeto Teragrid (www.teragrid.org)
Projeto Eurogrid (www.eurogrid.org)
Projeto DataGrid (www.eu-datagrid.org)
Projeto DOE Science Grid (www.doesciencegrid.org)

Projeto NASA Informations Power Grid (www.ipg.nasa.org)

O poder da computação grid também chegou ao ambientes das empresas. Os orçamentos para os ambientes de TI estão cada vez mais reduzidos, os recursos de equipamentos e humanos são escassos e caros. A maioria dos ambientes computacionais das empresas possuem poder computacional ocioso. Vários estudos [36] mostram que a maioria dos computadores pessoais e estações de trabalho dos ambientes acadêmicos e comercias usam somente por volta de 30% de seu poder computacional instalado.

Muitas estratégias de Grid das empresas começaram com ambientes de alta clusterização ou Computação de Alta Desempenho (HPC[32]). Esses ambientes tem características muito específicas, pois buscam resolver problemas que podem ser altamente paralelizados, e isso favorece bastante os ambientes de computação grid, já que, sob gerenciamento adequado, múltiplos processos podem ser executados por vários computadores simultâneamente e posteriormente consolidados .

O amplo uso da Internet e sua grande abrangência geográfica, somado ao potencial existente de milhares de computadores pessoais interligado em rede viabilizaram que surgissem projetos associados a computação grid. Um dos primeiros projetos que aproveitou essa modalidade de computação distribuída, chamada de "*Internet computing*" ou "computação filantrópica", foi o projeto SETI@home.

O projeto SETI@home foi concebido, em 1995, por David Gedye da Universidade da California, Berkeley, para o Instituto SETI[33]. O instituto, criado em 1984, pretende buscar vida inteligente fora da Terra através da análise de sinais de rádio captados do espaço por meio de rádio-telescópios. O projeto SETI@home, que começou a operar em Maio de 1999, pretende usar o tempo ocioso de uma grande quantidade de computadores pessoais voluntários ligados à Internet para analisar os sinais recebidos e distribuídos pela central do Instituto SETI. Para se ter uma idéia, o resultado dessa computação filantrópica, em 2004 haviam 5 milhões de usuários/voluntários cadastrados, tendo contribuído coletivamente com mais de 19 bilhões de

[32] High Performance Computing
[33] Search for ExtraTerrestrial Intelligence - www.seti.org

horas de processamento. Se todo o trabalho efetuado pelo projeto fosse executado por um único computador pessoal com microprocessador Pentium seriam necessários mais de 1.300.000 anos para executá-lo.

Existiram e existem ainda vários projetos baseados em computação filatrôpica, muitos bem sucedidas e outros nem tanto. Podemos destacar alguns desses projetos: criação de um supercomputador virtual para jogar xadrez (www.chessbrain.net), previsão do tempo (www.climateprediction.net), pesquisa novas drogas para o combate à AIDS (www.fightaidsathome.org), análise de dados para o projeto Genoma (www.folderol.org), pesquisas relacionadas a AIDS, câncer e malária (www.find-a-drug.org), projeto matemático na busca de números primos especiais (www.mersenne.org), entre outros.

5.2. Componentes da Arquitetura Grid

Um conceito operacional muito importante em computação grid é o de organização virtual (*virtual organization*). Ela é definida como um conjunto dinâmico de indivíduos e/ou instituições regidos por um conjunto de regras de compartilhamento de recursos computacionais, *software*, ferramentas e protocolos. Membros de uma organização virtual, baseados nas regras de compartilhamento de recursos definidas, podem ter acesso a recursos, podem executar aplicações de forma controlada e segura [36]. Um dos grandes desafios técnicos da computação grid é o assinalamento de usuários, recursos, e organizações de diferentes domínios em territórios geograficamente distintos.

Para uma organização virtual, os recursos computacionais podem estar distribuídos globalmente e podem ser heterogêneos (*mainframes*, supercomputadores, estações de trabalho, servidores, computadores pessoais, clusters e ambientes multiprocessados). Os componentes podem ser abstraído por protocolos que controlam a alocação de recursos, conectividade, gerenciamento e os aspectos de segurança relacionados aos recursos.

Um novo modelo de arquitetuta foi criado pelos pesquisadores Ian Foster e Carl Kesselman para a definição e gerenciamento de recursos dentro da organizacional virtual [37].

Essa nova arquitetura, chamada “arquitetura grid”, identifica os componentes básicos de um sistema grid, define o propósito e funções de tais componentes e indica como esses componentes interagem entre si.

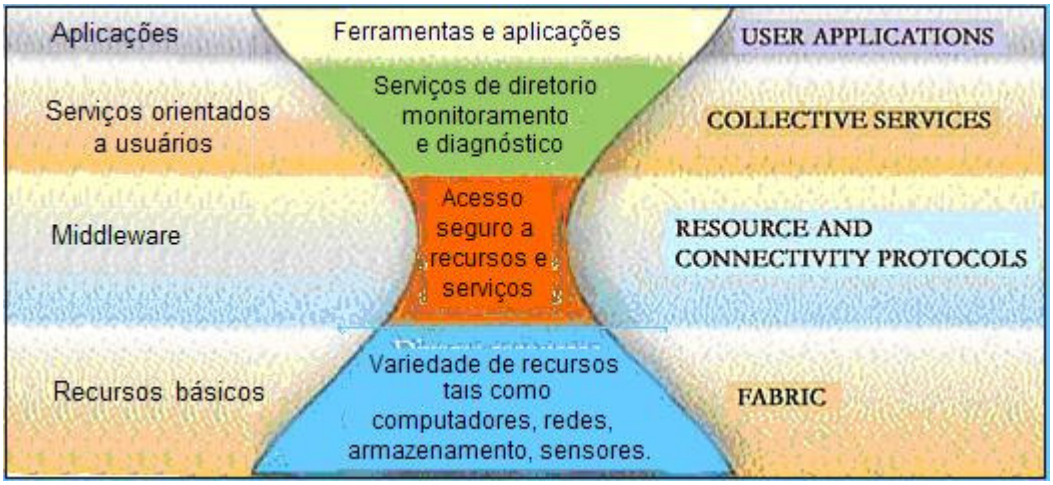


Figura 5.1 O modelo arquitetura grid em camadas, adaptado de [37]

Na Figura 5.1, pode-se ver como a arquitetura grid pode ser descrita na forma de camadas, numa estrutura aberta e extensível. Pode-se ver como os pesquisadores descrevem os componentes de cada uma camadas. Eles usam o modelo de ampulheta (*Hourglass Model*) [40] para especificar as várias camadas e descrever proporcionalmente a quantidade de componentes a serem usados em cada uma das camadas.

A parte estreita da ampulheta, associada a camada *middleware*, que é a camada de protocolos de recurso e de conectividade, define um pequeno conjunto de abstrações básicas e protocolos (por exemplo, TCP e HTTP), com os quais podemos mapear várias funcões de alto nível, parte superior da ampulheta, que é a camada de serviços orientados aos usuários (serviços coletivos). A mesma camada “*middleware*” pode ser mapeada sobre várias tecnologias básicas, que á a base da ampulheta, que é camada de recursos básicos (*Fabric*).

Pode-se observar que, por definição, o número de protocolos definidos na parte estreita da ampulheta deve ser pequeno, consistindo de protocolos de recursos e protocolos de conectividade que facilitem o compartilhamento de uma grande quantidade e variedade de recursos individuais. Esses mesmos protocolos podem ser usados para construir uma grande

variedade de serviços globais que envolvem o uso coordenado de múltiplos recursos. Na Tabela 5.1, apresenta-se um resumo das funcionalidades de cada camada da arquitetura grid.

Tabela 5.1 Descrição das camadas da arquitetura grid

Camada	Descrição
Aplicações dos usuários	Portais de acesso e programas que exploram a potencialidade das grids
Serviços orientados aos usuários	Ambientes de programação Grid
Middleware	Serviços básicos, como gestão de recursos distribuídos
Elementos básicos	Recursos computacionais, de armazenamento de dados, de redes, sensores, etc

Neste trabalho, o ponto focal em relação aos ambientes de computação grid está localizado na camada de elementos básicos, também chamada *Fabric*, que aqui não tem o mesmo significado do apresentado no Capítulo 2. Nesta camada, encontram-se os elementos que fornecem os recursos computacionais básicos para todas as camadas superiores da arquitetura grid. Um recurso pode ser uma entidade lógica, como um sistema de arquivos distribuído, um cluster de computadores, um sistema de armazenamento de dados, etc. Para ambientes grid, a experiência sugere que um recurso deve implementar mecanismos de introspecção, ou seja, que permita que se descubra sua estrutura, estado, e capacidade. Deve, também, implementar mecanismos de gerenciamento do recurso, permitindo algum controle na qualidade de serviço que ele pode oferecer.

5.3. Armazenamento de Dados em Grid

Um dos conceitos importantes que existe em ambientes grid é o de "Grid de Dados" (*Data Grid*). Os recursos de armazenamento de dados encontra-se na camada chamada "recursos básicos" do modelo de arquitetura grid. Um ambiente de computação grid requer acesso seguro e transparente a um conjunto de dados dinâmino e escalável. Requer, também, um sistema virtualizado global que possa apresentar um único "espaço de armazenamento" de dados para as aplicações e usuários, independente de sua localização ou de quais sistemas, tecnologias e equipamentos sejam usados para o armazenamento.

Os dados em grid são organizados como coleções estruturadas e compartilhadas de dados. Nessas coleções, os dados podem estar na forma de arquivos texto, arquivos binários estruturados, documentos XML, informações mantidas em banco de dados, etc. A manipulação, processamento e uso dessas coleções de dados distribuídos em larga escala requerem uma infra-estrutura compartilhada de dados, de recursos computacionais, e recursos de redes que devem ser usados de forma integrada, segura e flexível.

Os requisitos básicos de armazenamento de dados nos ambientes de computação grid são: flexibilidade, escalabilidade, transparência, desempenho, segurança e baixo custo. Com os esquemas tradicionais de armazenamento e administração de dados é impraticável atender os requisitos básicos ou manter um fluxo constante de dados para as aplicações, principalmente quando as coleções de dados são remotas e crescem com o tempo.

5.3.1 Transferência de Dados em Ambientes Grid

A primeira forma de compartilhar dados em um ambiente grid com equipamentos heterogêneos foi através da transferência de arquivos entre diferentes sistemas utilizando o protocolo FTP. Com isso foi possível formatar os dados para cada computador/sistema operacional a partir da aplicação de transferência de arquivos. Como exemplo, pode-se citar a transferência de um arquivo de um *mainframe,* que usa o fomato de caracteres EBCDIC[34], para um sistema aberto UNIX. A conversão de caracteres para formato ASCII[35], padrão para os sistemas abertos, pode ser feita automaticamente pelo FTP.

Contudo, o protocolo FTP era muito limitado para as necessidades dos ambientes grid. Assim, o primeiro esforço para a implementação de serviços de acesso e transporte de dados para ambientes de computação grid foi o desenvolvimento do protocolo GridFTP [41]. Ele é baseado no protocolo FTP, mas oferce um protocolo de transferência de dados de forma mais confiável, segura e de alto desempenho, sendo otimizado para redes amplas de alta velocidade. O protocolo

[34] *Extended Binary Coded Decimal Interchange Code* – esquema de codificação de caracteres usado nos sistemas operacionais de mainframe.
[35] *American Standard Code for Information Interchange* – esquema de codificação de caracteres usados na maiotia dos sistemas computacionais exceto mainframes.

FTP foi escolhido pela sua grande aceitação e uso na Internet, estando disponível na quase totalidade dos sistemas operacionais conhecidos para computadores pessoais, servidores, estações de trabalho e *mainframes*. Além das funcionalidades associadas ao protocolo FTP, o GridFTP pode ser usado como ferramenta de controle, monitoramento e transferência de dados entre dois outros sistemas. Ele permite a inclusão de código escrito pelo cliente, que pode ser usado para processar os dados antes da sua transmissão ou armazenamento. Para facilitar a interoperabilidade com outros serviços Grid, o GridFTP usa um mecanismo robusto e flexível de autenticação, integridade e confidencialidade baseado em GSI[36].

5.3.2 Acesso a Dados Remotos em Ambientes Grid

Uma das primeiras tentativas que surgiram com o propósito de permitir o acesso remoto a dados foi através do uso do protocolo NFS, padrão nos ambientes baseados em UNIX. Contudo esse protocolo não teve muita abrangência, porque possuía várias limitações. Entre as principais limitações cita-se: falta de escalabilidade, ser apropriado somente para redes locais (LAN) e não redes amplas (WAN), e possuir mecanismos de segurança considerados limitados.

Algumas limitações iniciais do protocolo NFS foram minimizadas com o desenvolvimento do protocolo AFS (*Andrew File System*). O AFS é sistema de arquivos distribuído que foi baseado em NFS, porém com grande preocupação de funcionamento em redes amplas e com forte apelo a segurança. Ele usa Kerberos[37] como protocolo de autenticação. Ele usa um sistema de manipulação de chaves e de criptografia para comprovar a identidade dos usuários. Contudo, isso não foi suficiente para que fosse um protocolo amplamente utilizado nos ambientes de computação grid, pois, entre outras coisas, se fosse adotado obrigaria a todos os sites na organização virtual adotassem esse mesmo mecanismo de segurança.

Uma das possibilidades sendo desenvolvidas para acesso remoto a dados vem do projeto GridNFS da Universidade de Chicago. No GridNFS é uma solução *middleware* que extende a

[36] *Grid Security Infrastructure* – é camada de segurança do Globus Toolkit 3. Provê serviços de autenticação e integridade de dados.

[37] Kerberos é um protocolo de autenticação em redes computadores desenvolvido pelo MIT (Massachusetts Institute of Tecnology

tecnologia de sistemas de arquivos distribuídos para atender às necessidades das organizações virtuais baseadas em grid. A base do GridNFS para o compartilhamento de arquivos é o protocolo NFS versão 4, que é o atual padrão do IETF para sistemas de arquivos distribuídos. Dentre os avanços implementados com NFS versão 4, cita-se: melhoria no acesso e desempenho na Internet, mecanismo de segurança forte com negociação embutida no próprio protocolo, melhor interoperabilidade entre diferentes plataformas, mecanismos de controle de acesso através de ACLs, etc.

5.3.3 Arquiteturas de Armazenamento de Dados em Ambientes Grid

Nos capítulos anteriores desse trabalho, foram descritas as arquiteturas DAS, NAS e SAN, que são as tecnologias típicas de armazenamento de dados dos ambientes tradicionais. Porém, uma simples adaptação dessas arquiteturas para trabalhar em grid não vai ser suficiente para atender os requisitos básicos necessários ao ambientes em grid. Vários desenvolvimentos devem ser acrescentados a essas arquiteturas para que se alcance um estágio onde se possam chamar de "sistemas avançados de armazenamento de dados", ou seja, sistemas de armazenamento totalmente integrados aos preceitos de computação grid.

A seguir, apresenta-se alguns dos desenvolvimentos que os dispositivos de armazenamento de dados para os ambientes grid devem implementar para se tornarem sistemas avançados de armazenamento de dados. Alguns desenvolvimentos, com algum grau de evolução, já fazem parte de produtos comerciais atuais.

- **Virtualização** - A separação e desligamento entre características físicas e lógicas é um requisito muito importante para implementação de grid. Assim, do mesmo modo que, deve existir virtualização de servidores e redes, a virtualização do armazenamento de dados é necessária para implementarmos a abstração que grid precisa na manipulação e gerenciamento de dados de forma trasparente.
- **Baixo custo –** Por causa da virtualização, as grids podem ser feitas de elementos escolhidos pelo baixo custo por capacidade. Em outras palavras, elas podem ser feitas de grandes volumes de elementos que são "comodites". Os elementos podem ter até baixa

disponibilidade quando comparados com elementos de maior custo e maior MTBF, mas, como demanda grid, eles podem ser usados se forem implementadas funcionalidade básicas no dispositivo de armazenamento, como continuar a funcionar, mesmo com a falha de algum de seus componentes e puder ser suportado por um contrato de manutenção mais barato, que incluam um tempo de resposta um pouco maior, mas sem impacto no desempenho geral do ambiente grid.

- **Escalável –** A arquitetura modular das grids significa que o investimento inicial pode ser modesto e com um investimento incremental mantendo baixo. Essa possibilidade de "*scale-out*", que implica em agregar mais dispositivos e fazê-los funcionar de maneira integrada. Isso é uma grande vantagem para os ambientes onde a capacidade de dados necessita frequentemente ser aumentada, diminuida, ou rapidamente modificada. Isso se opõe ao "*scale-up*' que seria trocar o equipamento por uma equipamento de maior capacidade ou mais poderoso quanto existisse a necessidade de maior capacidade ou desempenho.
- **Alta disponibilidade** – Os ambiente em grids devem ser auto-corretivos, com as cargas e dados sendo distibuídos entre vários elementos de armazenamento, de modo que uma falha em um elemento não produza impacto no desempenho da aplicação sendo executada. Isso significa que arquiteturas grid bem planejadas não devem ter tempo de parada em manuetnções planejadas ou mesmo em atualizações de elementos ou dispositivos. Manutenção preventiva e paradas programadas são coisas do passado.

Ao se analisar as arquiteturas de armazenamento de dados tradicionais pode-se perceber que somente as arquiteturas baseadas em redes, ou seja, as redes de armazenamento de dados (SAN e NAS) tem potencial para evoluirem e serem usadas na implementação dos "sistemas avançados de armazenamento de dados", pois elas se baseiam em infra-estruturas que podem ser decentralizados e dispersas geograficamente, isso pela própria natureza como são construídas. Consequentemente, a utilização da arquitetura de armazenamento DAS para os ambientes de computação grid não é praticamente usada, pois é limitada ao estar intimamente integrada e dependente de cada sistema computacional.

Alguns produtos de armazenamento de dados em grid baseiam-se em dispositivos de armazenamento de dados NAS. O dispositivo NAS oferece um primeiro nível de transparência de dados ao abstrair (ou virtualizar) para o sistema operacional e aplicações dos clientes onde as informações são armazenadas nos discos do dispositivo. Outro nível de abstração pode ser visto hoje, quando vários fornecedores de sistemas de armazenamento de dados NAS permitem que uma mesma informação armazenada no dispositivo possa ser manipuladas diretamente, sem nenhum tipo de conversão de dados, por clientes e aplicações usandos os diferentes protocolos do ambiente NAS, principalmente os protocolos NFS e CIFS. Com isso, quando se trata de serviços de arquivos, atendem simultaneamente ao universo de clientes UNIX e Windows.

Várias empresas já oferecem soluções embutidas nos dispositivos de armazenamento de dados, ou através de *software* rodando nos clientes, para virtualizar a localização e quantidade de dispositivos através de mecanismo de "*Global Name Space*". Com esse mecanismo pode-se virtualizar a localização de um dispositivo físico através de um nome, como se faz hoje com o mapeamento de nome de domínio em endereço IP na Internet através de DNS[38].

As arquiteturas SAN tem sido amplamente implementadas nos datacenters das empresas para centralizar o armazenamento de dados, visando facilitar o gerenciamento e proteção de dados para um grupo de servidores. Na fase inicial da consolidação das SAN tarefas como provisionamento de área de dados para novas aplicações, balanceamento de cargas nos discos para as aplicações, criação de novos volumes, configuração de grupos RAID, e assim por diante, consumiam tempo e eram tarefas manuais. Sem contar o alto custo da conectividade Fibre Channel, incluído comutadores e placas controladoras, e o alto custo do treinamemento de pessoal para administrar a SAN.

Com a evolução dos sistemas de provisionamento e gerenciamento de dados em SAN e com o advento de SAN implementadas com protocolo iSCSI, as empresas tem a oportunidade de introduzir armazenamento de dados uma conectividade que pode ser global, que é mais simples, mais barata, escalável, fácil de usar, baseada em infra-estruturas amplamente conhecidas, como Ethernet e TCP/IP, bases do protocolo iSCSI.

[38] Domain Name System – usado pela Internet para mapeamento de nomes de equipamentos/domínios em endereços IP.

5.4. Evolução das Tecnologias Grid

As tecnologias Grid se desenvolveram, desde o início da década de 1990, através de pesquisas e desenvolvimentos inicialmente acadêmico. Posteriormente, essas tecnologias se mostraram úteis também para os ambientes das empresas comercias. Segundo Ian Foster e Carl Kesselman [37], os principais momentos dessa evolução das tecnologias grid podem ser vistos na Figura 5.2.

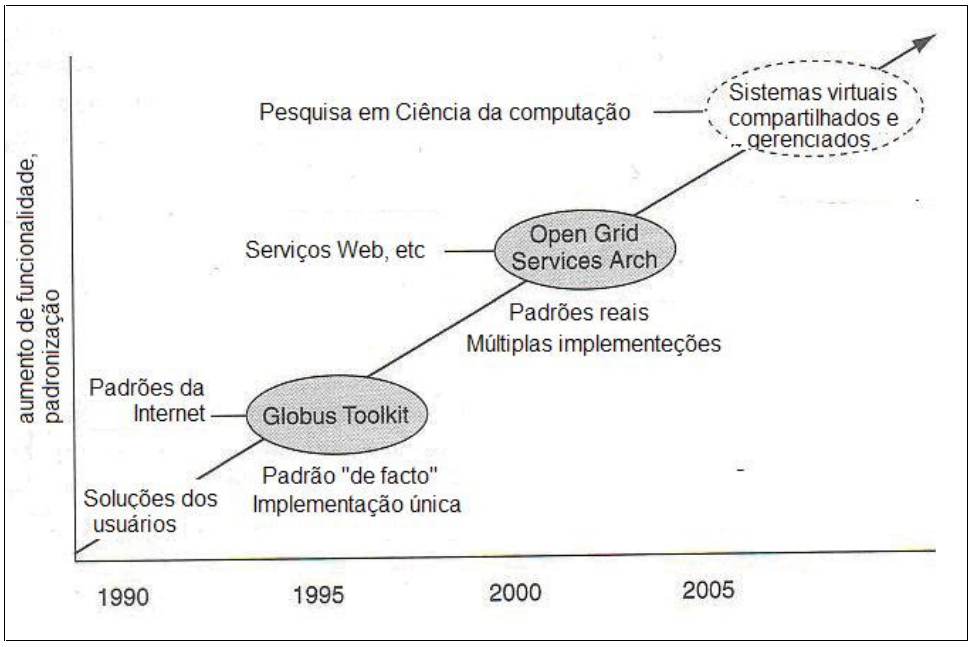


Figura 5.2 A evolução das tecnologias Grid, adaptado de [37].

A seguir, apresenta-se brevemente cada um dos momentos da evolução das tecnológicas Grid.

- **Soluções dos usuários**

  Começando no início da década de 1990, trabalhos em "metacomputacação" e assuntos correlatos levaram ao desenvolvimento de soluções customizadas para a solução de problemas de computação grid. O foco desses frequentemente esforços heróicos foram em fazer as coisas funcionassem e explorar o que era possível. As aplicações eram contruídas usando diretamente os protocolos da Internet com tipicamente somente funcionalidade

limitada em termos de segurança, escalabilidade, e robustez. Interoperabilidade não era uma preocupação significativa.

- **Globus Toolkit**

  Globus Toolkit é um conjunto de ferramentas de software aberto usadas para a construção de sistemas e aplicações baseadas em grid [39]. Ele tem sido desenvolvido pela entidade Globus Alliance e muitas outras entidades e indivíduos ao redor do mundo. O conjunto de ferramentas inclui software para segurança, infra-estrutura, gerenciamento de recursos, gerenciamento de dados, comunicação, deteção de falhas, e portabilidade. Seus serviços básicos, interfaces e protocolos permitem que os usuários acessem recursos remotos como se fossem locais. O Globus Toolkit versão 1 foi lançado em 1998, mas foi com a versão 2, lançada em 2002, que o Globus Toolkit se tornou uma solução de grande impacto na comunidade da computação de alto desempenho, pois focava em usabilidade e interoperabilidade. Ele definiu e implementou protocolos, interfaces com programas, e serviços usados em milhares de desenvolvimentos em grid ao redor do mundo, sendo que a arquitetura por ele definida tornou-se um "padrão *de facto*" como infra-estrutura para computação grid. O Globus Toolkit versão 3 foi lançado em 2003. A versão mais atual é o Globus Toolkit versão 4 que foi lançado em 2005.

- **Arquitetura OGSA (Open Grid Services Architecture)**

  Em 2002, um dos produtos gerados pelo entidade GGF (*Global Grid Forum*), que congrega mundialmente organizações voltadas para o desenvolvimento de computação grid, foi a padronização de uma arquitetura de protocolos abertos, baseados em serviços web (*Web services*), chamada OGSA. Ela é chamada de arquitetura, pois é responsável pela descrição e construção de um conjunto específico de interfaces a partir das quais é possível construir sistemas grid.

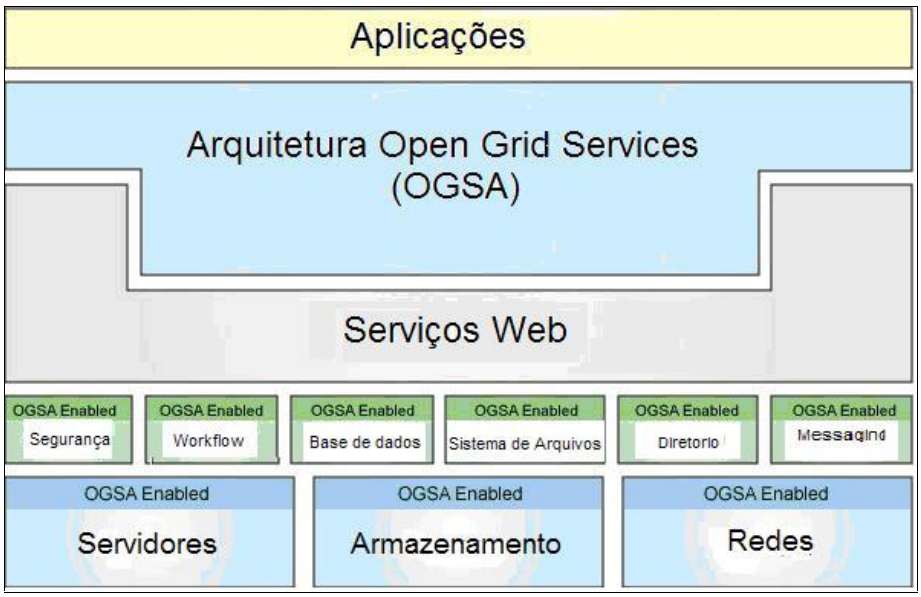


Figura 5.3 Estrutura da arquitetura OGSA

Na Figura 5.3, ilustra-se a estrutura da arquitetura OGSA, que é um verdadeiro padrão comunitário com múltiplas implementações, incluindo, em particular, o Globus Toolkit versão 3, disponibilizado em 2003, extendendo de maneira significativa os conceitos e tecnologias apresentadas na Globus Toolkit versão 2. A arquitetura OGSA alinha firmemente computação grid com iniciativas amplas da indústria em arquiteturas voltadas a serviços e serviços web. Oferece ainda uma forma que permite a definição ampla de serviços portáveis e interoperáveis.

- **Sistemas Virtuais Compartilhados e Gerenciados**

  Os futuros sistemas em grid ainda fazem parte das pesquisas atuais da Ciência da Computação. Tudo indica que deverão ser baseados na arquitetura orientada à serviço OGSA. Veremos uma expansão dos serviços de interoperabilidade, os sistemas escalando em número de entidades, altos níveis de virtualização, formas mais ricas de compartilhamento de recursos e um aumento na quantidades de serviços através de várias formas de gerenciamento ativo.

5.5. Computação Grid nas Empresas

A computação grid nos ambientes empresariais é motivada pela grande necessidade de poder computacional e pela manipulação de grandes quantidades de dados. As organizações inovadoras tem se beneficiado pela reposição das tradicionais soluções computacionais por soluções baseadas no conceito de grid. A razão para isso fica clara quando se percebe que o modelo de computação grid é uma estratégia para a utilização atual e futura de recursos, pois se baseia na idéia que se pode agregar recursos de *hardware* e *software*, locais ou remotos, possibilitando uma melhor utilização de recursos computacionais e retorno de investimento (ROI[39]) mais rápido.

É natural que no ambiente empresarial não se cogite usar computadores pessoais para tarefas críticas ao negócio, assim os ambientes de computação grid empresariais deverão ser constuídas a partir de servidores, pois esses são mais estáveis que os computadores pessoais. Um ambiente de negócio depende fortemente da confiabilidade e disponibilidade dos sistemas computacionais. No entanto, os inúmeros computadores pessoais existentes nas empresas representam uma grande fonte de recursos computacionais quase sempre ociosos. Eles podem ser uma grande fonte de poder computacional para as empresas que adotem ambientes baseados em computação grid.

A seguir, apresenta-se alguns motivadores para que as empresas adotem os conceitos de computação grid:

Preço / Desempenho

- Maximizar os recursos existentes ao usar os ciclos de CPU[40] disponíveis. As empresas gastam muito dinheiro anualmente em servidores e, mesmo assim a utilização não passa de 30% da capacidade de processamento disponível.
- Nas configurações tradicionais de TI[41], os recursos computacionais geralmente são dedicados a organizações e aplicações específicas, ficando ociosos quando essa

[39] Return Of Investment .
[40] Central Processing Unit
[41] Tecnologia da Informação

aplicações não estão em uso. A computação grid fornece um mecanismo para compartilhar recursos por toda a empresa.

- Sistemas baseados em grid são um caminho para o "*computer de utility*" no qual os recursos computacionais são disponibilizados sob demanda em oposição à constante criação ou ajuste de arquiteturas específicas das aplicações.
- As arquiteturas de computação grid são construídas através de servidores clusterizados de baixo custo ao invés de grandes servidores monolíticos. Isto é possível, em parte, através do uso de dispositivos de armazenamento de dados consolidados que possuem seu próprio sistema operacional e executam várias tarefas que normalmente seriam processadas pelos servidores. Essas arquiteturas têm um custo efetivo melhor que infra-estruturas tradicionais de TI.

Compartilhamento de Custos

- Múltiplos grupos podem contribuir e se beneficiar de recursos para um projeto.
- Já existe um conjunto de padrões para computação grid, que estão em contínua evolução, abrangendo um grande conjunto de sistemas e arquiteturas diferentes.

Melhoria no Gerenciamento de Recursos

- As arquiteturas de TI das empresas tradicionais são muito ineficientes em termos do nível de manutenção que elas necessitam. Muitas empresas chegam a utilizam até 90% de seus investimentos em TI para manter seus sistemas funcionando.
- A computação grid permite que as empresas promovam, de modo transparente, a modificação de seus processos computacionais sem ter que mudar suas infra-estruturas de TI. Um bom nível de crescimento e distribuição de recursos pode ser feito automaticamente pela própria infra-estrutura, permitindo que os gerentes de TI se concentrem em assuntos mas importantes ao invés de tarefas de manutenção e reconfiguração do dia-a-dia.
- Uma infra-estrutura grid pode, efetivamente, gerenciar aplicações com requisitos de capacidade variáveis sem a necessidade de constantes interversões e reconfigurações manuais.

- Os ambientes grid podem começar pequenos e crescer na medida que a demanda aumenta, sem a necessidade de refazer a arquitetura e sem interromper as aplicações que já estavam sendo executadas.
- Num ambiente de computação grid, o administrador pode gerenciar de modo fácil centenas de servidores e de petabytes de armazenamento de dados.

Nova Classe de Capacidades

- A computação grid tem o potencial de resolver, de modo mais rápido, alguns problemas muito grandes ou complexos, principalmente aqueles que podem ser quebrados em processamentos paralelos.
- Existem novas aplicações nas áreas de pesquisa e simulação que podem se beneficiar dos ambientes de computação grid.

Existem várias características que devem ser atendidas nos planejamentos e projetos voltados para computação grid, independente se é um sistema novo que está se iniciando, a expansão de um sistema existente, ou se é a mudança de arquitetura centralizada baseada em servidores para a arquitetura grid. As características mais importantes são:

- **Escalabilidade** – A escalabilidade se aplica de duas maneiras: deve ser possível crescer os dispositivo físico; por exemplo: adicionar desempenho, adicionar discos, etc, ou o crescimento deve ser ao acrescentar novos dispositivos. Em ambos os casos, o gerenciamento do sistema como um todo deve acomodar o crescimento de maneira transparente ao ambiente, aplicações e usuários.
- **Disponibilidade** – Além da confiabilidade esperada dos componentes do sistema, ele deve também permitir movimentações, ajustes e adições imediatas, sem afetar usuários e sem interromper aplicações.
- **Gerenciamento** – A interface de gerenciamento deve apresentar uma visão unificada do sistema distribuído, incluindo todas aplicações e recursos computacionais e de armazenamento de dados que compõe o sistema, apresentando como se fosse um grande sistema.

- **Servicibilidade** – Se um recurso do sistema deve ser removido ou desligado temporariamente para manutenção, deve ser possível fazê-lo sem afetar os usuários e sem interromper as aplicações que não dependem daquele recurso.

5.6. Organizações voltadas à Padronização de Computação Grid

Vários esforços tem sido realizados na padronização e desenvolvimento de computação grid. O objetivo é gerar padrões que possam ser usados para alavancar a utilização e disseminação de tecnologias Grid. Existe uma série de organizações que formam criadas com esses objetivos, algumas das principais organizações são:

Global Grid Forum (GGF) – Organização criada a partir de uma série de seminários, encontros e workshops sobre computação grid do evento "Supercomputing 1998". É uma comunidade de usuários, desenvolvedores e fabricantes que buscam a padronização mundial de computação grid. Hoje são mais de 400 representantes da indústria e centros de pesquisa em mais de 50 países. Seu objetivo é promover e suportar o desenvolvimento e implementação de tecnologias grid e aplicações através da criação e documentação de "melhores práticas" – especificações técnicas, experiência de usuários e guias de implementação. Site: http://www.ggf.org .

Enterprise Grid Alliance (EGA), Consórcio de fornecedores e usuários, criado em abril de 2004. Seu objetivo é acelerar o desenvolvimento e adoção de soluções grid específicas para os ambientes comercial e técnico das empresas. Site: http://www.gridalliance.org .

Globus Alliance – Comunidade de organizações e indivíduos. Foi criada em 2003, fica localizada no Laboratório Nacional Argonne, congrega renomadas universidades e centros de pesquisa mundiais. Conduz pesquisa e desenvolve tecnologias, padrões, e sistemas fundamentais à construção de grids computacionais. Associação dedicada a desenvolver tecnologias fundamentais necessárias a construção de infra-estruturas para computação grid. Um dos grandes produtos sendo desenvolvidos pela Globus Alliance e muitos outros contribuintes ao redor do mundo é o Globus Toolkit – conjunto de

ferramentas de *software* aberto usado para a criação de aplicações e sistemas grid. Site: www.globus.org .

W3C – *World Wide Web Consortium* – é um consórcio internacional, criado em 1994 por Tim Berners-Lee[42] e outros, busca desenvolver padrões de tecnologias baseadas de interoperabilidade (especificações, programas, guias, e ferramentas) para conduzir a Web a seu potencial pleno. Site: www.w3c.org.

OASIS – *Organization for the Advancement of Structured Information Standards* – é um consórcio internacional, sem fins lucrativos, que foi fundado em 1993, possui representantes de mais de 600 organizações e indivíduos de mais de 100 países. O consórcio busca direcionar o desenvolvimento, convergência e adoção de padrões de e-business. O grande infoque tem sido o desenvolvimento de padrões de "Web services". Site: www.oasis-open.org.

DMTF – *Distributed Management Task Force, Inc* – fundada em 1992, é uma organização da indústria composta por mais de 200 organizações, que direciona o desenvolvimento, adoção e interoperabilidade de padrões de gerenciamento e tecnologia de integração para os ambientes empresariais e de Internet. Entre as tecnologias desenvolvidas pela DMTF destaca-se o modelo CIM (*Common Information Model*) - modelo comum de dados usado para implementar formas de gerenciamento de informações em ambientes de redes e empresariais. Site: www.dmtf.org .

SNIA – *Storage Network Industry Association* – associação cujos membros são dedicados a assegurar que redes de armazenamento de dados se tornem soluções completas e confiáveis por toda comunidade de TI. Site www.snia.org.

Os conceitos de computação grid tem muito em haver com com arquiteturas de redes de armazenamento de dados, pois os ambientes de computação grid presupõe que, tanto os sistemas computacionais, quanto os dados armazenados estão disponíveis através de redes. As formas pelas quais esses sistemas são conectados, ou que protocolos são usados para comunicação entre eles não importa, basta que sejam baseados em padrões que possam ser adotados pela grande maioria das empresas de forma fácil e econômica.

[42] Tim Berners-Lee foi o inventor a World Wide Web em 1989.

6. Conclusões

"NAS e SAN estão convergindo para um armazenamento de bloco de dados e arquivos baseado em fabric (FAS) ... Acreditamos que 70% de todo armazenamento será FAS na metade da década"

Roger W. Cox, Gartner Dataquest, 2002.

As redes de armazenamento de dados representam uma mudança de paradigma em como os dados são armazenados, acessados e gerenciados pelos sistemas computacionais. Consistente com qualquer mudança tecnológica, elas trazem consigo o desafio de entender novos modelos de arquiteturas de armazenamento e, mais importante, como integrar essas novas soluções dentro dos dinâmicos ambientes computacionais das organizações.

As redes de armazenamento podem ser consideradas entre as principais melhorias nas arquiteturas de computadores desde a revolução vinda com o ambiente cliente/servidor no final da década de 1980. Desde aquela época, a necessidade de armazenar grandes quantidades de informações foi sempre crescente e dinâmica. O sucesso do ambiente cliente/servidor impulsionou o poder computacional dos componentes tecnológicos do servidor a novos níveis que, por sua vez, excedeu os limites dos dispositivos de armazenamento tradicionais.

Essa situação direcionou sistemas e fabricantes de dispositivos de armazenamento de dados a desenvolverem produtos e soluções, tanto na área de *hardware* quanto em *software*, para atender à crescente necessidade de armazenamento. A busca de dispositivos de dados de maior capacidade e acesso mais rápido levaram fabricantes e comunidade de usuários ao desenvolvimento de novos padrões.

Somente dispositivos de maior capacidade e de acesso mais rápido não ofereciam, de maneira eficiente, o gerenciamento da grande demanda de armazenamento que começou a crescer durante os anos 90. A indústria e o ambiente de armazenamento de dados das organizações da época buscaram por novos meios de implantação, acesso e gerenciamento dessa grande capacidade de dados. Isso só aconteceu após a mudança de paradigma de conectividade do armazenamento do modelo de conexão direta para o modelo baseado em redes. Com os conceitos

de redes de armazenamento emergindo como uma solução viável de armazenamento, que permitiu:

- Primeiro, a redescoberta de uma forma de armazenamento em rede, criada sob a orientação dos sistemas especializados Unix, os dispositivos NAS;
- Em segundo lugar, ampliação e encorajamento da integração de novas tecnologias com a criação de redes separadas para armazenamento, como foi o desenvolvendo da arquitetura SAN.

Com base em todo o histórico exposto neste estudo, as redes de armazenamento justificam os investimentos e apresentam melhores resultados operacionais, gerenciais e financeiros para as organizações, pelas seguintes razões:

- Maior quantidade de dados – Os dados armazenados *on-line* e transferidos entre o servidor e o armazenamento têm aumentado em volume e tamanho e, a quantidade de dados transmitidos entre o servidor e os seus clientes é muito maior, direcionado por dados estruturados (bancos de dados) combinados com dados não-estruturados (texto, imagens, áudio e vídeo);
- Mais quantidade de fontes de dados – As aplicações devem trabalhar com várias fontes de dados para satisfazer às transações dos clientes. Isso significa que há várias unidades de armazenamento *on-line* as quais o servidor deve se conectar para processar a aplicação;
- Estratégia de distribuição simples – Os resultados das aplicações devem ser colocados em local centralizado e consolidado para ser acessado por todos.

Deriva-se, assim, algumas conclusões:

O problema referente ao volume e taxa de transferência de dados, é um assunto discutido pelas organizações, que buscam meios pelos quais elas vão endereçar esses problemas, ou seja, a conexão entre servidor e unidade de armazenamento requer uma taxa de transferência de dados mais eficiente. O modelo armazenamento usa o dispositivo de armazenamento ligado diretamente ao servidor é limitado, principalmente nos ambientes multi-usuário. O problema referente ao tamanho dos dados é resolvido com os recursos das redes de armazenamento, e para a questão

das taxas de transferência de dados, a solução é adoção de meios de transporte de dados rápidos, como é o caso de soluções baseadas em Gigabit Ethernet e Fibre Channel.

Por fim, a integração do armazenamento pode ajudar a fornecer a vantagem competitiva que organizações precisam para sobreviver num mundo globalizado e competitivo. Com a alta taxa de crescimento da indústria de rede de armazenamento, qualquer previsão para o futuro deve ser tratada com cuidado. Certamente, são esperados dispositivos de velocidades de acesso mais altas, 4 Gigabit Fibre-Channel e 10 Gigabit Ethernet são hoje uma realidade tecnológica.

A habilidade para as organizações implementarem as "redes de armazenamento abertas" e fazer um mistura de dispositivos e componentes de armazenamento de fabricantes heterogêneos esta aumentando dia a dia. Soluções novas, como o protocolo iSCSI, deve acelerar muito a adoção de armazenamento de dados em rede. O protocolo o TCP/IP, depois de se tornar o protocolo "*de-facto*" das redes locais e da Internet, começa se consolidar na comunicação de voz através do VOIP[43] e, também, como protocolo para o transporte das informações dos dispositivos de armazenamento dados. Deve-se esclarecer que as redes de armazenamento de dados baseadas em Fibre Channel não irão desaparecer de imediato, pelo contrário ainda oferecem uma base tecnológica muito eficiente para grandes sistemas.

Com a chegada de computação grid, a busca pela maximização de utilização de recursos, com uma redução considerável nos investimentos, traz novas oportunidades para os ambientes de armazenamento de dados em rede. Estes farão, com certeza, parte da infra-estrutura básica do armazenamento de dados do futuro.

[43] Voice Over Internet Protocol

Referências Bibliográficas

[1] Artigo "43% *of companies that suffer huge data loss never re-open*" escrito por Brendan Killeen para o jornal online "Thomas Crosbie Media" em 2001, http://archives.tcm.ie/businesspost/2001/09/23/story489475828.asp

[2] Estudo "How *Much Information? 2003*" realizado pelo "School of Information Management and Systems" da Universidade da Califórnia - Berkeley, em 2003 http://www.sims.berkeley.edu:8000/research/projects/how-much-info-2003/

[3] Press release do IDC "*External Disk Storage Systems Archieves Record Growth in the Third Quarter of 2005*", Dezembro de 2005, http://www.idc.com/getdoc.jsp?containerId=prUS20019605

[4] Press release do IDC Latin America "*Armazenamento de dados com controle das informações e investimentos reduzidos está entre os grandes desafios corporativos*", Setembro de 2005, http://www.idclatin.com/miami/telas/pagina.asp?id_area=3&n=199

[5] Merrill Lynch & Company e McKinsey & Company, "*The Storage Report Customer Perspectives and Industry Evolution*", Junho 2001.

[6] PRESTON, W. Curtis. "*Using SANs and NAS*", 1ª edição, Editora O´Reilly & associates, Fevereiro de 2002.

[7] LAKATOS, Eva Maria e MARCONI, Marina de Andrade, "*Metodologia Científica*", Atlas editora, 1992.

[8] WIKIPEDIA, "*Timeline of Computing*", Disponível em: <http://en.wikipedia.org/wiki/Timeline_of_computing >, Acessado em 25 de Maio de 2006.

[9] IBM Archives, "*History of IBM Archives:1956*", Disponível em <http://www.ibm.com/ibm/history/history/year_1956.html>, Acessado em 25 de Maio de 2006.

[10] MASSIGLIA, Paul, "*Disk Storage Management*", 2ª edição, Veritas Software Corporation, 2001.

[11] BLUNDEN, Mark; ALBRECHT, Bernd; BARDINET, Jean-Paul e MELLISH, Barry, "*Monitoring and Managing IBM SSA Disk Subsystems*", 1ª edição, IBM RedBook, 1998.

[12] ATA/ATAPI History, "*The Brief History of ATA and ATAPI*", Disponível em: <http://www.ata-atapi.com/hist.htm>. Acessado em 25 de Maio de 2006.

[13] SATA-IO, "*Serial ATA; A Comparation with Ultra ATA Technology*", Disponível em <http://www.sata-io.org/docs/serialata_a_comparation_with_ultra_ata_technology.pdf>, Acessado em 25 de Maio de 2006.

[14] The Official SCSI FAQ, "*Frequently Asked Questions List for comp.periphs.scsi*", Disponível em: <http://home.comcast.net/~SCSIguy/SCSI_FAQ/scsifaq.html>, Acessado em 25 de Maio de 2006.

[15] WILLIAMS, Bill, *"The Business Case for Storage Networks"*, 1ª edição, Cisco Press, 2005.

[16] WORDEN, Daniel J., "*Storage Networks*", 1ª edição, Editora Apress, 2004.

[17] PATTERSON, David A.; GIBSON, Garth; KATZ, Randy H., "*A Case for Redundant Arrays of Inexpensive Disks (RAID)*", International Conference on management of data, 1988.

[18] TANENBAUM, Andrew S. "*Redes de Computadores*", 3ª edição, Editora Campus, 1997.

[19] WIKIPEDIA, "*OSI model*", Disponível em http://en.wikipedia.org/wiki/OSI_model>, Acessado em 25 de Maio de 2006.

[20] J.F. Kurose and K. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet*, Addison-Wesley, Reading, MA, 2001.

[21] IEEE, "*IEEE 802.3 CSMA/CD (ETHERNET)*", Disponível em: http://grouper.ieee.org/groups/802/3/ , Acessado em 25 de Maio de 2006.

[22] ABRAMSON, Norman, "*Development of the ALOHANET*", IEEE Transactions on Information Theory, Vol II31, 1985.

[23] FARLEY, Marc, "*Building Storage Networks*", 1ª edição, Editora Osborne/McGraw-Hill, 2000.

[24] SIMITCI, Huseyin, *"Storage Network Performance Analysis"*, 1ª edição, Wiley Publishing, 2003.

[25] STEVENS, W. Richard. *"TCP/IP Illustrated – Volume 1"*, 19ª edição, Addison-Wesley, 2001.

[26] LEACH, Paul e PERRY, Dan, "*CIFS: A Common Internet File System*", Microsoft Interactive Developer, 1996.

[27] SNIA CIFS Technical Work Group, "Common Internet File System (CIFS) Technical Reference", SNIA, 2002.

[28] CLARK, Tom, "*IP SANs – A Guide to iSCSI, iFCP, and FCIP Protocols for Storage Area",* Networks. 2002

[29] CLARK, Tom, "*Designing Storage Area Networks*", 2ª edição, Addison-Wesley, 2003.

[30] COX, Roger, "*The Centerpiece of the New IT Infrastructure Universe*" Conferência Gartner PlanetStorage 2003, Las Vegas, EUA.

[31] SPALDING, Robert, "*Stotage Networks: The Complete Reference*", 1ª edição, McGraw-Hill, 2003.

[32] POLLACK, Daniel, "*Practical Storage Area Networking*", 1ª edição, Addison-Wesley, 2003.

[33] TAURION, Cezar; "*Grid Computing: Um Novo Paradigma Computacional*", Editora Brasport, 2004.

[34] IDC, "*Worldwide HPC Technical Grid 2005-2008 Forecast Adoption Rates in Mid-2005 (IDC #33870)*", IDC, Outubro 2005.

[35] Thomas A. DeFanti, I. Foster, M.E. Papka, R. Stevens, e T. Kuhfuss, "*Overview of the I_WAY: Wide Area Visual Supercomputing*", International Journal of Supercomputing Aplications, 1996.

[36] Foster, Ian e Kesselman, Carl, "*The Grid: Blueprint for a New Computing Infrastructure*". Morgan Kaufmann Publishers, 1999.

[37] Foster, Ian e Kesselman, Carl, "*The Grid2: Blueprint for a New Computing Infrastructure*". Morgan Kaufmann Publishers, 2004.

[38] Foster, Ian; Kesselman, Carl e Tuecke, Steve. "*The Anatomy of the Grid: Enabling Scalable Virtual Organizations*", International Journal of Supercomputer Applications, 2001.

[39] JOSEPH, Joshy e FELLENSTEIN, Craig, " Grid Computing", 3ª edição, IBM Press, 2004.

[40] National Academy of Sciences, "*Realizing the Information Future: The Internet and Beyond*". Disponível em : < http://www.nap.edu/readingroom/books/rtif/ >. Acesso em: 25 de Maio de 2006.

[41] W. Allock, J. Bester, J. Bresnaham, A. Chervenak, I. Foster, C. Kesselman, S. Meder, V. Nefedova, D. Quesnel, e S. Tuecke, "*Data Management and Transfer in High-Performance Computational Grid Environments*", Parallel Computing, 2002.