

**BACKUP E RECUPERAÇÃO DE INFORMAÇÕES
UTILIZANDO A FERRAMENTA IBM TIVOLI STORAGE
MANAGER NO BANCO XYZ**

FERNANDA ZIMMER

**TRABALHO DE CONCLUSÃO EM GESTÃO DE
SEGURANÇA DA INFORMAÇÃO**

**FACULDADE DE TECNOLOGIA
UNIVERSIDADE DE BRASÍLIA**

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**BACKUP E RECUPERAÇÃO DE INFORMAÇÕES
UTILIZANDO A FERRAMENTA IBM TIVOLI STORAGE
MANAGER NO BANCO XYZ**

FERNANDA ZIMMER

ORIENTADORA: DRA. EDNA DIAS CANEDO

**TRABALHO DE CONCLUSÃO EM GESTÃO DE SEGURANÇA DA
INFORMAÇÃO**

BRASÍLIA/DF: JULHO - 2017

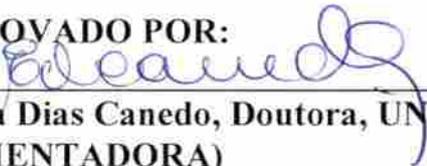
**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**BACKUP E RECUPERAÇÃO DE INFORMAÇÕES UTILIZANDO A
FERRAMENTA IBM TIVOLI STORAGE MANAGER NO BANCO
XYZ**

FERNANDA ZIMMER

**TRABALHO DE CONCLUSÃO SUBMETIDO AO DEPARTAMENTO
DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA
DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
PÓS-GRADUAÇÃO EM GESTÃO DE SEGURANÇA DA
INFORMAÇÃO.**

APROVADO POR:



**Edna Dias Canedo, Doutora, UNB/FGA
(ORIENTADORA)**

**Laerte Peotta de Melo, Doutor, UNB
(EXAMINADOR INTERNO)**

**Eliane Carneiro Soares, Mestre, SEEDF
(EXAMINADORA EXTERNO)**

BRASÍLIA/DF, JULHO DE 2017.

FICHA CATALOGRÁFICA

Zimmer, Fernanda.

Backup e Recuperação de Informações Utilizando a Ferramenta Tivoli Storage Manager no Banco XYZ [Distrito Federal], 2017.

56p.

Trabalho de Conclusão – Especialização Gestão de Segurança da Informação – Universidade de Brasília, Faculdade de Tecnologia.

Departamento de Engenharia Elétrica.

- | | |
|---------------------------|-------------------------------|
| 1. Backup de Dados | 2. Recuperação de Informações |
| 3. Tivoli Storage Manager | 4. Armazenamento de Dados |
| 5. Mainframe | |

REFERÊNCIA BIBLIOGRÁFICA

Zimmer, Fernanda. (2017). Backup e Recuperação de Informações Utilizando a Ferramenta Tivoli Storage Manager no Banco XYZ. Trabalho de Conclusão, Publicação UnBLabRedes.MFE.007/2017, Faculdade de Tecnologia, Universidade de Brasília, Brasília, DF, 56p.

CESSÃO DE DIREITOS

AUTOR: Fernanda Zimmer

TÍTULO DO TRABALHO DE CONCLUSÃO: Backup e Recuperação de Informações Utilizando a Ferramenta Tivoli Storage Manager no Banco XYZ.

GRAU / ANO: Especialização / 2017

É concedida à Universidade de Brasília permissão para reproduzir cópias desse trabalho de conclusão de curso e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. A autora reserva outros direitos de publicação e nenhuma parte desse trabalho de conclusão pode ser reproduzida sem autorização por escrito da autora.



Fernanda Zimmer

Avenida Pau Brasil Lote 14 Apartamento 1201, Via Brisa– Águas Claras

CEP: 71.926-000 - Brasília - DF

Tel. 55 – 61 – 992321254 / feunisc@gmail.com

RESUMO

BACKUP E RECUPERAÇÃO DE INFORMAÇÕES UTILIZANDO A FERRAMENTA IBM TIVOLI STORAGE MANAGER NO BANCO XYZ

Autora: Fernanda Zimmer

Orientadora: Professora Dra. Edna Dias Canedo

Programa de Pós-graduação em Gestão de Segurança da Informação

Brasília, Julho de 2017.

Este trabalho tem como objetivo analisar os aspectos da ferramenta IBM Tivoli Storage Manager e apresentar as vantagens de seu uso para o backup e recuperação de informações, no ambiente do Banco XYZ. Essa análise possibilita avaliar as suas vantagens e as aplicações no mundo corporativo atual. O objetivo é demonstrar que as informações críticas de empresas de negócio devem ser gerenciadas e armazenadas de forma segura e responsável, e a aplicação da ferramenta IBM Tivoli Storage Manager provê essas premissas. Assim, como, em caso de desastres de qualquer natureza, a informação deve ser recuperada de forma rápida, eficiente e segura, a fim de evitar grandes prejuízos à administração. Trata-se de um levantamento na bibliografia sobre o assunto e a apresentação de uma ferramenta desenvolvida para esse fim, Tivoli Storage Manager (TSM), ferramenta licenciada, desenvolvida pela IBM para atender à grande demanda na área. Além da revisão, exemplifica-se com a aplicação em um ambiente corporativo atual, o ambiente mainframe, utilizado por diversos ramos de negócios (como exemplo, o ramo financeiro), e que se caracteriza por grande volume de dados gerenciados, processados e armazenados.

ABSTRACT

BACKUP AND RECOVERY INFORMATION USING THE IBM TIVOLI STORAGE MANAGER TOOL IN BANK XYZ

Author: Fernanda Zimmer

Supervisor: Professora Dra. Edna Dias Canedo

Programa de Pós-graduação em Gestão de Segurança da Informação

Brasília, July of 2017.

This work aims to analyze aspects of IBM Tivoli Storage Manager tool and present the advantages of its use for the backup and recovery of information on the Bank XYZ environment. The goal is to demonstrate that critical business information companies must be managed and stored in a safe and responsible manner, and the implementation of IBM Tivoli Storage Manager tool provides these premises. Thus, as in the case of disasters of any kind, information must be recovered quickly, efficiently and safely in order to avoid major damage to the administration. This is a survey of the literature on the subject and the presentation of a tool developed for this purpose, Tivoli Storage Manager (TSM), licensed tool developed by IBM to meet the large demand in the area. In addition to the review, exemplified with the application in an actual enterprise environment, the mainframe environment, used by several business areas (for example, the financial sector), and which is characterized by large volume of data managed, processed and stored.

SUMÁRIO

1 - INTRODUÇÃO	1
1.1 - MOTIVAÇÃO.....	1
1.2 - OBJETIVOS DO TRABALHO.....	2
1.2.1 - <i>Objetivo Geral</i>	2
1.2.2 - <i>Objetivos Específicos</i>	2
1.3 - METODOLOGIA DE PESQUISA.....	2
1.4 - ORGANIZAÇÃO DO TRABALHO.....	3
2 - O ARMAZENAMENTO DE INFORMAÇÕES	4
2.1 - DADOS E INFORMAÇÕES.....	4
2.2 - ARMAZENAMENTO.....	5
2.2.1 - <i>Dispositivos de Armazenamento</i>	6
2.2.2 - <i>Evolução da Tecnologia de Armazenamento</i>	8
2.2.3 - <i>Gerenciamento da Infraestrutura de Armazenamento</i>	10
2.2.4 - <i>Ambiente do Sistema de Armazenamento</i>	13
2.2.4.1 - Host.....	13
2.2.4.2 - Conectividade.....	14
3 - O BACKUP E A RECUPERAÇÃO DE INFORMAÇÕES	16
3.1 - CARACTERÍSTICAS DO BACKUP.....	16
3.2 - OBJETIVOS DO BACKUP.....	18
3.3 - TIPOS DE BACKUP.....	19
3.4 - ESTRATÉGIA E PROCESSO DE BACKUP.....	20
3.5 - RECUPERAÇÃO DE INFORMAÇÕES.....	23
4 - O MAINFRAME	28
4.1 - CARACTERÍSTICAS DO MAINFRAME.....	30
4.2 - VANTAGENS DO MAINFRAME.....	34
5 - A FERRAMENTA IBM TIVOLI STORAGE MANAGER (TSM)	37
5.1 - CARACTERÍSTICAS DO TSM.....	38
5.1.1 - <i>Arquitetura do TSM</i>	41
5.2 - VANTAGENS DO TSM.....	43

6 - O TSM NO AMBIENTE DO BANCO XYZ	47
7 - CONCLUSÕES	53
7.1 – TRABALHOS FUTUROS	53
REFERÊNCIAS BIBLIOGRÁFICAS	54

LISTA DE FIGURAS

Figura 2.1 – Dispositivos de Armazenamento – Dispositivos Removíveis.....	7
Figura 2.2 - Dispositivos de Armazenamento – Dispositivos não Removíveis	7
Figura 2.3 - Dispositivos de Armazenamento – Capacidade, Velocidade de Leitura e Gravação	8
Figura 2.4 – Evolução das Arquiteturas de Armazenamento.....	9
Figura 3.1 – Arquitetura e Processos de Backup.....	22
Figura 3.2 – Operação de Backup	26
Figura 3.3 – Operação de Restauração	27
Figura 4.1 – Evolução dos Modelos de Mainframe	29
Figura 4.2 – Principais Componentes do Mainframe	32
Figura 4.3 – Processamento Online e Batch.....	35
Figura 5.1 – Arquitetura do TSM	42
Figura 5.2 – TSM para z/OS.....	48
Figura 6.1 – Topologia do TSM no Banco XYZ	50
Figura 6.2 – Registro de node no TSM Server	51
Figura 6.3 – Associação de cloptset ao node.....	52
Figura 6.4 – Associação de schedule ao node	52

LISTA DE ACRÔNIMOS

ACP	<i>Airline Control Program</i>
API	<i>Application Programming Interface</i>
CD	<i>Compact Disk</i>
CMS	<i>Conversational Monitor System</i>
CPU	<i>Central Processing Unit</i>
DAS	<i>Direct-Attached Storage</i>
DRM	<i>Disaster Recovery Management</i>
DVD	<i>Digital Versatile Disk</i>
ETR	<i>External Time Reference</i>
FCP	<i>Fibre Channel Protocol</i>
FICON	<i>Fibre Connectivity</i>
GPFS	<i>General Parallel File System</i>
HBA	<i>Host Bus Adapter</i>
HSM	<i>Hierarchical Storage Management</i>
IDE/ATA	<i>Integrated Device Electronics/ Advanced Technology Attachment</i>
IP-SAN	<i>Internet Protocol SAN</i>
iSCSI	<i>Internet SCSI</i>
ISPF	<i>Interactive System Productivity Facility</i>
JES	<i>Job Entry Subsystem</i>
LAN	<i>Local Area Network</i>
LPAR	<i>Logical Partition</i>
LTO	<i>Linear Tape-Open</i>
LV	<i>Logical Volume</i>
LVM	<i>Logical Volume Manager</i>
MD5	<i>Message Digest 5</i>
MEC	<i>Ministério da Educação</i>
MVS	<i>Multiple Virtual System</i>
NAS	<i>Network-Attached Storage</i>
NDMP	<i>Network Data Management Protocol</i>
PCI	<i>Peripheral Component Interconnect</i>
RAID	<i>Redundant Array of Independent Disk</i>

RAM	<i>Random-Access Memory</i>
ROM	<i>Ready-Only Memory</i>
RPO	<i>Recovery Point Object</i>
RTO	<i>Recovery Time Object</i>
SAN	<i>Storage Area Network</i>
SCSI	<i>Small Computer System Interface</i>
SHA-1	<i>Secure Hash Algorithm</i>
SMIT	<i>System Management Interface</i>
TPF	<i>Transaction Processing Facility</i>
TSM	<i>Tivoli Storage Manager</i>
TSO	<i>Time Sharing Option</i>
UC	Unidade de Controle
ULA	Unidade Lógica Aritmética
VG	<i>Volume Group</i>
VM	<i>Virtual Machine</i>
WAN	<i>Wide Area Network</i>

1 INTRODUÇÃO

Desde o ataque de 11 de setembro de 2001, a proteção de dados corporativos mudou em muitos aspectos. Aumentou a importância de as empresas possuírem um backup dos seus dados em mais de um local. Além disso, um plano de contingência de negócio, descentralização do site recovery e testes de backup e recuperação foram aprimorados para evitar a descontinuidade do negócio.

Tradicionalmente, o mundo corporativo busca soluções para o armazenamento de suas informações, assim como a disponibilidade rápida e eficiente em caso de recuperação dessas informações. Tal solução vem de encontro a uma das grandes vantagens do uso da ferramenta IBM Tivoli Storage Manager (TSM), que possibilita o backup, armazenamento e recuperação de dados em caso de desastres.

O TSM é uma ferramenta licenciada, desenvolvida pela IBM, com arquitetura cliente-servidor que fornece serviços de gerenciamento do armazenamento de informações, em ambientes com diversas plataformas, tais como AIX, Linux, Windows e z/OS. Possui aplicação em diversos ambientes, desde laptops até ambientes como os mainframes, que gerenciam e processam um grande volume de dados.

O presente trabalho apresenta a utilização da ferramenta TSM pelo Banco XYZ para backup, restore e recovery dos dados corporativos, que são importantes para o negócio. Informa quais os tipos de backup realizados, vantagens e exemplos práticos do uso do software.

1.1. Motivação

O ambiente de backup corporativo utiliza diversas ferramentas para armazenamento e recuperação de informações, sendo uma delas a ferramenta IBM – Tivoli Storage Manager. Os estudos relacionados ao tema não apresentam a aplicação em ambiente bancário, de alta disponibilidade e criticidade. Apresentar a aplicabilidade da ferramenta no processo de backup, armazenamento e recuperação de informações em ambiente tecnológico em empresa do ramo financeiro, como o Banco XYZ.

1.2. Objetivos do Trabalho

1.2.1 Objetivo geral

O principal objetivo deste trabalho é analisar os aspectos da ferramenta IBM Tivoli Storage Manager e apresentar as vantagens de seu uso para o backup e recuperação de informações em um ambiente tecnológico de uma empresa do ramo financeiro, o Banco XYZ.

1.2.2 Objetivos específicos

- a. Explicar a importância de armazenar as informações;
- b. Apresentar os conceitos backup, recuperação, restauração e armazenamento de informações;
- c. Apresentar o conceito de ambiente mainframe; apresentar formas, dispositivos e ferramentas de backup e recuperação de informações;
- d. Descrever as características e funcionalidades da ferramenta da IBM, o TSM - Tivoli Storage Manager e sua aplicação no ambiente do Banco XYZ;
- e. Apresentar as vantagens da ferramenta para o backup e recuperação de informações;
- f. Demonstrar o uso da ferramenta através de comandos e figuras de exemplo.

1.3. Metodologia de Pesquisa

A metodologia adotada neste estudo é a pesquisa bibliográfica exploratória, onde a coleta de informações e conceitos é realizada em livros, artigos científicos, revistas relacionadas ao tema e publicações eletrônicas. A análise de dados é realizada através da análise documental, com a organização e interpretação dos conceitos apresentados, e consolidação dos dados para a apresentação neste trabalho.

A obtenção de telas da ferramenta é realizada junto ao Banco XYZ e apresentadas informações sobre a estrutura na empresa, com números para exemplificar o ambiente através de dados de backup e servidores.

Serão apresentados os conceitos de backup, recuperação de informações, apresentação da ferramenta Tivoli Storage Manager, com seus conceitos, e a vantagem de uso por parte do Banco XYZ.

1.4. Organização do Trabalho

O trabalho está organizado como segue.

O Capítulo 2 apresenta o conceito de dados e informações. Detalhe-se armazenamento, com suas características, dispositivos, ambiente, gerenciamento e evolução.

No Capítulo 3 se demonstra o conceito de backup, demonstrando seus tipos, objetivos, estratégias e o processo de backup, além de sua importância para as empresas. Em complemento, são apresentados o processo e operação de recuperação de informações.

O Capítulo 4 detalha o ambiente mainframe, o mais utilizado em bancos.

O Capítulo 5 apresenta detalhadamente a ferramenta IBM TSM e sua aplicabilidade no ambiente do Banco XYZ.

O Capítulo 6 apresenta as considerações finais deste trabalho.

2 O ARMAZENAMENTO DE INFORMAÇÕES

Este capítulo tem como foco a revisão dos principais conceitos de armazenamento de informações, apresentado seus componentes de armazenamento. Na seção 2.1 é apresentado o conceito de dados e informações. Na seção 2.2 são abordados os conceitos relacionados a armazenamento, tais como: dispositivos, evolução, gerenciamento, ambiente e informações sobre conectividade.

2.1. Dados e Informações

De acordo com Setzer (2009), os conceitos de dados e informações são diferentes e complementares. Dado pode ser definido com uma sequência de símbolos, tais como um texto, imagem, sons e animação. Sua descrição pode ser realizada através de representações formais, estruturais e, por serem quantificáveis, podem ser armazenados e processados por um computador.

Já o conceito de informação define que é uma abstração informal que, através de texto, imagem, sons e animação que representam algo, possui semântica. A informação não é passível de processamento em computador, somente de armazenamento, quando é formada por dados. “Um dado é puramente objetivo - não depende do seu usuário. A informação é objetivo-subjetiva no sentido que é descrita de uma forma objetiva (textos, figuras, etc.), mas seu significado é subjetivo, dependente do usuário” (Setzer, 2009, np).

Atualmente, a informação é de grande importância para o desenvolvimento do mundo corporativo. Conforme Silva *et al.* (1999 apud Monte; Lopes, 2004), informação é sinônimo de dados do conhecimento registrado, ou seja, é o registro das atividades humanas em documentos (informação documental). Em outra definição, “dados são um conjunto de fatos em estado bruto a partir dos quais conclusões podem ser tiradas. [...] Informação é a inteligência e o conhecimento derivados dos dados” (Somasundaram *et al.*, 2010, p.27 e 29).

Segundo Somasundaram *et al.* (2010), existem algumas premissas que justificam o aumento de dados digitais, conforme segue:

- Aumento da capacidade de processamento de dados: os computadores atuais possuem um alto poder de processamento e armazenamento, o que permite a conversão de vários tipos de conteúdo para o formato digital;

- Menor custo do armazenamento digital: foram desenvolvidas soluções de baixo custo para o armazenamento em função dos avanços tecnológicos e diminuição do custo com a matéria prima, tais como dispositivos como o pen-drive e o HD externo. O que resultou no aumento da taxa de geração e armazenamento de dados;
- Tecnologias de comunicação mais rápidas e acessíveis: a velocidade de transmissão de informações e compartilhamento de dados vêm aumentando progressivamente, facilitando o acesso à informação.

2.2. Armazenamento

Na área de tecnologia, essa informação deve ser armazenada de alguma forma, seja para fins de backup ou para recuperação. Entende-se por armazenamento o conceito de guardar e manter as informações em dispositivos de armazenamento (Somasundaram *et al.*, 2010).

Segundo Laundon (2001), o armazenamento pode ser primário ou secundário. No armazenamento primário o processamento de informações e programas é imediato, para isso são guardados para uso imediato, em estado volátil (que requerem energia elétrica). A tecnologia utilizada é a mais rápida e, por consequência, a mais cara. O acesso é imediato, eletrônico e quase à velocidade da luz. Os tipos de armazenamento primário são: registrador, memória RAM (Random Access Memory), memória cache e memória ROM (Read-only Memory).

Já, no armazenamento secundário, o estado é não volátil, ou seja, não requer energia elétrica para seu funcionamento, assim, acaba retendo as informações mesmo com a máquina desligada. O volume de dados é muito grande e o acesso a esses dados é mais lento, se comparado ao acesso ao armazenamento primário. Como exemplo de armazenamento secundário, podem-se citar os discos magnéticos, ópticos e as fitas magnéticas (Laundon, 2001).

Conforme observado em Dell (2007, np), o conceito de armazenamento “se refere a um dispositivo que hospeda dados”. A classificação do armazenamento pode ser definida por meio da arquitetura, protocolo de rede ou sistemas de arquivos de armazenamento.

A classificação por arquitetura, de acordo com Dell (2007), é definida através das seguintes premissas:

- Armazenamento de conexão direta (DAS – Direct-Attached storage): o sistema de armazenamento é conectado diretamente no servidor, sem necessidade de uma rede de armazenamento para realizar a conexão;

- Armazenamento de conexão com a rede (NAS - Network-Attached storage): o servidor de armazenamento é especializado, possui seu próprio IP, e se comunica com clientes e outros servidores em uma rede local ou de longa distância;
- Rede de área de armazenamento (SAN – Storage area network): os dispositivos de armazenamento são compartilhados e disponibilizados a outros servidores em uma rede local ou de longa distância.

Conforme Dell (2007), a classificação por protocolo de rede é dividida em:

- Internet Protocol Suite (TCP/IP): todos os protocolos de comunicação utilizados na internet. A arquitetura NAS utiliza o protocolo TCP/IP sobre Ethernet;
- Fibre Channel Protocol (FCP): os comandos SCSI são transportados em redes Fibre Channel, um cabeamento com finalidade especial;
- Protocolo Internet SCSI (iSCSI): os servidores enviam comando SCSI para dispositivos de armazenamento SCSI. Está sendo adota por muitas empresas, não necessita de cabeamento especial, apenas que a Ethernet se comunique com redes locais ou de longa distância.

A classificação por sistema de arquivos de armazenamento é dividida, de acordo com Dell (2007), conforme segue:

- Armazenamento em nível de arquivo: os dados são organizados em um sistema de arquivos. Ou seja, para o usuário, os dados são representados como arquivos individuais. A arquitetura NAS utiliza este tipo de armazenamento;
- Armazenamento em nível de bloco: os dados são transformados em uma sequência de bytes ou bits de tamanho fixo, e os sistemas conectados ficam responsáveis por gerenciar os dados brutos. Este tipo de armazenamento é utilizado pela arquitetura SAN;
- Armazenamento baseado em objeto: os dados são organizados em uma estrutura de tamanho flexível denominada objeto, composto por dados (sequência de bytes não interpretada) e metadados (os atributos que descrevem os dados).

2.2.1 Dispositivos de armazenamento

Para que os dados sejam armazenados são utilizados dispositivos de armazenamento. As Figuras 2.1 e 2.2 mostram alguns exemplos de dispositivos utilizados para a guarda de informações.



Figura 2.1: Dispositivos de armazenamento – dispositivos removíveis (Madeira,2010).



Figura 2.2: Dispositivos de armazenamento – dispositivos não removíveis. (Madeira,2010).

Segundo Madeira (2010), dispositivo de armazenamento “é o hardware que possui a finalidade de armazenar o software. Os dados ou informações gravadas em um computador ficam armazenados nesses dispositivos”.

Conforme Soares (2011, np) “os meios de armazenamento que permitem acesso mais rápido, em geral, são mais caros do que aqueles mais lentos. O custo referente à capacidade de armazenamento e portabilidade adicionais varia amplamente [...]”.

Ainda, conforme Madeira (2010), os dispositivos podem ser classificados em:

- Dispositivos removíveis: aqueles dispositivos que podem ser transportados, junto com seus dados. Possuem acesso mais lento de leitura e escrita, e menor capacidade de armazenamento. Exemplos: disquete, cartão flash, pen drive, minidisc, Compact Disk (CD) e fita magnética;
- Dispositivos não removíveis: aqueles dispositivos que não podem ser transportados. Seu acesso de leitura e escrita é mais rápido e possui grande capacidade de armazenamento. Exemplos: disco rígido e disco de estado sólido.

A figura 2.3 mostra a capacidade de armazenamento, velocidade de leitura e velocidade de gravação de alguns dispositivos de armazenamento. De acordo com Madeira (2010, np), “com a evolução tecnológica os dispositivos de armazenamento estão cada vez menores no tamanho e maiores na capacidade”.

Dispositivo	Capacidade	Velocidade de Leitura	Velocidade de Gravação
Disquete	1,44 MB	62,5 KB/s	62,5 KB/s
MiniDisc	160 MB		352 KB/s
CD	700 MB	7800 KB/s (52x)	7200 KB/s (48x)
DVD	4,7 GB e 8,5 GB	32,4 MB/s (24x)	32,4 MB/s (24x)
Blu-ray	25 GB e 50 GB	54 MB/s (12x)	54 MB/s (12x)
Cartão flash	até 48 GB	40 MB/s	12 MB/s
Pendrivel	até 64 GB	30 MB/s	16 MB/s
Fita magnética	até 400 GB	80 MB/s	80 MB/s
Disco de estado sólido	até 256 GB	700 MB/s	250 MB/s
Disco rígido	até 2,5 TB	70 MB/s	70 MB/s

Figura 2.3: Dispositivos de armazenamento – capacidade, velocidade de leitura e gravação (Madeira, 2010).

Quanto aos métodos de acesso aos dados nos dispositivos de armazenamento podem ser divididos conforme observado em Soares (2011):

- Acesso sequencial: os dados são acessados na ordem em que são armazenados, ou seja, sequencialmente;
- Acesso direto: os dados são acessados diretamente, independentes de sua ordem de armazenamento. Normalmente, este acesso é mais rápido que o sequencial.

2.2.2 Evolução da tecnologia de armazenamento

De acordo com Somasundaram *et al.* (2010), a tecnologia de armazenamento evoluiu de armazenamento interno não inteligente para armazenamento em rede inteligente.

Conforme IBM (2012), a evolução desta tecnologia inclui:

- RAID (Redundant Array of Independent Disk): tecnologia que gerencia os requisitos de custo, desempenho e disponibilidade de dados. É utilizada em todas as arquiteturas de armazenamento;
- DAS (Direct-Attached Storage): para o armazenamento, é realizada uma conexão direta a um host ou aos servidores em cluster, sendo a informação armazenada interna ou externamente;

- SAN (Storage Area Network): consiste em uma rede dedicada e de alto desempenho utilizada na comunicação entre servidores e o armazenamento, que é dividido e os acessos aos seus dados são feito através de um servidor. Como vantagem, podemos citar escalabilidade, disponibilidade, desempenho e custo, em comparação ao DAS;
- NAS (Network-Attached Storage): consiste em uma conexão a uma rede LAN já existente e fornece o acesso às informações. Como é arquitetado para aplicativos de servidor de arquivos, possui maior escalabilidade, disponibilidade, desempenho e vantagens de custo em relação a servidores de arquivos em geral;
- IP-SAN (Internet Protocol SAN): evolução das tecnologias utilizadas em SAN e NAS, que fornece comunicação em nível de blocos através de uma LAN ou uma WAN, resultando em uma maior consolidação e disponibilidade dos dados.

Podemos verificar na figura 2.4 a evolução das arquiteturas de armazenamento. Essa evolução é contínua e, conforme Somasundaram *et al.* (2010, p.32), “o que permite às organizações consolidar, proteger, otimizar e alavancar seus dados para obter maior retorno em ativos de informação”.

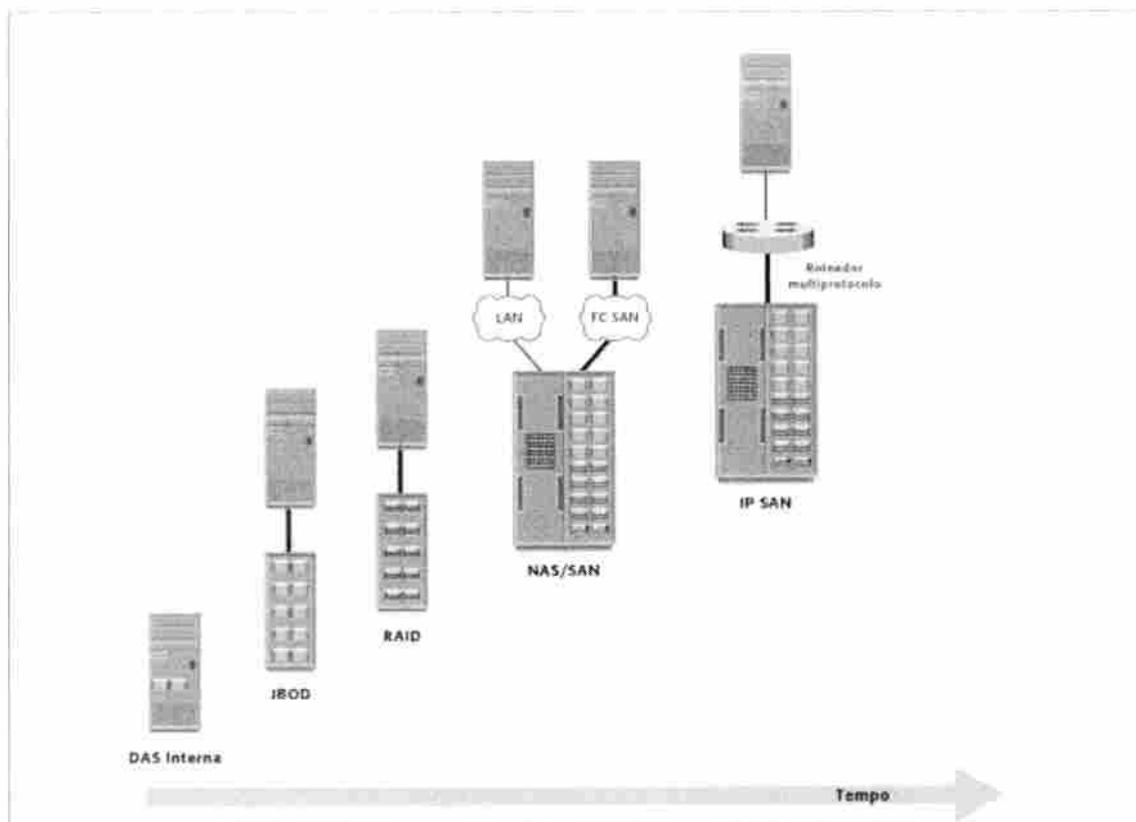


Figura 2.4: Evolução das arquiteturas de armazenamento (Somasundaram *et al.*, 2010).

Conforme Rezende (2012), o conceito de RAID (Redundant Array of Independent Disks – Disposição Redundante de Discos Independentes) consiste em dispor vários discos de baixo custo em uma arquitetura transparente que possua boa capacidade de armazenamento, desempenho e segurança. O servidor “enxerga” essa arquitetura como uma única unidade lógica. Nos níveis de RAID forem divididos de 1 a 5, com características e desempenho variáveis, porém todas com mecanismos de tolerância a falhas. E existe uma arquitetura não redundante chamada de RAID 0.

Segundo IBM (2012), cada vez que os dados mudam, eles precisam ser copiados para os discos de armazenamento, e essa cópia, além de não poder ser feita manualmente, deve ser tolerante a falhas. Assim, um controlador RAID pode manter os discos sincronizados e também gerenciar todas as operações de I/O para o disco.

A combinação de vários discos em uma unidade lógica de armazenamento é denominada stripping. Esse método divide o espaço de armazenamento em faixas alternadas de cada disco que compõe o conjunto. Assim, a carga de I/O do computador é balanceada para diversos discos, rendendo uma melhor performance ao sistema operacional. Sem essa técnica, corre-se o risco das operações de I/O utilizarem somente um disco, subutilizando os demais (Rezende, 2012).

A arquitetura de IP-SAN é a mais atual, pois foi a evolução do Fibre Channel SAN para a consolidação das redes de armazenamento e Ethernet. Conforme destacado em IBM (2012), essa evolução permite:

- Simplificar o gerenciamento de armazenamento;
- Aumentar a utilização dos ativos nas redes de armazenamento;
- Melhorar a flexibilidade dos ambientes de armazenamento para adotar novas soluções;
- Reduzir o custo de infraestrutura por meio da consolidação de ativos;
- Melhorar a eficiência de armazenamento e de rede ethernet.

2.2.3 Gerenciamento da infraestrutura de armazenamento

De acordo com definido por IBM (2012), um dos maiores motivos do gasto de verba em TI é a busca por cada vez mais capacidade de armazenamento de dados. Estes custos não abrangem somente os dispositivos, mas também os custos com espaço físico, eletricidade e o sistema de gerenciamento, sendo este de grande importância.

Conforme Somasundaram *et al.* (2010), existem algumas tarefas importantes no gerenciamento da infraestrutura de armazenamento, conforme observado abaixo:

- **Monitoração:** é a tarefa responsável por coletar e revisar continuamente as informações. A adoção de tal prática inclui segurança, desempenho, acessibilidade e capacidade;
- **Geração de relatórios:** é a tarefa realizada periodicamente, com insumos na utilização de recursos, na capacidade e no desempenho da infraestrutura do armazenamento. Os relatórios baseiam decisões sobre o investimento no negócio e nas opções de armazenamento;
- **Provisionamento:** é a tarefa responsável por fornecer hardware, software e demais recursos necessários ao funcionamento da infraestrutura de armazenamento;
- **Planejamento de capacidade:** é a tarefa onde é assegurado que, tanto os usuários, quanto os aplicativos, tenham suas necessidades satisfeitas, de forma eficaz e controladas;
- **Planejamento de recursos:** é a tarefa onde são avaliados e identificados os recursos necessários para alocação de pessoal, local de armazenamento e tecnologia a ser utilizada.

Com base nessas premissas, Somasundaram *et al.* (2010, p.36) exemplifica:

[...] a utilização da capacidade de armazenamento alocado de um aplicativo pode ser monitorada. Assim que ela atingir um valor crítico, a capacidade adicional de armazenamento poderá ser provisionada para o aplicativo. Se a utilização da capacidade de armazenamento for monitorada e informada apropriadamente, o crescimento do negócio pode ser percebido e futuros requisitos de capacidade podem ser previstos. Isso auxilia uma política pró ativa de gerenciamento de dados.

De acordo com citado em IBM (2012, np), “nem todos os dados são criados da mesma forma. Existem muitos tipos em um ambiente típico de TI, e seu valor muda durante o seu ciclo de vida”. Assim, mesmo que o arquivo de dados seja menos crítico em um momento, nem por isso deixa de ser importante. Em IBM (2) o conceito de tiering é citado como um princípio de gerenciamento do ciclo de vida dos dados, e consiste em uma forma de armazenar dados em vários tipos de mídia, baseado em desempenho, disponibilidade e recuperação. A infraestrutura deste conceito é baseada em camadas, podendo ser de duas até seis camadas.

Em geral, os dados mais recentes e os dados que devem ser acessados com mais frequência são disponibilizados em mídias de armazenamento mais rápidas, porém

mais caras, enquanto os dados menos críticos são armazenados em uma mídia mais barata, porém mais lenta. Os dados destinados à restauração em caso de perda ou corrupção podem ser armazenados localmente – para uma rápida recuperação –, enquanto os dados armazenados somente com propósitos regulamentares podem ser arquivados em mídia de custo mais baixo (IBM, 2012, np).

Entretanto, conforme Somasundaram *et al.* (2010, p.36), existem alguns desafios que devem ser vencidos para garantir uma boa política de gerenciamento de infraestrutura de armazenamento, conforme segue:

- O universo digital em crescimento explosivo: para garantir a disponibilidade dos dados, se faz necessário sua duplicação, o que faz com que a taxa de crescimento da quantidade de informações cresça exponencialmente;
- O aumento na dependência das informações: o mundo corporativo utiliza as informações para a tomada de decisões estratégicas, o que acaba se tornando uma vantagem competitiva no mercado;
- O valor inconstante das informações: conforme o passar do tempo, o valor das informações varia se tornando de valioso a menos importante.

Também devemos considerar a durabilidade dos dispositivos de armazenamento. Enquanto que as informações guardadas em papel são mantidas por um longo período de tempo, os dispositivos de armazenamento em meio eletrônico não dispõem de todo esse tempo. O que se conclui que “um dos grandes desafios dos profissionais da informação será como manter a integridade das informações geradas em computador e como conservá-las para o futuro” (Monte *et al.*, 2004, p.64).

Melhores práticas para um bom desempenho no gerenciamento do armazenamento de informações são listadas por IBM (2012), conforme segue:

- Armazenar o necessário pelo tempo necessário: utilizando tecnologias de compactação de dados, deduplicação e processos de gerenciamento de demanda;
- Obter mais da infraestrutura de armazenamento: através de virtualização, consolidação e monitoramento;
- Armazenar os dados no local apropriado: otimizar o espaço utilizado pelos dados e a hierarquia de armazenamento.

Com esses princípios, conforme IBM (2012, np) “[...] as organizações estarão preparadas para assumir uma abordagem mais inteligente para o gerenciamento de armazenamento, que trata do grande crescimento de armazenamento e do aumento dos custos”.

Concluindo, os dispositivos de armazenamento necessitam de um gerenciamento efetivo para garantir a confiabilidade e disponibilidade das informações a qualquer tempo. Por ser um recurso específico, necessita de especialização em sua implementação e gerenciamento, de acordo com Monte *et al.* (2004).

2.2.4 Ambiente do sistema de armazenamento

De acordo com Somasundaram *et al.* (2010) um ambiente do sistema de armazenamento é um conjunto de componentes que formam o caminho por onde as informações passam, desde o seu aplicativo até o seu armazenamento. Tais componentes são: o host, a conectividade e o armazenamento (sendo este já descrito anteriormente).

2.2.4.1 Host

De acordo com Viana (2012, np):

Por definição, host é qualquer computador ou máquina conectado a uma rede, que conta com número de IP e nome definidos. Essas máquinas são responsáveis por oferecer recursos, informações e serviços aos usuários ou clientes. Por essa abrangência, a palavra pode ser utilizada como designação para diversos casos que envolvam uma máquina e uma rede, desde computadores pessoais à roteadores.

- Os componentes físicos principais de um host são, de acordo com Somasundaram *et al.* (2010):
- A unidade central de processamento (CPU): que se constitui da Unidade Lógica e Aritmética (ULA) que, de acordo com Freire (2017), é responsável pela manipulação dos dados, da Unidade de Controle (UC), registradores e cache de nível 1;
- O armazenamento (memória interna e dispositivos de disco): guardam os dados, de forma temporária ou persistente. Normalmente, existem dois tipos de memória, a RAM (memória de acesso aleatório) e a ROM (memória somente leitura). Conforme Freire (2017), qualquer local da memória pode ser acessado aleatoriamente, a fim de recuperar uma informação em intervalos de tempo fixos;
- Os dispositivos de entrada e saída (I/O): possibilitam a comunicação entre hosts. Esta comunicação é pode ser entre o usuário e o host, entre host e host e entre host e dispositivo de armazenamento. Estes processadores possuem circuitos eletrônicos

que são responsáveis por comunicar a controlar a transferência de dados do computador para o usuário, host e/ou dispositivo de armazenamento (Freire, 2017).

Conforme Freire (2017), os componentes citados fazem parte do hardware do computador. Os componentes lógicos principais de um host são, de acordo com Somasundaram *et al.*(2010):

- Sistema Operacional: possibilita aos aplicativos o acesso aos dados. Realiza a organização e controle dos componentes do hardware e o gerenciamento da alocação de recursos. Provê segurança básica para o acesso à informação e utilização dos recursos gerenciados. Executa, também, as tarefas básicas de gerenciamento de armazenamento dos dados. Segundo Lopes (2008, np), “[...] é uma coleção de programas que inicializam o hardware do computador. Fornece rotinas básicas para controle de dispositivos. Fornece gerência, escalonamento e interação de tarefas. Mantém a integridade de sistema”.
- Driver de dispositivo: permite que o sistema operacional se comunique com algum dispositivo, como a impressora ou o próprio disco rígido. O sistema operacional reconhece o dispositivo a ser utilizado e utiliza uma Application Programming Interface (API), uma interface padrão, para o acesso e controle do dispositivo;
- Gerenciador de volume: também conhecido como gerenciador de volume lógico (LVM – Logical Volume Manager). Aloca espaço do disco rígido em diversos volumes físicos, que são, por sua vez, compostos de volumes lógicos. Tais volumes podem ser redimensionados, permitindo um melhor aproveitamento do espaço disponível. Os volumes lógicos podem ser configurados para obterem um melhor desempenho para os aplicativos e, ao serem espelhados, fornecem uma maior disponibilidade dos dados. Segundo Red Hat (2017, np) “LVM é um método de alocar espaço do disco rígido em volumes lógicos que podem ser facilmente redimensionados, ao contrário das partições”.

2.2.4.2 Conectividade

De acordo com Somasundaram *et al.*(2010), a conexão entre o *host* e outro *host*, usuário, dispositivo periférico ou dispositivo de armazenamento é denominado conectividade. Na conexão entre um *host* e um dispositivo de armazenamento existem os componentes físicos e os componentes lógicos de conectividade.

Os componentes físicos de conectividade são os elementos de hardware: barramento, porta e cabo. O barramento constitui no caminho entre as partes de um computador, facilitando a transmissão dos dados entre elas, de acordo com Somasundaram *et al.*(2010).

Segundo Freire (2017, np), “o barramento [...] consiste em linhas de dados, linhas de endereço e linhas de controle”. Essas linhas conectam-se a todas as interfaces de periféricos, como impressora, teclado, mouse, disco, etc.

A porta é o dispositivo que disponibiliza a conectividade entre o host e os dispositivos externos. Pode ser definida como entrada ou saída, de acordo com Freire (2017). O cabo possibilita a conexão do host com dispositivos internos e externos, e é composto de cobre ou fibra óptica, conforme Somasundaram *et al.*(2010).

Os componentes lógicos de conectividade são os protocolos de comunicação entre o host e os dispositivos periféricos: o PCI (Peripheral Component Interconnect); e aos dispositivos de disco: IDE/ATA (Integrated Device Electronics/ Advanced Technology Attachment) e o SCSI (Small Computer System Interface). O PCI especifica a padronização com que as placas de rede e modems trocam informações com a CPU. O IDE/ATA é um protocolo de interface muito utilizado atualmente, e o SCSI é um protocolo mais utilizado em computadores high-end (Somasundaram *et al.*,2010).

3 O BACKUP E A RECUPERAÇÃO DE INFORMAÇÕES

Este capítulo introduz aos conceitos de guarda e recuperação de documentos. A seção 3.1 demonstra as características de backup. A seção 3.2 apresenta os objetivos para a guarda de informações. A seção 3.3 demonstra os tipos de backup. Complementando, a seção 3.4 apresenta estratégias de backup. E a seção 3.5 mostra o processo e operação de recuperação de informação e a importância destes para as empresas.

Segundo Somasundaram *et al.* (2010, p.273), “backup é a cópia dos dados de produção criada e mantida com o único propósito de recuperar dados apagados e corrompidos”. Atualmente, com o avanço das tecnologias e do mercado de negócios, as empresas necessitam cada vez mais realizar backup de um grande volume de dados, assim como a recuperação desses dados deve ser rápida e consistente.

Conforme a norma NBR ISO/IEC 17799, norma que estabelece as melhores práticas para segurança da informação, a criação de cópias de segurança está prevista: “[...] 10.5.1 Cópia de segurança das informações [...] Convém que as cópias de segurança das informações e dos softwares sejam efetuadas e testadas regularmente conforme a política de geração de cópias de segurança definida”.

De acordo com Løest (2009, pg.19), o conceito de backup é definido como “[...] a cópia de dados de um dispositivo para outro com o objetivo de posteriormente os recuperar no caso de perda de dados”.

Segundo Diaz (2011) o conceito de backup se resume a realizar uma cópia de segurança, para manter salvos informações importantes e arquivos pessoais protegidos, caso ocorra algum problema, como perda de dados, dados corrompidos ou exclusão de algum arquivo.

3.1. Características do Backup

De acordo com o descrito em Loest (2009, pg.19):

Um sistema de backup tradicional é normalmente composto por um servidor remoto (muitas vezes isolado e protegido contra fogo, intempéries e outros riscos físicos), acessível através da rede e com uma grande capacidade local de armazenamento de dados. A organização e manutenção do processo de backup é uma tarefa complexa.

De acordo com Barros (2007), o backup possui alguns elementos básicos que o compõem, conforme relacionado a seguir:

- Job: procedimento para a realização do backup, que só é executado com sucesso após a finalização do job;
- FileSet: definição das unidades, arquivos, diretórios, expressões regulares e partições que serão parte da rotina do backup;
- Schedule: é o agendamento da execução do job, permitindo a configuração do dia e horário que iniciará o procedimento;
- Pool: volumes utilizados pelo backup, que possuem características como onde será armazenada a cópia (dispositivo), período de reciclagem e tempo de retenção (tempo que será mantido);
- Catálogo: índices que facilitam a localização da cópia para fins de restauração. Contém informações como nome do volume que foi gravado, data de gravação, data de expiração, etc. Esse catálogo é armazenado em uma base de dados;
- Storage: dispositivo de armazenamento utilizado pelo backup, podendo ser uma unidade ou uma biblioteca de fitas, conhecido como tape library (fitoteca). Pode ser dispositivo de armazenamento externo ou interno;
- Período de retenção: tempo que os dados do backup permanecerão em storage. É variável e definido nas políticas de backup. Quanto maior o período de retenção, maior o custo de armazenamento e maior é o número de backups realizados. Quanto menor o período de retenção, menores custos, porém, menos backups disponíveis;
- Data de expiração: define quando os dados serão descartados, para a reutilização do dispositivo de armazenamento. Também é variável e definida nas políticas de backup. Normalmente, são excluídos os índices da base de dados, não necessariamente o próprio dado.

Conforme Loest (2009), as principais características de um sistema de backup são listadas a seguir:

- Desempenho: qual a performance do sistema de backup, levando em conta alguns quesitos, como tempo de cópia, tempo de recuperação e quantidade de informação a ser comportada, por exemplo;
- Integridade e consistência: o sistema deve garantir a integridade e consistência da informação, tanto ao realizar a guarda, quanto ao recuperá-la;

- Confidencialidade e privacidade: a confidencialidade dos dados deve ser mantida, assim como a privacidade dos seus usuários;
- Disponibilidade: os dados devem estar disponíveis a qualquer tempo, assim como o próprio sistema de backup deve ser resistente a falhas, estar disponível para os usuários.

De acordo com Calvano (2017), para que as premissas acima sejam acertadas, alguns fatores envolvidos no sistema de backup devem ser levados em conta. Tais fatores são observados a seguir:

- Recursos de infraestrutura: que devem ser analisados através da arquitetura, dimensionamento e utilização;
- Janelas de backup: que consiste no tempo de execução do backup. Sua definição deve levar em conta o ciclo da informação, a necessidade de sua cópia e o impacto da atividade de backup;
- Política de backup: é a estratégia e periodicidade do backup;

3.2. Objetivos do Backup

De acordo com Somasundaram *et al.* (2010, p.274) “backups são feitos por três motivos: recuperação de desastres, backup operacional e arquivamento”. Já, de acordo com Loest (2009), a sua utilização é justificada por duas situações: na restauração do sistema ao estado anterior a um desastre e na recuperação de alguma informação que foi apagada indevidamente ou corrompida.

Conforme observado por IBM (2012), os principais motivos do backup são:

- Recuperação de desastres: geralmente esse tipo de backup envolve a movimentação de fitas para outros locais. A estratégia de recuperação de desastres é feita com base nos requisitos de RPO (Recovery point object) e RTO (Recovery time object). Também pode ser utilizada a estratégia de replicação remota dos dados;
- Backup operacional: O backup operacional é geralmente utilizado para restabelecer a situação normal de funcionamento em caso de arquivos corrompidos, apagados inconsistentes etc. Exemplo: Backup de e-mail, Bancos de dados, arquivos de log, Rede Man, etc;

- Arquivamento: Geralmente utilizado para guardar informações por longos períodos, atender a legislação específica, guardar informações financeiras etc. Exemplo: arquivamento de e-mail, arquivamento do balanço da empresa, etc.

Conforme Barros (2007, pg. 22), em caso de recuperação de desastres, uma boa estratégia é a adoção de um site backup, que “[...] nada mais é que um local físico, distinto do ambiente real de produção em que se possa rapidamente restabelecer os serviços em caso de um desastre”. Alguns tipos de site backup, de acordo com Barros (2007), são listados a seguir:

- Cold Site: não possui a cópia dos dados do ambiente original e nem o hardware configurado, possui menor custo de montagem, porém o restabelecimento dos serviços demora muito tempo;
- Warm Site: possui hardware configurado como o ambiente original, porém não possui a cópia dos dados;
- Hot Site: possui a cópia dos dados do ambiente original e o hardware configurado, possui maior custo de montagem, porém o restabelecimento dos serviços é rápido.

O procedimento de backup é importante no caso de recuperação de desastres para restaurar os dados em outro local, quando o local onde estavam localizadas as informações está com o acesso inviável devido a algum desastre. Algumas empresas ainda utilizam o backup para realizar o arquivamento de suas informações. Normalmente, o fazem para manter uma cópia de transações, mensagens de e-mail e outros arquivos considerados necessários para a organização (Somasundaram *et al.*, 2010).

3.3. Tipos de Backup

Calvano (2017, np), resume os tipos de backup levando em conta a política de backup utilizada. Segundo o autor, “[...] o backup pode ser completo quando todos os dados são salvos, diferencial quando somente os dados que mudaram depois do último backup são salvos ou incremental quando os dados modificados entre o último backup completo e cada incremental são salvos”.

De acordo com Barros (2007), os tipos de backup podem ser divididos em:

- Completo, normal ou full: copia todos os dados definidos na política de backup para o dispositivo de armazenamento. Segundo IBM (2012), o backup completo é feito

sobre os dados integrais nos volumes de produção em um determinado momento. A sua vantagem é que, na recuperação, os dados estão em uma mesma mídia, o que torna a operação mais rápida. Porém, dependendo do tamanho da área a ser guardada, o backup full pode se tornar lento e custoso;

- Cópia: como o backup completo, porém não é feito uma marcação nos dados como se eles tivessem passado por um backup.
- Incremental ou cumulativo: copia todos os dados adicionais que não estavam do último backup realizado, seja completo ou incremental. De acordo com IBM (2012), o backup incremental guarda os dados que foram alterados a partir do último backup completo ou incremental, o que tiver ocorrido por último. Diferente do backup completo, o backup incremental é mais rápido visto que copia apenas os últimos dados modificados. Já, em caso de recuperação os dados podem estar espalhados por mais de um tipo de mídia, o que tornaria a operação lenta. Nesse caso para recuperação seriam necessárias a mídia do último backup full e todos os backups incrementais seguintes.
- Diferencial: copia todos os dados adicionais que não estava no último backup completo realizado. A vantagem desse tipo de backup, conforme descrito em IBM (2012), é que é um misto de completo e incremental, ou seja, em caso de recuperação seriam utilizadas a mídia do backup completo mais a mídia do último backup diferencial. Porém, em ambientes com elevado número de alterações, com o passar dos dias o backup diferencial equivale ao backup completo.

3.4. Estratégia e Processo de Backup

Conforme IBM (2012, np) “A quantidade de dados perdidos e o tempo inativo de um recurso são as preocupações primárias na seleção e implementação de uma estratégia específica de backup”.

De acordo com Barros (2007), conforme os tipos de backup são definidos algumas estratégias, listadas abaixo:

- Backup centralizado: existe um sistema central, composto de software, base de dados e dispositivo de storage, que gerencia o backup de todos sistemas que estejam em uma rede. Esse gerenciamento objetiva facilitar a administração do backup, e reduzir

os custos com unidades de armazenamento e mídias, visto que reduz as unidades utilizadas;

- Backup distribuído: no caso de impossibilidade ou inviabilidade do backup centralizado, pode-se realizar o backup distribuído, ou seja, dos sistemas envolvidos de forma individual. Cada sistema utilizará o seu próprio gerenciamento de backup, com dispositivos de armazenamento próprios. Essa estratégia encarece o processo, pois são utilizadas unidades de armazenamento e mídias para cada sistema. Além disso, se o sistema for proprietário, ainda existe o custo com licenças de sistemas de backup;
- Risco: caso ocorra uma indisponibilidade do sistema de backup ou um dispositivo de armazenamento é perdido, ocorre um prejuízo que afeta toda a empresa.

Ao administrar um backup corporativo, conforme ressalva Calvano (2017), deve-se levar em conta que se trata de um processo com estratégias dinâmicas, visto que a quantidade de informação aumenta relativamente e a busca por ela aumenta na mesma proporção.

Concluindo, os principais desafios dos ambientes de backup são, conforme IBM (2012):

- Cada vez mais dados são criados e guardados;
- Custo tem que ser baixo;
- Recursos mínimos;
- Restauração rápida e segura.

A figura 3.1, a seguir, demonstra a arquitetura e o processo de backup.

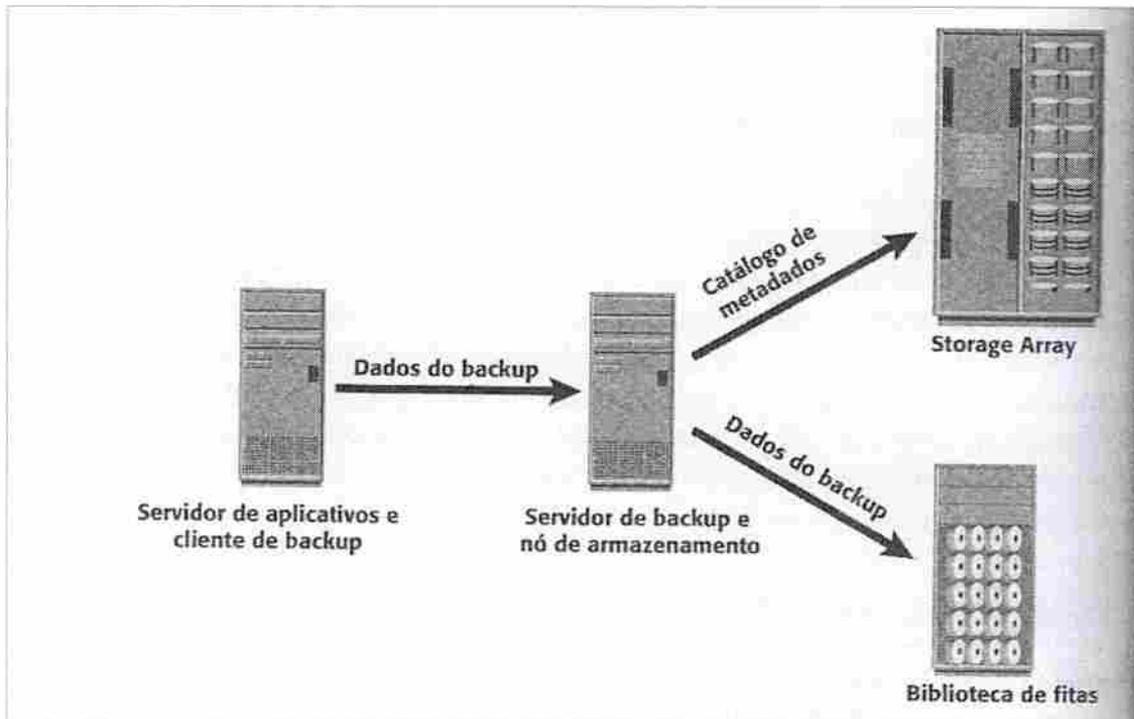


Figura 3.1: Arquitetura e processo de backup (Somasundaram *et al.*, 2010).

Como resumido em IBM (2012), o sistema de backup utiliza a arquitetura Cliente/Servidor. Um cliente de backup utiliza um agente específico para coletar as informações de backup e enviá-los ao servidor. Já o servidor armazena os dados e os metadados em um catálogo de backup conforme as políticas de backup. Geralmente a conexão é iniciada pelo servidor conforme agendamento.

As principais tecnologias para armazenamento de backup são a seguir descritas por IBM (2012):

- Backup em fita: as fitas são consideradas tecnologia de baixo custo e alta capacidade de armazenamento e uma boa opção para armazenamento dos backups por longos períodos. Em contraponto ao baixo custo o seu acesso é sequencial e a montagem de um cartucho também é uma operação lenta. Ainda há a desvantagem da leitura de toda a fita para se obter uma pequena parte da informação e a fragmentação dos dados, diminuindo a capacidade útil do dispositivo;
- Biblioteca de fitas: Dispositivo de armazenamento que contém uma ou mais unidades de fita, certa quantidade de slots para guardar cartuchos de fitas magnéticas, leitor de código de barras para identificar cartuchos de fita e um método automatizado para inserir as fitas na unidade de leitura (um robô) e drives que efetuam a leitura das fitas. A biblioteca de fita física, ou simplesmente fitoteca, fornece hospedagem e energia

para os drives e cartuchos de fita magnética, junto com um braço robótico ou mecanismo de captura. Os softwares de backup possuem inteligência para gerenciar o robô e todo o processo de backup ou restauração. Os drives de fita são responsáveis pela leitura e gravação dos dados nos cartuchos. Quando um processo de backup ou restauração é iniciado, o braço do robô é instruído para carregar uma fita em um drive. Esse processo aumenta o atraso dessas operações. Dependendo do tipo de hardware a montagem de um cartucho varia de 5 a 10 segundos, porém, a leitura ou escrita do dado solicitado, o retorno da fita para a posição inicial e a gravação da informação em buffer pode levar minutos. Para se ter uma ideia, o acesso a uma informação em memória leva até 60 ns (sessenta bilionésimos de segundo). Em disco o acesso demora por volta de 15ms (quinze milissegundos);

- Backup em disco: A tecnologia de backup em disco é utilizada para armazenar os dados com maior probabilidade de recuperação, criticidade da informação armazenada e menor tempo de guarda. É um dispositivo com acesso muito mais rápido que as fitas, porém o custo é mais elevado. Atualmente, com a popularização e evolução dessas tecnologias existem mídias com valor bem acessível, como exemplo os discos SATA, o que justifica os custos da manutenção de backups em disco. Esse tipo de tecnologia é considerado como dispositivo primário de armazenamento de backup;
- Cartuchos LTO (Linear Tape-Open): Esses cartuchos possuem grande capacidade de armazenamento, ideal para a guarda de grandes quantidades de dados e por longos períodos, porém, o acesso aos dados é mais lento, o que faz com que as fitas sejam utilizadas como dispositivo secundário de armazenamento de backup.

3.5. A recuperação de Informações

A recuperação do backup, ou, como é chamado, *recovery*, envolve a recuperação de informações que foram armazenadas. Caso a informação atual esteja danificada, ou o dado tenha sido apagado, é o *restore* (restauração) que possibilita essa recuperação (Barros, 2007).

Quanto aos termos *recovery* (recuperar) e *restore* (restaurar), ambos se referem a recuperação de arquivos a partir de um dispositivo de backup, como um disco rígido, ou outros meios de armazenamento. No entanto, a recuperação de arquivos normalmente se refere a salvar um ou mais arquivos, enquanto que uma restauração geralmente refere-se à

substituição de um sistema completo ou disco rígido a partir de um backup completo do sistema (IBM, 2012).

Conforme IBM (2006), o restore permite aos usuários recuperarem todos os dados que têm sido guardados anteriormente através de backup. E, também, restaurar versões mais antigas de dados perdidos, excluídos ou corrompidos.

De acordo com Barros (2007), mesmo quando não é solicitada, é importante realizar rotinas de recuperação de tempos em tempos, pois, as mídias de armazenamento também possuem um tempo de vida útil. Para esse procedimento adota-se a nomenclatura de testes de integridade. A adoção desses testes depende da política de backup e restore adotado pela empresa.

Algumas considerações sobre recuperação de dados devem ser analisadas, conforme IBM (2012). O RPO e o RTO são considerações importantes ao se planejar uma estratégia de backup. O RPO define o limite tolerável de perda de dados e especifica o intervalo entre dois backups, e esses podem ser diário, semanal, mensal etc. O período de retenção desses backups também são requisitos do RPO.

A empresa pode reter alguns backups por um longo período de tempo a fim de atender as legislações específicas. A maioria dos usuários também pode optar por reter seus backups pelo tempo máximo disponível na estratégia de backup. Porém, a guarda de muitos dados por longos períodos pode demandar a estrutura de armazenamento gerando assim um custo elevado. Sendo assim, é importante definir uma estratégia de retenção que combine um tempo racional de armazenamento com o orçamento disponível para tal (IBM, 2012).

Já O RTO refere-se ao tempo gasto na recuperação. E isso depende da estratégia e granularidade de backup adotada. A recuperação do backup full, por estar provavelmente em uma mesma mídia de fita, será mais rápida. As empresas costumam guardar mais backups full que o necessário. A combinação de backups incrementais e diferenciais costuma atender as demandas em um tempo baixo e custo razoável, porém são mais complexas que a recuperação de um backup completo, visto que podem utilizar mais de uma mídia durante o processo (IBM, 2012).

Outra situação a ser pensada é a certeza de que a versão restaurada não sobrescreva a versão atual, causando uma perda irreparável dos dados. Como o procedimento de restauração é solicitado com base em uma necessidade de informações que não estão disponíveis em produção, por diversos motivos, é importante confirmar com o solicitante se a versão atual pode ser descartada ou não (Barros, 2007).

Assim que for definido se a restauração irá ou não sobrescrever o arquivo atual, é importante verificar as características da informação que se deseja, tais como versão, tipo, tamanho, administrador, ente outras. Após essa verificação, localizar no catálogo o volume onde estão os dados desejados. A partir daí o sistema de backup disponibilizará o volume para o restore (Barros, 2007).

Conforme IBM (2006), para recuperar um arquivo, um diretório, ou até mesmo uma máquina inteira, é necessário saber duas coisas: o que se deseja recuperar (nome do arquivo, diretório, etc) e, opcionalmente, a partir de quando, se deseja restaurar um objeto que não seja o mais recente. Não é necessário saber o que os dados realmente são.

Segundo Barros (2007), depois de localizado o volume e disponibilizado para a restauração, o próximo passo é verificar se o acesso aos dados será sequencial ou não, o que é definido pelo tipo de mídia e tamanho dos dados. Após iniciado o processo de restore, os seguintes passos são realizados:

1. O software de backup envia a solicitação para a unidade de armazenamento de backup;
2. É realizada a leitura do volume / mídia onde estão os dados;
3. Os dados são copiados para uma determinada área do disco, seja uma área temporária, definida previamente, ou então no local de origem, sobrescrevendo o dado que lá estava.

Para garantir a integridade dos dados restaurados, visto que eles passam por diversos componentes (software para gerenciamento de backup, mídia de armazenamento, hardware para a gravação do backup, cabeamento de rede, etc), é gerado um resumo de mensagem dos dados (hash). Esse resumo permite a verificação da assinatura e do status da recuperação. Existem alguns algoritmos para a geração de resumo de mensagem, tais como: Message Digest 5 (MD5) e Secure Hash Algorithm (SHA-1) (Barros, 2007).

Para entender o processo de recuperação, é importante rever a operação de backup. Quando ocorre o processo de backup, é realizada uma grande comunicação entre todos os componentes que fazem parte da estrutura de backup. O servidor de backup inicia a conexão com os clientes, de acordo com a política de backup de cada um deles (Somasundaram et. al., 2010).

A figura 3.2 ilustra e explica o processo de operação de backup.

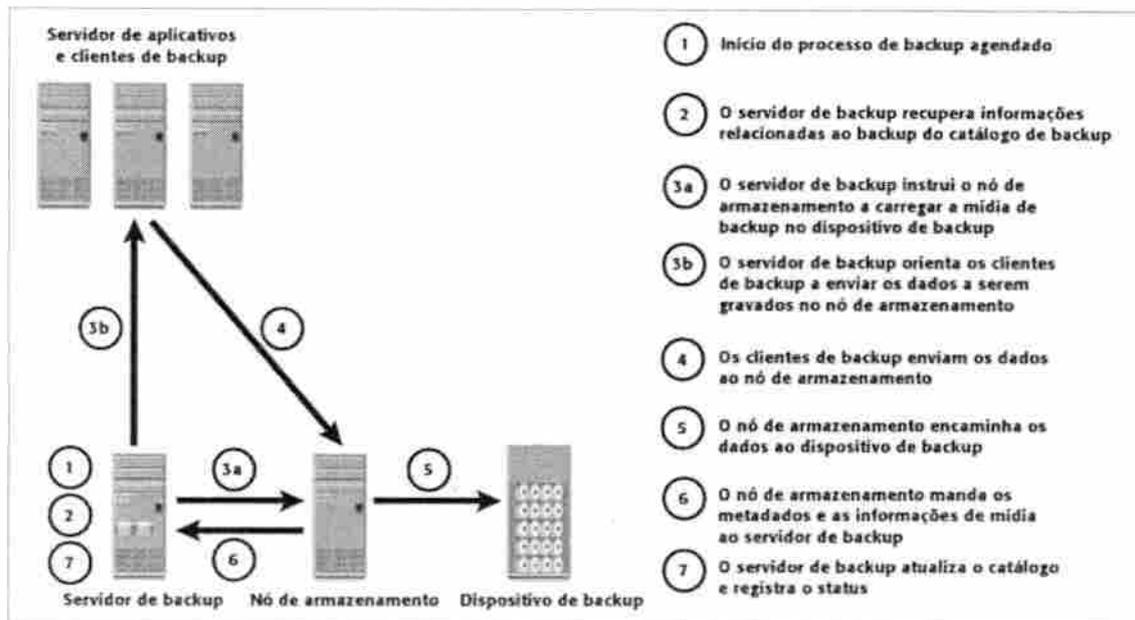


Figura 3.2: Operação de backup (Somasundaram *et al.*, 2010).

Conforme pode ser observado, é o servidor de backup quem realiza a coordenação do processo de backup, envolvendo todos os componentes da configuração. Após o processo, é ele quem atualiza o catálogo de backup com as informações sobre o processo (Somasundaram *et. al.*, 2010).

Depois de realizado o processo de backup, os dados já estão disponíveis para restauração sempre que necessário. Conforme Somasundaram *et al.* (2010, p.283), “um processo de restauração deve ser iniciado manualmente. Alguns produtos de software de backup têm um aplicativo separado para operações de restauração. Os aplicativos de restauração são acessíveis apenas para os administradores”. A figura 3.3 ilustra e explica o processo de operação de restauração.

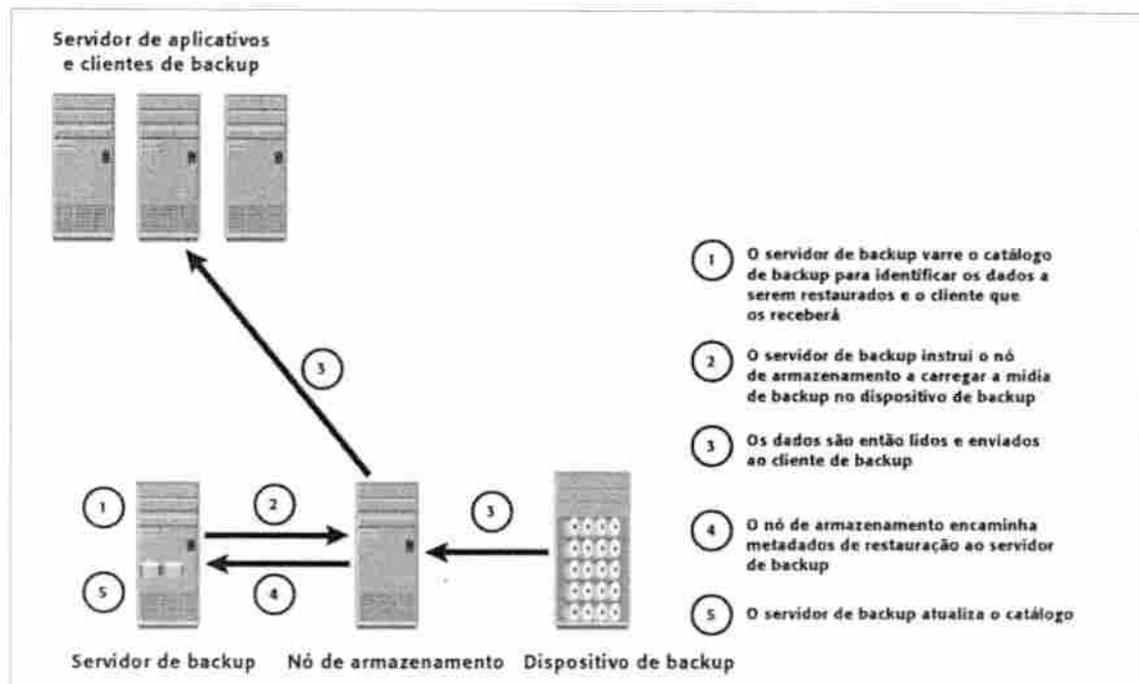


Figura 3.3: Operação de restauração (Somasundaram *et al.*, 2010).

O administrador, após selecionar os dados a serem restaurados, dá início ao processo de restauração. O servidor de backup acessa a mídia onde estão os dados e envia para a montagem no dispositivo de backup. Os dados são lidos e gravado no cliente que fez a solicitação. Após, o catálogo de metadados é atualizado pelo servidor de backup (Somasundaram *et. al.*, 2010).

4 O MAINFRAME

Este capítulo detalha o ambiente mainframe, amplamente utilizado pelas empresas do ramo financeiro e principal ambiente do Banco XYZ. A seção 4.1 apresenta as características do mainframe e a seção 4.2 demonstra as vantagens da utilização deste ambiente.

O mainframe é definido como um computador de grande porte e é responsável pelo processamento de um grande volume de informações. Atualmente, é muito utilizado em ambientes corporativos, para fins comerciais e econômicos. Realizam operações em grande velocidade sobre um grande volume de dados (IBM,2017).

Segundo Oliveira *et al.* (2010, prefácio):

Os computadores de grande porte, comumente conhecidos por mainframes já tiveram a sua Era Áurea. Foi numa época não muito distante, onde estes dominavam o mundo do processamento de informação, acerca de três décadas, quando falar em computadores era o mesmo do que falar em mainframes, pois estes eram os únicos da sua espécie.

Já, conforme Laundon (2001, p.108) “[...] é o maior computador, uma casa de força com grande memória e com poder de processamento extremamente rápido. É usado em grandes aplicações empresariais, científicas ou militares [...]”, visto que é necessário movimentar um grande volume de dados e realizar funções muito complexas.

Os mainframes são fabricados pela IBM, desde o princípio dos anos 40, e, com o lançamento do System/360 em 1964, iniciou-se um mercado em constante evolução, sendo acrescentado de outros fabricantes, como a HP e a Burroughs (atual Unisys). Diversos ramos de negócio começaram a utilizar a tecnologia mainframe, sendo mais utilizada para a transmissão, leitura e gravação de dados (Oliveira *et al.*, 2010).

Naquela época, apenas algumas empresas podiam pagar pelos mainframes. A arquitetura S/360 (ou System/360) é a arquitetura padrão que viria a ser utilizada em todos os modelos desenvolvidos pela IBM. Segundo Oliveira *et al.* (2010, p.4), algumas das inovações apresentadas foram: “[...] memória de tamanho variável, endereçamento relativo, multiprocessamento, subsistema de Input e Output”.

Ainda, conforme Oliveira *et al.*(2010), alguns modelos lançados pela IBM, da série S/360, são vistos a seguir:

- 1964 – S/360 Series Model 40;
- 1970 – S/370 Series Model 145;
- 1977 – S/370 Compatible Model 3031;

- 1979 – 4300 Processing Systems Model 4341;
- 1984 – 4300 Processing Systems Model 4381;
- 1985 – 3090 Processor Series Model 3090;
- 1990 – S/390 Series Model ES/900;
- 2000 – z/Series Model z900.
- 2003 – z/Series Model z990.

Na figura 4.1 abaixo, podemos atestar a evolução dos modelos de mainframes fabricados pela IBM. À esquerda, está o IBM 705, projetado em meados dos anos 1950 e apresentado como o servidor com o processamento de dados mais potente do mercado. À direita, observa-se o z990, introduzido em maio de 2003 como "servidor mais sofisticado do mundo" (IBM, 2017).

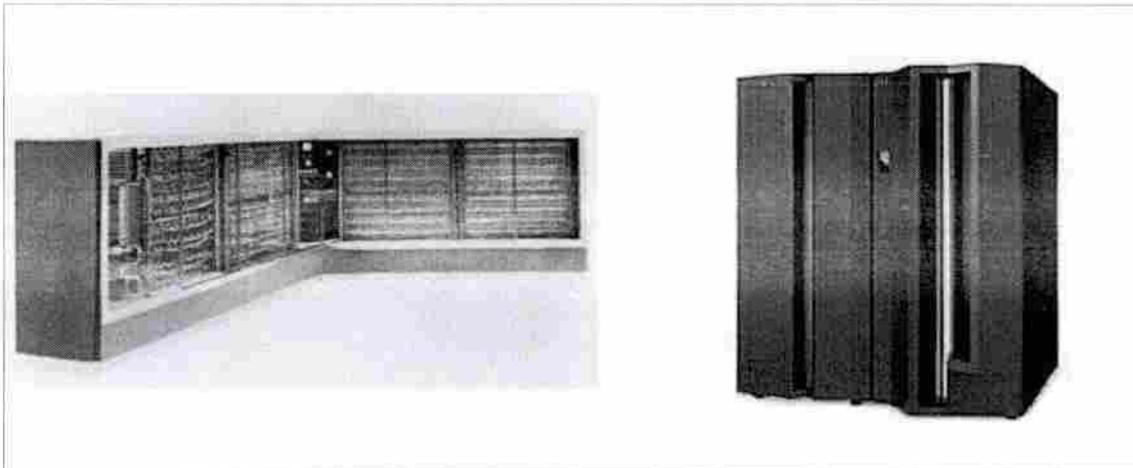


Figura 4.1: Evolução dos modelos de mainframes (IBM – *Mainframes photo album*, 2017).

Segundo IBM (2017), o nome S/360 refere-se ao conceito de 360° para cobrir todo o círculo de possíveis arquiteturas e utilizações. O S/360 também foi o primeiro desses computadores a usar microcódigo para implementar muitas das suas instruções de máquina, em vez de ter todas as suas instruções de máquina em seu circuito. Microcódigo (ou firmware, como às vezes é chamado) é constituído por micro instruções armazenadas, não disponíveis para os usuários, que fornecem uma camada funcional entre o hardware e o software. A vantagem de microcódigo é a flexibilidade, em que qualquer correção ou nova função pode ser implementada apenas mudando o microcódigo existente, em vez de substituir o computador.

4.1. Características do Mainframe

Os mainframes são computadores de maior porte, sendo assim, possuem um hardware semelhante ao de computadores menores (servidores e desktops). Memória, processadores e disco rígido possuem as mesmas funcionalidades, porém, em uma escala maior, definidos em sua arquitetura.

Conforme descrito em IBM (2017), uma arquitetura é um conjunto de termos definidos e regras que são utilizadas como instruções para construir produtos. Cada geração de computadores mainframe incluiu melhorias na arquitetura, mantendo-se mais estável, segura e compatível com todas as plataformas de computação. Os modelos mais recentes podem lidar com as cargas de trabalho de clientes mais avançados e exigentes e, no entanto, continuar a executar aplicações que foram escritas na década de 1970 ou anterior.

Para realizar a troca, retirada e/ou inclusão de processadores, a máquina não precisa ser necessariamente desligada. O mainframe possui um banco de processadores que, quando necessário, supre as atividades do processador que está sofrendo intervenção (Oliveira *et al.*, 2010).

O backup das informações constantes nos mainframes normalmente é executado nas fitas magnéticas. Os discos armazenam o sistema operacional. Assim como com os processadores, os discos também podem ser substituídos sem desligar o mainframe (IBM, 2017).

Conforme Oliveira *et al.* (2010), os mainframes possuem quatro principais sistemas operacionais, conforme listado a seguir:

- MVS (Multiple Virtual System): possui a carga de serviço dividida em várias partições, as Address Space. Essas partições não são fixas ou limitadas em sua quantidade e tamanho. Suporta processamento batch e conexão online com usuários. Em cada Address Space existe um programa denominado task, que, por sua vez, pode executar outros subprogramas em paralelo, denominados subtasks. O MVS evoluiu para o z/OS. Conforme Rouse (2005), MVS (e seu sucessor) pode ser considerado um "mundo" de sistema operacional, pois executa, por exemplo, folha de pagamento, contas a receber, processamento de transações, gerenciamento de banco de dados e outros programas essenciais para grandes empresas. Embora o MVS tem sido muitas vezes visto como um sistema de informação monolítico, centralmente controlado, a IBM nos últimos anos reposicionou-o como um "grande servidor" em um ambiente distribuído orientado à rede, utilizando um modelo de aplicação em 3 camadas. A

interface principal do usuário em sistemas MVS é o Time Sharing Option (TSO). O Interactive System Productivity Facility (ISPF) é um conjunto de menus para compilar e programas de gestão e para a configuração do sistema. O principal sistema de gestão do trabalho é o Job Entry Subsystem 2 ou 3 (ou JES2 JES3) (Rouse, 2005);

- TPF (Transaction Processing Facility): inicialmente, era um sistema operacional para atender a demanda de transações online de empresas aéreas, chamado de Airline Control Program (ACP), passando, com o tempo, a atender outros mercados, como empresas de cartão de crédito, redes de hotéis, bancos e redes de varejo, por exemplo. Através de uma rede de comunicação, localizada em uma extensa localização geográfica, altamente disponível, realiza um grande volume de transações (cerca de 10.000 transações por segundo), com rápido processamento. Evoluiu para o z/TPF. Conforme IBM (2017), z/TPF pode usar vários mainframes em um ambiente flexível para rotineiramente lidar com dezenas de milhares de transações por segundo enquanto experimenta disponibilidade ininterrupta que é medido em anos.
- VM (Virtual Machine): realiza a virtualização de operação e arquitetura de uma ou mais máquinas. Cria ambientes lógicos dentro de uma máquina física, cada qual com memória e unidades de I/O próprias. Executa qualquer sistema operacional, facilitando a execução de diversos aplicativos em diversas plataformas. A grande vantagem da VM é a racionalização de espaço físico e manutenção de máquinas e sistemas. Evoluiu para o z/VM. De acordo com IBM (2017), Como um programa de controle, o z/VM é um hypervisor porque ele é executado em outros sistemas operacionais das máquinas virtuais que cria. Possui dois componentes básicos: um programa de controle (PC) e um sistema operacional de um único usuário, Conversational Monitor System (CMS). O programa de controle cria artificialmente várias máquinas virtuais a partir dos recursos de hardware reais. Para usuários finais, ele aparece como se eles têm utilização dedicada dos recursos reais compartilhados. Os recursos compartilhados incluem impressoras, dispositivos de armazenamento de disco e da CPU. O programa de controle garante a segurança dos dados de aplicação entre os sistemas convidados. O hardware real pode ser compartilhado entre os convidados, ou dedicados a um único convidado por razões de desempenho. Para a maioria dos usuários, a utilização de sistemas dessa maneira evita a necessidade de configurações maiores de hardware.
- z/OS: sistema operacional organizado em elementos, onde cada um executa uma função específica no sistema. Essa organização facilita a manutenção, aumenta a

confiabilidade e evita a replicação de código. Cada unidade básica de programa do z/OS é um módulo de carga, que possui dados e instruções, e um ou mais pontos de entrada, objetivando a carga na memória e modo de execução do sistema. Possui um sistema de arquivos que o diferencia de outros sistemas operacionais como Windows e UNIX. No z/OS, os arquivos são chamados de datasets, e são orientados a registros e o sistema define várias organizações e formatos diferentes para eles. Nas demais plataformas, os arquivos são orientados a byte e não possuem formato específico. De acordo com Rouse (2005), o ambiente de desenvolvimento mais comum hoje em dia usa as linguagens C e C++, e utiliza banco de dados DB2, que é o principal banco de dados relacional da IBM. Aplicações Java podem ser desenvolvidas e executadas em ambientes UNIX OS/390.

Conforme Oliveira *et al.*(2010, p.8) “um mainframe consiste basicamente em uma memória central, uma ou mais unidades de processadores centrais, recursos de operador, um subsistema de canal e unidades de I/O”.

Podemos observar esse conjunto na figura 4.2 a seguir, e listados posteriormente.

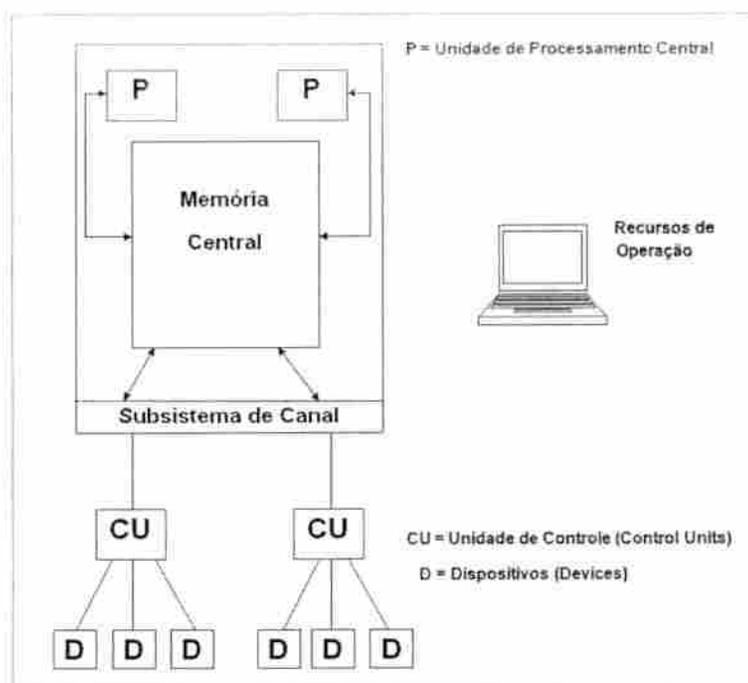


Figura 4.2: Principais componentes do mainframe Oliveira *et al.*(2010). Adaptado pela autora.

- Memória: o acesso e o processamento dos dados são rápidos, pois o acesso à memória é direto, tanto pelos processadores quanto pelo subsistema de canal. Nas unidades de I/O são armazenados as informações e os programas e, antes de serem processados,

são enviados para a memória central. O tamanho da memória é definido conforme o modelo do mainframe, e é dividido em blocos de 4k bytes. Também, de acordo com o modelo da máquina, além da memória central, existe uma memória expandida, que é acessada pelos processadores e realiza a comunicação com a memória central através de instruções especiais (IBM, 2017);

- Unidades de Processamento Central (CPU): componente considerado o principal no mainframe, pois é responsável por manter os recursos para sequenciar e processar informações, executar instruções, realizar ações de interrupção, exercer funções de tempo, entre outras funções inerentes ao mainframe. O processamento dos dados pode ser de forma paralela ou sequencial e as instruções processadas são classificadas em: geral, decimal, de ponto flutuante, de I/O e de controle. Utilizam a memória interna quando da execução de algumas instruções, memória esta que está na mesma estrutura física da memória central, mas não pode ser acessada por programas. Segundo IBM (2017), o processador central contém os processadores, memória, circuitos de controle, e interfaces para canais. Um canal fornece um caminho de dados e controle independente entre os dispositivos de I/O e memória. Os primeiros sistemas tiveram até 16 canais; as maiores máquinas de grande porte de hoje podem ter mais de 1000 canais;
- Referência Externa de Tempo (ETR): esse componente existe dependendo do modelo do mainframe. Consiste em um dispositivo de referência externa de tempo, sendo muito utilizado em situações onde várias máquinas estão conectadas e o ETR é utilizado para definir um tempo para elas. Como exemplo de ETR, podemos citar o Sysplex Timer (IBM, 2017);
- Subsistema de Canal: controla a transferência de dados entre as unidades de I/O e a memória central, diminuindo a carga de trabalho para os processadores. Assim, os processadores ficam livres para executar as instruções e o subsistema de canal executa a transferência dos dados, tudo sendo realizado simultaneamente. Para realizar a conexão com as unidades de I/O, o subsistema de canal utiliza os caminhos de canal (Channel Path) (IBM, 2017);
- Unidades de I/O e Unidades de Controle: as unidades de I/O são compostas por equipamentos responsáveis pela entrada e saída dos dados, tais como impressoras, unidades de fitas magnéticas e cartuchos, discos, terminais, entre outros. E as unidades de controle gerenciam as operações desses equipamentos, e provém os

recursos físicos e lógicos para esse gerenciamento. Existem canais para ligar unidades de controle. Uma unidade de controle contém a lógica para trabalhar com um tipo particular de dispositivo I/O. Por exemplo, uma unidade de controle para uma impressora teria muitos circuitos internos e lógica diferentes do que uma unidade de controle para uma unidade de fita. Algumas unidades de controle podem ter várias conexões de canal fornecendo vários caminhos para a unidade de controle e seus dispositivos (IBM, 2017);

- Recursos de Operação: responsáveis pelo controle da máquina, e, para realizar as funções de zerar, iniciar, parar, recomeçar, alterar e mostrar utiliza uma console de operação (IBM, 2017).

4.2. Vantagens do Mainframe

A capacidade de dividir um grande sistema em vários sistemas menores, chamados de partições lógicas (LPARs), é agora um requisito fundamental e uma grande vantagem em praticamente todas as instalações mainframe. A flexibilidade do desenho de hardware permite que qualquer processador acesse e aceite interrupções para qualquer canal, como a unidade de controle e o dispositivo ligado a uma certa LPAR, o que contribui para a flexibilidade, confiabilidade, e o desempenho do sistema completo (IBM, 2017).

De acordo com IBM (2017), as partições lógicas (LPARs) são, na prática, equivalentes a mainframes separados. Cada LPAR executa seu próprio sistema operacional, que pode ser qualquer sistema operacional de mainframe. Os dispositivos de I/O podem ser compartilhados através de vários LPARs, mas esta é uma decisão da empresa. O administrador do sistema pode atribuir um ou mais processadores de sistema para o uso exclusivo de uma LPAR.

Além da LPAR, de acordo com Oliveira *et al.*(2010), outras vantagens do mainframe são listadas a seguir:

- Capacidade: grande capacidade de armazenamento e processamento de grandes volumes de informações. Alguns exemplos de empresas que os utilizam são os bancos, empresas aéreas, administradoras de cartão de crédito e algumas grandes redes de varejo;
- Escalabilidade: capacidade de se adaptar a uma grande demanda de processamento, variando conforme a necessidade (diminuindo ou aumentando) a capacidade de

processamento e armazenamento. Essa variação é baseada no número de processadores, rede e memória disponíveis;

- **Integridade e Segurança:** compartilhar muitos dados, de forma simultânea, entre vários usuários, de forma segura;
- **Disponibilidade:** manter o sistema disponível o tempo todo, sendo essa a sua maior prioridade. Essa disponibilidade baseia-se em três princípios:
 - **Confiança:** os componentes de hardware e software realizam auto testes e, caso necessário, se autocorrigem;
 - **Disponibilidade:** um componente que esteja com problema é isolado, a fim de evitar comprometer os demais componentes do sistema;
 - **Manutenção:** o componente com erro é identificado e logo trocado, de forma a reduzir o impacto sofrido pelo sistema todo.
- **Acesso à informação:** acessar uma grande quantidade de informação e conectar com vários meios de armazenamento (fitas, cartuchos, discos rígidos, etc.);
- **Gerenciamento de sistema:** acompanhar a performance e a capacidade do sistema, através de diversos mecanismos internos;
- **Processamento:** os mainframes possuem dois tipos de processamento, conforme pode ser visto na figura 4.3, a seguir.

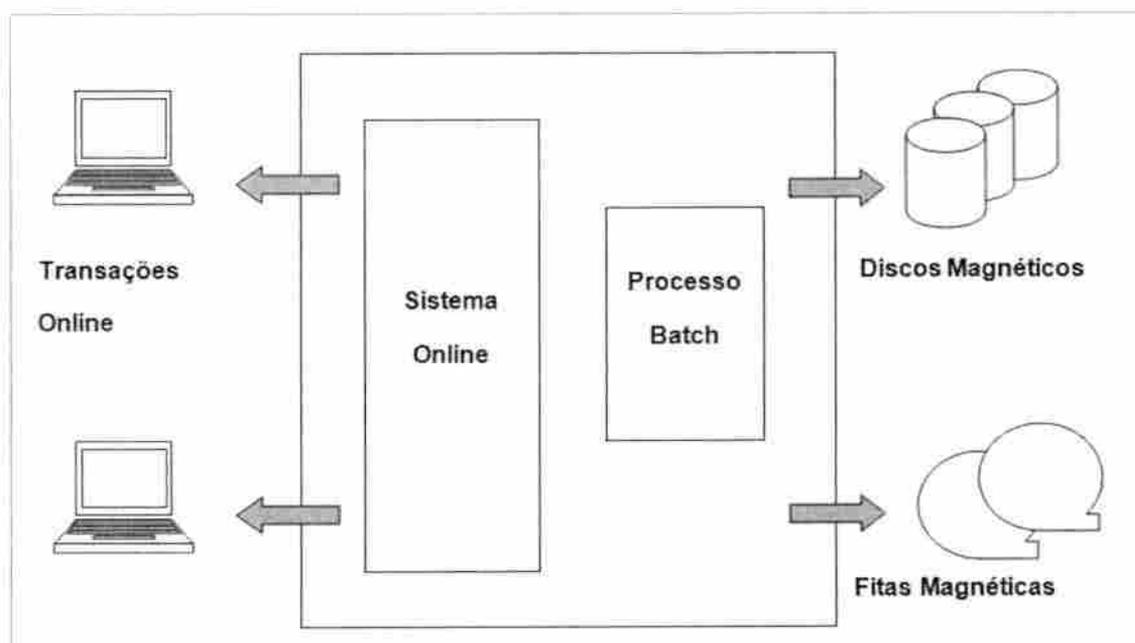


Figura 4.3: Processamentos online e batch (Oliveira *et al.*, 2010). Adaptado pela autora.

Conforme explica Laudon (2001):

- Batch: é um processamento sequencial, de um grande volume de informações, em lote, como, por exemplo, a folha de pagamento de uma grande empresa. Sua maior preocupação é com a quantidade de dados processados, do que com o tempo em que esses dados são processados e suas instruções executadas;
- Online: esse processamento é composto de transações rápidas e pequenas, individuais e com um volume pequeno de informações. Como por exemplo, podemos citar as transações com cartão de crédito. Nesse tipo, o que é levado em conta é seu tempo de processamento.

Cada nova geração de computadores mainframe incluiu melhorias em uma ou mais das seguintes áreas da arquitetura (IBM, 2017):

- Processadores mais rápidos;
- Mais memória física e maior capacidade de endereçamento de memória;
- Capacidade dinâmica para a atualização de hardware e software;
- Aumento da automação de verificação de erro de hardware e recuperação;
- Dispositivos avançados de I/O e canais mais rápidos de comunicação entre estes dispositivos e os processadores;
- Anexos de I/O mais sofisticados, tais como adaptadores de rede local com ampla transformação interior;
- Uma maior capacidade de dividir os recursos de uma máquina em vários sistemas, logicamente independentes e isolados, cada um executando seu próprio sistema operacional;
- Tecnologia de cluster avançada, como Parallel Sysplex, e capacidade de compartilhar dados entre vários sistemas.

5 A FERRAMENTA IBM – TIVOLI STORAGE MANAGER (TSM)

Este capítulo apresenta detalhadamente a ferramenta IBM TSM e sua aplicabilidade no ambiente do Banco XYZ. A seção 5.1 mostra as características e arquitetura da ferramenta TSM. A seção 5.2 demonstra quais as vantagens do uso da ferramenta no backup e recuperação de informações. A seção 5.3 apresenta o TSM no ambiente do Banco XYZ.

IBM Tivoli Storage Manager (TSM) consiste em um software de armazenamento de informações, aplicado em sistemas distribuídos heterogêneos. Armazena e gerencia uma grande quantidade de dados, a partir de estações de trabalho para um ambiente de servidores. Suporta mais de 44 sistemas operacionais, e sua utilização é feita a partir de uma interface gráfica de usuário (IBM, 2006).

De acordo com Sasso (2010), o TSM protege os dados de uma empresa contra falhas de hardware e outros erros de armazenamento, de backup e cópias de arquivos no armazenamento off-line. Pode ser configurado para prover serviços a uma grande variedade de computadores, desde laptops até mainframes, executando diversos sistemas operacionais, conectados através da Internet, em redes WANs, LANs ou SANs.

O TSM é um aplicativo de gerenciamento de armazenamento em rede para empresas. Ele fornece serviços automatizados de gerenciamento de armazenamento para estações de trabalho de vários fornecedores, computadores pessoais e servidores de arquivos de rede local (LAN). Inclui os seguintes componentes: servidor; cliente administrativo; cliente de backup; gestão hierárquica de armazenamento; interface de programação de aplicativos; entre outros (IBM, 2000).

Conforme Neduziak (2010), a ferramenta propõe uma solução flexível para o backup de informações em empresas de médio e grande porte, devido ao seu custo. Para seu gerenciamento, exige habilidades técnicas mais avançadas, porém, com o tempo, as suas políticas e funcionalidades facilitam sua aplicação e manutenção. É considerada pioneira por utilizar um banco de dados DB2 (sistema gerenciador de banco de dados) independente.

De acordo com Dorion (2007), o TSM é elogiado por uns como a ferramenta de backup mais flexível existente no mercado. Já, por outros, é criticado como o menos intuitivo. O Tivoli Storage Manager da IBM (TSM) já existe há muitos anos. Introduzido pela primeira vez no início dos anos 90 como ADSM (ADSTAR Distributed Storage Manager), o software foi construído sobre conceitos de gerenciamento de armazenamento de dados de mainframe.

IBM Tivoli Storage Manager é um sistema de backup e arquivamento de aplicação de classe empresarial. Como todos os produtos de software de backup corporativo, é projetado para fazer cópias de dados de uma organização e para proteger contra a perda destes dados. Semelhante a outras plataformas de software de backup de empresa, permite backup automatizado de dados, baseado em políticas (Rouse, 2013).

Exemplos de outros softwares de backup (fornecedor/ferramenta):

- IBM: Spectrum Protect
- EMC: Avamar
- HP: HP Data Protector

Basicamente: todos fazem a mesma coisa, pois descendem da teoria da organização de arquivos e estrutura de dados, assim, não há o que alterar, cabe a cada empresa analisar qual o mais adequado ao seu propósito.

5.1. Características do TSM

Segundo IBM (2006) o TSM fornece:

- Administração centralizada de dados e gerenciamento do armazenamento destes dados;
- Gerenciamento de recuperação de dados em caso de desastres;
- Proteção totalmente automatizada das informações;
- Gestão eficiente do crescimento de operação;
- Recuperação automatizada de servidores;
- Compatibilidade com vários dispositivos de armazenamento, assim como redes LAN, WAN e SAN;
- Soluções de backups personalizados para grandes grupos empresariais, empresas de Enterprise Resource Planning (ERP) e produtos de banco de dados.

Possui gestão centralizada, baseada na web, técnicas de movimentação e armazenamento de dados, e abrangente automação baseada em políticas de backup, que, juntas, trabalham para realizar a proteção dos dados e minimizar os custos de administração. Os módulos opcionais da ferramenta permitem que os aplicativos críticos de negócio possam ser executados em escala 24x7x365 (24 horas, 7 dias por semana, 365 dias por ano) e utilizarem o TSM para uma proteção centralizada dos dados, sem interrupção dos seus serviços (IBM, 2006).

De acordo com IBM (2006), as opções de backup e recuperação, fornecidas pelo TSM, são listadas a seguir:

- Imagem de instalação completa do sistema (mksysb) para ser instalada em outro sistema com configuração de hardware igual ou diferente (clonagem);
- Backup de um grupo de volumes (volume group – VG). Um VG é responsável por agrupar um ou mais discos ou partições;
- Backup de um sistema de arquivos (file system);
- Backup de um arquivo ou diretório;
- Backup de um volume lógico (logical volume – LV). Um LV é a partição lógica de um volume físico;
- Recuperação de todo ou parte de um sistema.

O TSM possui uma administração central e ferramentas de automação, tais como utilitários para criação de scripts de backup e horários definidos para automação de tarefas de mais simples. As operações de backup são listadas e verificadas através de um log de registro, com o status de concluído. Além disso, recursos instalados no cliente permite que o administrador gerencie as operações de backup de forma centralizada (local ou remotamente) a partir de um único servidor (IBM, 2006).

Segundo IBM (2006), o armazenamento de dados é baseado em políticas de gerenciamento, que auxiliam as empresas a atender as exigências regulamentares de várias agências governamentais e da indústria. Alguns novos regulamentos exigem garantias adicionais sobre a retenção de dados. O TSM fornece políticas de retenção de dados que ajudam a atender a essas novas regras.

Conforme Sasso (2010), as políticas de backup definidas pelo TSM possuem as seguintes definições:

- Policy domain: é a mais alta definição na hierarquia da política de backup. Permite que os clientes sejam agrupados por políticas comuns, como por exemplo, separação por sistema operacional. Uma policy domain contém uma ou mais policy set, mas apenas uma delas pode ser ativada a cada momento (chamada active). O servidor usa apenas a policy set active para administrar os backups das máquinas associados àquela policy domain;
- Policy set: especifica cada management class disponível. A policy set contém uma ou mais management classes, mas apenas uma deve ser definida com default. Apenas um policy set, a active, controle as operações de política de domínio;

- Management class: associa grupos de arquivos de backup e archive e especifica se e como os arquivos são migrados para as áreas de armazenamento. Um management class pode conter um backup copy group e um archive copy group, ambos ou nenhum copy group;
- Backup copy group: durante o backup, controla o processamento dos arquivos associados com a management class. Determina os seguintes itens:
 - Com que frequência um arquivo pode passar pelo backup;
 - Como agir com arquivos que estão em uso durante o backup;
 - Onde o servidor armazena inicialmente as versões de backup de arquivos e diretórios;
 - Quantas versões do backup o servidor armazena;
 - Por quanto tempo o servidor armazena versões do backup;
- Archive copy group: controla o processamento do arquivamento de arquivos associados com a management class. Determina os seguintes itens:
 - Como agir com arquivos que estão em uso durante o archive;
 - Onde o servidor armazena os arquivos que passaram pelo archive;
 - Por quanto tempo o servidor armazena os arquivos arquivados.

O TSM pode gerenciar dados de backup e arquivamento através de arrays de disco, bibliotecas de fitas ou de armazenamento óptico e pode mover os dados automaticamente para dispositivos de armazenamento de menor desempenho. À medida que vida útil da informação acaba. Esta abordagem é chamada de armazenamento em camadas ou gerenciamento de armazenamento hierárquico. As políticas podem especificar o tempo de retenção, o tipo de mídia de armazenamento, a frequência de backup, tipo de dados e hardware de backup (Rouse, 2013).

Conforme IBM (2006), as características do TSM para armazenamento de informações são listadas a seguir:

- Políticas de validade: os dados expiram quando não mais necessários liberando a mídia de armazenamento e fornecendo uma relação eficiente com os custos de armazenamento;
- Proteção de dados off-site: é padrão, cópias off-site podem ser criadas em qualquer das mídias disponíveis e a política de validade é igual à original;

- Programa cliente de arquivo: permite que os usuários utilizem arquivos a partir de suas estações de trabalho ou de servidores de dados com as informações armazenadas, além de recuperar dados para suas estações de trabalho;
- Suspensão de exclusão e expiração: permite que você coloque uma espera incondicional nos dados. Isso significa que os dados não podem ser apagados ou modificados até que a suspensão seja liberada;
- Gerenciamento de retenção baseada em eventos: o tempo em que os dados são retidos é baseado no cálculo de um intervalo de tempo que inicia após o evento de retenção ocorrer. Os dados não podem ser excluídos até que o tempo limite expire. Por exemplo, você pode manter os dados de um funcionário específico por até um ano após a saída deste da empresa;
- Proteção de retenção de dados: os dados não serão apagados até que os critérios de retenção sejam satisfeitos.

Conforme Rouse (2013), o TSM usa um esquema de backup para guardar dados, chamado de incremento progressivo, ou "incremental para sempre". O software tem um backup completo inicial e todos os backups subsequentes são incrementais - isto é, só são copiados os dados que sofrem alguma alteração. Este conceito era exclusivo para a IBM, mas se tornou mais comum ao longo do tempo, especialmente entre os fornecedores de backup em nuvem. Além do backup incremental, o TSM apresenta o backup do tipo archive, que simplesmente guarda (arquiva) o dado por um tempo definido e o remove no final

5.1.1 Arquitetura do TSM

TSM é implementado como uma aplicação cliente-servidor, que consiste de um componente do TSM com software servidor, o componente do TSM com software cliente de backup e outros componentes (IBM, 2006).

O servidor de TSM constrói os dados de gerenciamento via backbone, conforme listado a seguir, de acordo com IBM (2006):

- Gerenciado o hardware de armazenamento;
- Providenciando um ambiente seguro;
- Fornecendo funções de monitoramento, automação e geração de relatórios;
- Implementando políticas de gerenciamento de armazenamento;
- Armazenando todos os dados do inventário na base de dados do TSM;

- Fornecendo uma única interface para gerenciar múltiplos servidores de TSM;
- Gerenciando as mídias e planos em caso de recuperação de desastres.

A figura 5.1 a seguir ilustra a arquitetura do TSM.

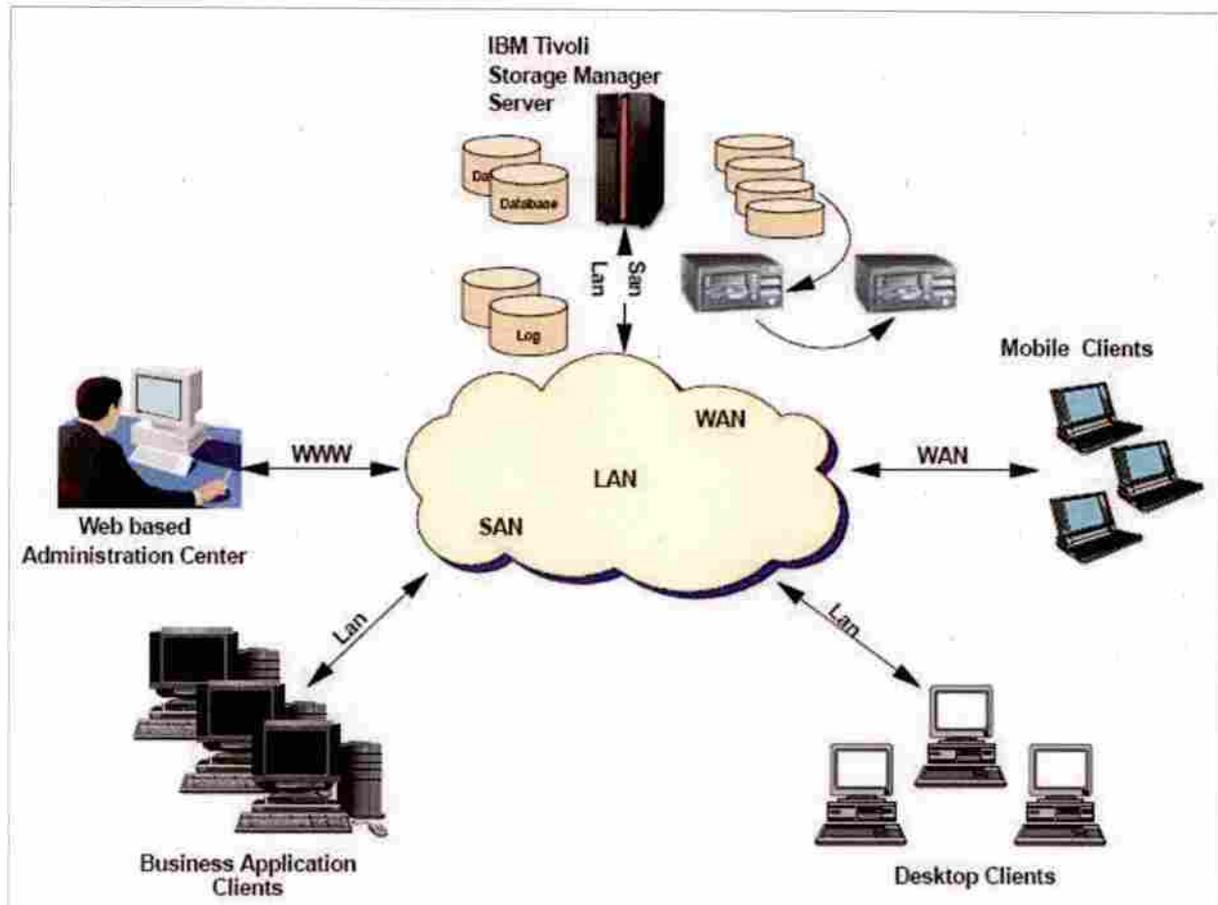


Figura 5.1: Arquitetura do TSM (IBM, 2006).

O TSM client e os demais componentes da arquitetura possuem todas as funções de gerenciamento de dados, como backup e recuperação de informações, arquivamento e gerenciamento hierárquico do espaço (Rouse, 2013).

Ainda, de acordo com Sasso (2010), o TSM server possui schedules administrativas, que consistem em processos destinados à manutenção automatizada e segurança. As mais utilizadas são: backup do Banco de Dados (backup DB); auditoria dos dados armazenados pelo servidor e suas licenças (audit lic); migração dos dados de backup guardados em disco para fita; exclusão definitiva dos dados marcados para este fim (expire inventory) e processo de reciclagem das fitas com áreas fragmentadas (reclamation).

5.2. Vantagens do TSM

Tivoli Storage Manager pode ser configurado de muitas maneiras diferentes para atender às necessidades de uma organização. É mais comumente implantado em empresas com grandes quantidades de dados de backup e uma equipe para gerenciá-lo. IBM TSM é compatível com uma variedade de hardware, incluindo AIX, HP-UX, Linux, Solaris e Windows Server. Clientes de backup do TSM estão disponíveis para versões mais compatíveis dos principais sistemas operacionais (Rouse, 2013).

Conforme descrito por Dorion (2007), as maiores diferenças entre o TSM e os backups tradicionais são listados a seguir:

- Backups incrementais progressivos: uma vez que é realizado o backup de um dado, arquivo ou sistema, ele só será acessado novamente pelo TSM caso sofra alguma alteração. A primeira vez que é feito o backup de um servidor de arquivos, todos os arquivos são copiados, uma vez que nunca foram copiados para o TSM. Daquele ponto em diante, somente os arquivos que foram alterados são, posteriormente, submetidos a backup. O objetivo principal consiste em reduzir a quantidade de dados que são transferidos através da rede;

Versões de backup: para manter informações específicas, o TSM usa a opção de versões (versionamento) de backup. Ao contrário dos produtos de backup completo/incrementais, o TSM não retém (ou expira) backups com base em jobs ou meios de comunicação, mas sim sobre as versões de arquivos. Para cada arquivo de backup, há uma versão ativa (arquivo tal como ele existe atualmente em um sistema) e versão inativa (arquivo como era antes de ser feito o backup novamente).

- O TSM manterá um número definido pelo usuário de versões para um determinado período de tempo;
- Arquivos do TSM: os usuários podem agrupar arquivos em pacotes que carregam um “carimbo” de tempo e descrição. Estes pacotes são conhecidos como arquivos de terminologia TSM e são apagados após um período definido pelo usuário de tempo (isto é, um, cinco, sete anos, etc.) Cada pacote de arquivo é considerado único e conservada por um período específico de tempo, independentemente de cópias previamente armazenados ou versões idênticas em TSM. Estes arquivos são independentes de backups;

- Hierarquia do conjunto de armazenamento: armazena os objetos de backup em pools de armazenamento. Estes estão associados a classes específicas de dispositivos (disco, fita, ópticos). Os pools de armazenamento podem usar dispositivos de acesso aleatório (disco) ou dispositivos de acesso sequencial, tais como fitas, óptico, arquivo, etc. Pools de armazenamento primários são organizados em uma hierarquia, o que significa que um pool de armazenamento pode apontar para outro conjunto como "next pool de armazenamento" permitindo que os dados sejam migrados automaticamente quando um limite de capacidade definida pelo usuário é atingido. A configuração mais comum consiste em um conjunto de discos apontando para um conjunto de fitas de backup de dados, onde são inicialmente armazenados em disco e mais tarde migram para a fita. TSM também suporta a associação de um "servidor" com um pool de armazenamento, ou seja, dados de backup pode ser migrados de um dispositivo de armazenamento local para um servidor TSM remoto;
- Recuperação de mídia de fita: como os objetos de backup são expirados em de backups inteiros, acaba criando "buracos lógicos" em um volume de fita. Uma vez que o limite de espaço livre lógico é atingido em certos volumes de fita, os dados válidos restantes são consolidados em uma fita que tenha espaço suficiente. A fita a partir do qual os dados foram movidos pode então ser recuperada para utilização;
- Interface de linha de comando: o TSM fornece uma interface de linha de comando completo (CLI - Command Line Interface) que permite que sejam realizados scripts de gestão, backup, restauração, monitoramento e operação de relatórios. A CLI está disponível no servidor e no cliente TSM;
- Backups seletivos: o TSM oferece um backup seletivo que permite que sejam selecionados os arquivos que se deseja, mesmo que eles não tenham sido alterados desde o último backup incremental.

O uso do TSM permite reduzir o custo total com propriedades de gerenciamento de armazenamento nas seguintes áreas, conforme IBM (2006):

- Administração: economiza o tempo do administrador de armazenamento para manter o controle dos arquivos de backup enquanto eles são movidos de uma mídia a outra durante a recuperação, esse procedimento é realizado de forma automatizada pelo TSM. Permite que vários servidores TSM sejam gerenciados a partir de um único ponto, uma única estação de trabalho, por exemplo;

- Fitotecas e uso de mídias: o TSM utiliza várias técnicas de redução das mídias de armazenamento, incluindo um banco de dados embutido. Utiliza metodologia progressiva de backup, colocação e recuperação de dados de forma a consolidar mais informações na mesma mídia. Assim, menos fitas são necessárias para suportar uma determinada quantidade de backup e dados arquivados, reduzindo os custos com meios de armazenamento. Aumenta a eficiência com a utilização das fitotecas, aumentando o espaço para outros volumes de trabalho;
- Tempo do operador: como são utilizadas menos mídias de armazenamento, são menos volumes que devem ser verificados pelo operador. Quando o TSM detecta um erro de leitura na mídia, ao copiar ou realizar backup de dados, ele solicita automaticamente um backup desses dados e realiza o arquivamento, o operador não precisa procurar o local da cópia de segurança. Além disso, o TSM mantém um registro de todos os backups realizado, volume, local, datas, etc, não sendo necessário ter esse controle manualmente;
- Migração e/ou revezamento de mídia: o TSM utiliza seu módulo de HSM (Hierarchical Storage Management) para migrar dados de um tipo de mídia para outro (por exemplo, de um disco rígido para uma fita magnética). Utilizando esse recurso, pode facilmente mover dados de volumes antigos para novos, do mesmo tipo, eliminando a necessidade do administrador de ter que rastrear volumes e arquivos para esse fim;
- Gerenciamento de fitas off-site: automatiza a programação de dados de backup e arquivados, tanto no local de armazenamento quanto fora. Os administradores e operadores não precisam gerenciar esses dados em uma base volume por volume, o que economiza tempo e reduz a chance de erro;
- Recuperação de desastres: empresas que fazem backup de dados todos os dias também devem atualizar o plano de recuperação de desastres para refletir os números de volume diário de fitas. O TSM rastreia todas essas informações, consolidando-as com outras informações armazenadas no plano de recuperação de desastres e reduzindo o custo de atualizar manualmente o plano.

Um exemplo da aplicação do TSM pode ser visto em IBM (2013). A empresa Piedmont Healthcare, da área de saúde nos EUA, adquiriu, entre os anos de 2010 e 2012, quatro hospitais nos arredores de Atlanta. Com esta aquisição, o tamanho de seu ambiente de dados cresceu para cerca de 400 Tb. Após uma ação judicial que foi movida contra o

hospital, devido um vazamento de 300.000 arquivos de pacientes, e que resultou em uma multa de mais de 200 milhões de dólares, a segurança dos dados tornou-se uma questão importante. Piemonte implantou o software IBM Tivoli Storage Manager para acelerar e reforçar a sua infra-estrutura de backup e recuperação de dados e aumentar a segurança de dados. O resultado foi que essa implantação reduziu o tempo de backup em 86%, além de reforçar a segurança dos dados com criptografia em nível de fita.

Outra empresa que utiliza o TSM é a Colgate-Palmolive, conforme IBM (2012). A empresa necessitava implantar uma solução que gerenciasse o grande crescimento de dados associado ao desenvolvimento do negócio, mantendo os custos com TI baixos. A equipe de TI descobriu que gastavam muito tempo na administração do banco de dados que suportam as aplicações SAP da empresa. Funções críticas de negócios relacionados, tais como backups diários e processos de recuperação de desastres, também foram impactadas pelo crescimento do volume de dados. Com a aquisição do DB2 e TSM, Colgate reduziu seus volumes de banco de dados em uma média de 65 por cento, e o TSM fica responsável pelo backup e recovery destes dados.

De acordo com Neduziak (2010, np), existem alguns pontos que devem ser considerado pelas empresas no momento de sua implantação, tais como: complexidade da ferramenta (instalação, configuração e gerenciamento); preço das licenças; necessidade de implantação de uma rede SAN, visto que uma grande solicitação de backup e recuperação de dados acaba por sobrecarregar a rede.

6 O TSM NO AMBIENTE DO BANCO XYZ

Essencialmente, o TSM ainda é uma solução inovadora na proteção e gerenciamento de dados e permite que as empresas protejam, armazenem e garantam a disponibilidade de seus ativos de dados digitais. O TSM continua a ser uma boa solução de proteção de dados e abrangente recuperação, e disponibiliza aos clientes aproveitar para resolver suas próprias necessidades de proteção e gestão da informação (Rodhes, 2009).

Aplicações de negócios estratégicos são tipicamente coleções complexas de componentes interdependentes, tanto de software comercial e proprietária que extensão do desktop, ambientes distribuídos e ambientes de computação de mainframe. Aplicação de proteção está relacionada com a disponibilidade de dados, desempenho e capacidade de recuperação, e integra a gestão de dados do aplicativo em proteção de dados corporativos. O TSM torna-se uma opção de gerenciamento integrado que transforma tecnologia da informação em um recurso estratégico de negócios (IBM, 2006).

Conforme Rodhes (2009), o TSM disponibiliza muitos relatórios pré-configurados e totalmente personalizáveis, o que torna o gerenciamento do armazenamento mais eficiente, e o monitoramento da disponibilidade dos dados é realizado em tempo real. As ferramentas disponíveis no TSM permitem que se melhore e aperfeiçoe os servidores de TSM em grandes ambientes, como em mainframes, e causa menos impacto para os servidores gerenciados.

Como o servidor de TSM não rodava sobre o sistema operacional z/OS (sistema operacional do mainframe), uma possibilidade alternativa de uso é, ao invés de ter o servidor de TSM na plataforma z/OS, o servidor poderia ser hospedado em uma plataforma distribuída executando o AIX, Linux ou Windows, e um cliente TSM, que fosse executado em z/OS. Assim você ainda poderia fazer backup de seus dados de mainframe. Além disso, o servidor de TSM poderia ser executado em uma partição z/Linux do mainframe. Nas últimas versões lançadas, o TSM já inclui a plataforma z/OS em seu rol de suportadas (IBM, 2006).

Podemos visualizar na figura 6.1, a seguir, a representação de como funciona o TSM versão 6 para o sistema operacional z/OS, de mainframe (Wambeke, 2011).

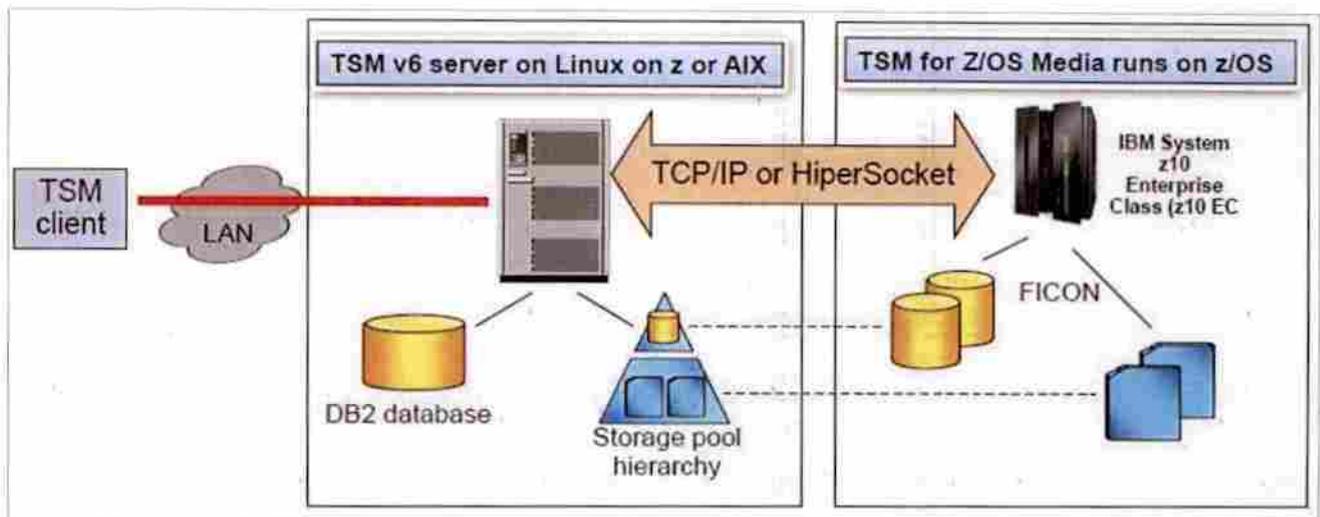


Figura 6.1: TSM para z/OS. Fonte: Wambeke (2011).

O TSM para z/OS possibilita que o banco de dados tenha responsabilidade sobre os nodes (máquinas), administradores, política de monitoramento dos dados de comunicação com os clientes, o centro de administração, a hierarquia do conjunto de armazenamento e outros servidores TSM. Além disso, o TSM executa I/O para a fibre connectivity (FICON) - de fita e disco. A FICON é uma interface I/O de alta velocidade para conexão do mainframe para dispositivos de armazenamento (Wambeke, 2011).

- **Consolidação de armazenamento:** Em um ambiente composto de servidores distribuídos capacidade de armazenamento fragmentada, muitas vezes os recursos de armazenamento são subutilizados, e realocar ou reconfigurar esses recursos, muitas vezes faz com ocorram interrupções e muito tempo de inatividade. O TSM possibilita melhorar a utilização de ativos, baixar os custos operacionais, e realocar automaticamente os recursos de armazenamento, conforme suas necessidades de mercado;

Os mainframes são utilizados em vários nichos de mercado, pela sua disponibilidade e possibilidade de processamento e armazenamento de um grande volume de dados. Segundo IBM (2006) o uso do TSM, no ambiente mainframe, se justifica pela sua aplicação nas seguintes áreas:

- **Proteção de dados:** com o aumento do valor das informações estratégicas, aumenta o valor com backup e armazenamento rápido e confiável. O TSM disponibiliza o servidor e a rede rapidamente para outros processos, fornece cópias de segurança e

restauração de dados críticos e utiliza da melhor forma os caros recursos de armazenamento;

- Recuperação de desastres: Perder o acesso a uma divisão chave ou repositório de dados da empresa por causa de um desastre pode custar um negócio, tanto no curto e longo prazo. Quando informações estratégicas são conduzidas 24x7x365, qualquer tempo de inatividade pode ser caro. O TSM permite projetar uma estratégia de tolerância a desastres, com nenhum ponto único de falha, que pode facilmente e com flexibilidade acomodar um grande volume de dados, como os armazenados em ambientes mainframes.

A relação do backup com os atributos de segurança da informação seguem explicados abaixo:

- Integridade: Geralmente, um software de backup corporativo só faz o backup após a verificação da integridade da informação na origem, no disco, filesystem. Nesse processo é verificada a integridade do bloco e qualquer erro de i/o é alarmado e a informação não é gravada. A empresa possui processo de monitoração para esses erros. De outra forma, há ainda processo de validação da informação armazenada em mídia do tipo LTO, onde rotinas periódicas oxigenam e auditam as fitas mais antigas, garantindo sua integridade;
- Disponibilidade: O software de backup serve como apoio à disponibilidade dos dados. Uma vez que ocorra algum incidente ou mesmo um desastre, a informação pode ser recuperada do backup;
- Confidencialidade: Os operadores e responsáveis pelas rotinas e dados de backup assinaram termos de confidencialidade. Além disso, pode-se configurar backup criptografado para informações críticas, tais como, bancos de dados de informações de negócios, arquivos que contenham regras de firewall, drives de rede confidenciais. Dessa forma, a informação só poderá ser lida se restaurada no local de origem. Caso seja extraviada ou restaurada em local diversos, não será lida devido a falta da chave criptográfica.

A figura 6.2 a seguir demonstra uma topologia resumida do TSM no ambiente do Banco XYZ.

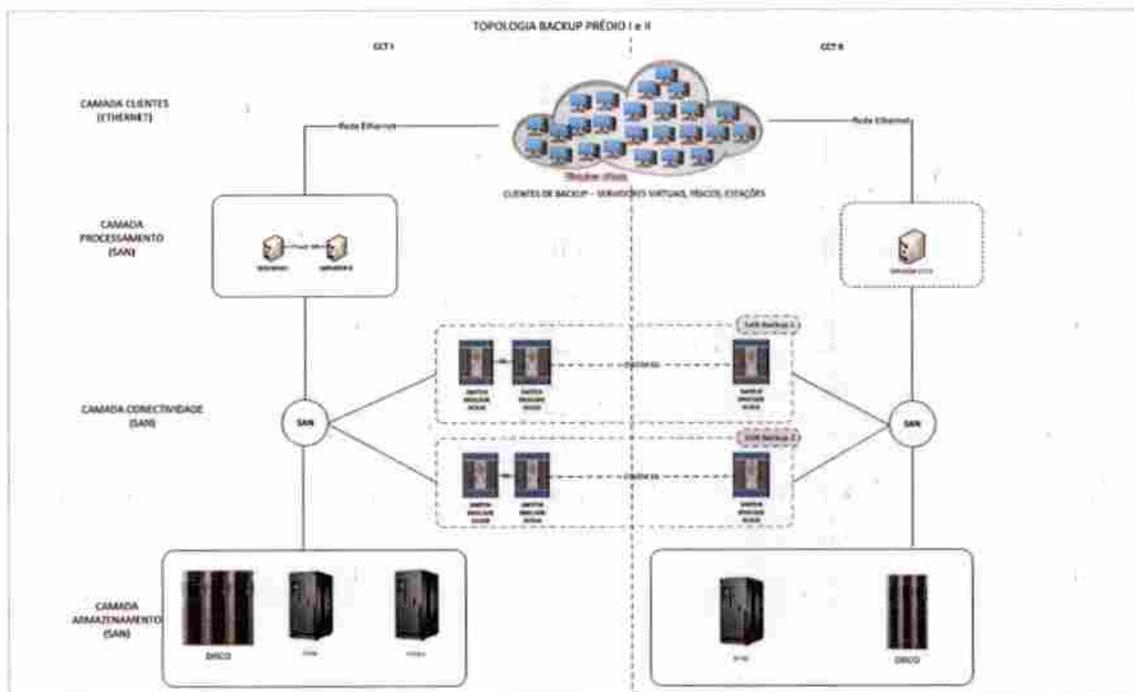


Figura 6.2: Topologia TSM no Banco XYZ. Fonte: Elaborada pela autora.

A figura demonstra a operação de backup de dados através de um rede SAN, que é uma rede dedicada de armazenamento utilizada para melhorar o desempenho do processo. Ao integrar o TSM a uma rede SAN, procura-se compartilhar dispositivos de armazenamento entre os servidores TSM, mover diretamente os dados dos clientes para os servidores, compartilhar unidades de discos e fitas e utilizar o cluster GPFS (General Parallel File System) que permite que vários clientes sejam consolidados em um único node no TSM Server.

Segue listada a seguir a estrutura do ambiente TSM do Banco XYZ:

- Área armazenada: 8 Petabytes (8Pb);
- Servidores: 18 servidores de backup, em cluster, e 18 aplicações TSM Server, de Arquitetura IBM Power 7, Processadores Risk de 6 cores com 100Gb de memória;
- Clientes: 5.504 nodes de backup;
- Média de backup: 50Tb de backup/dia;
- Fitas: 11.000 volumes LTO5 e LTO6.

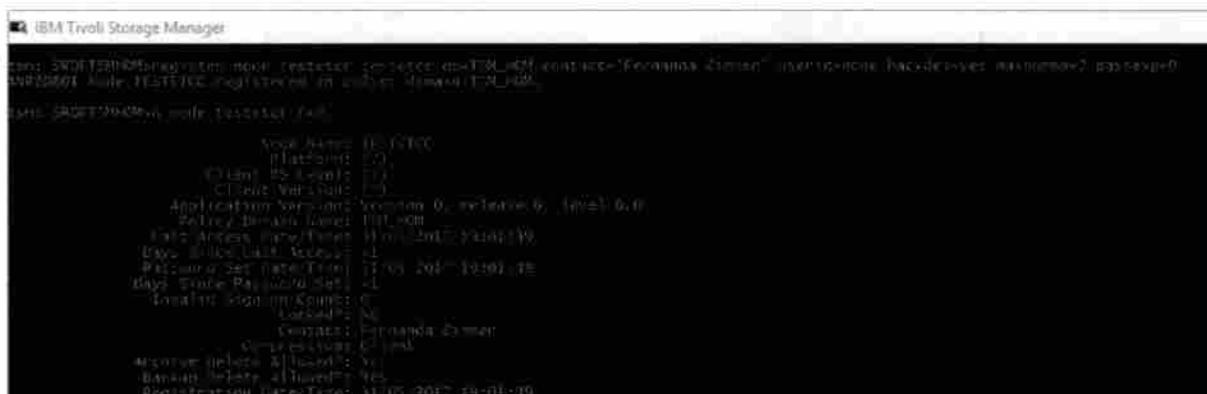
O Banco XYZ possui um TSM4VE que disponibiliza backup para mais de 4000 servidores virtuais, além de uma recuperação instantânea. Se um servidor físico com diversas máquinas virtuais associadas for perdido em casos de ataques do tipo ransomware, por

exemplo, o TSM consegue recuperá-las em um novo servidor físico, sem perda de informação. O dado retorna íntegro pois são realizados testes de backup e restore.

O processo físico do backup do TSM envolve a guarda de documento inicialmente em storage pool de disco e depois em storage pool de fita. E o processo lógico envolve política de backup, que define quantas versões do dado serão armazenadas e qual o tempo de retenção.

Para executar o backup é necessário o registro da máquina (node) no servidor do TSM que irá gerenciar esse processo. O comando utilizado para esse registro é “register node” e ele define um nó cliente para o servidor e configura opções para este.

A figura 6.3 abaixo demonstra o registro de node no TSM Server.



```
IBM Tivoli Storage Manager
> q node SRDFTSMHOM; register node testetcc -serverid=SRDFTSMHOM -contact="Fernanda Camar" -userprofile=BACKUPUSER -maxspace=2 -optset=OPTSET_PADRAO
> q node SRDFTSMHOM; node testetcc /q
Node Name: TESTETCC
Platform: AIX
Client Type: Client
Client Version: 6.2.3
Application Version: Version 6, release 9, level 6.0
Policy Domain Name: D01_000
File Access Type: Full
File Access Type: Full
File Access Type: Full
Password Set Date Time: 11/05/2017 19:01:39
Data Source Password Set: 1
Inhibit Migration Count: 0
Contact: Fernanda Camar
Contact: Fernanda Camar
Archive Deletion Allowed: Yes
Backup Deletion Allowed: Yes
Registration Date Time: 11/05/2017 19:01:39
```

Figura 6.3: Registro de node no TSM Server. Fonte: Elaborada pela autora.

O exemplo criado é do registro de um node chamado TESTETCC no servidor TSM denominado SRDFTSMHOM. Para listar as informações do node, utiliza-se o comando “q node testetcc f=d”, que exibe informações parciais ou completas sobre o node.

Após o registro do node, deve ser associada uma política de backup e retenção, chamada cloptset. O exemplo utiliza a política padrão denominada OPTSET_PADRAO e com o comando “q clop OPTSET_PADRAO” é possível exibir informações sobre um conjunto de opções do cliente. Nesse caso, é realizado o backup de todos os filesystems do servidor, com política de retenção de 2 versões ou um mês, excluindo os filesystems /tmp e /home.

A associação da política ao node é feita através do comando “update node testetcc cloptset=OPTSET_PADRAO”.

A figura 6.4 a seguir demonstra a associação do cloptset ao node registrado.

7 CONCLUSÕES

O presente trabalho apresentou um estudo acerca das vantagens do armazenamento, backup e recuperação de dados, além de descrever o ambiente mainframe e a ferramenta IBM TSM, com suas características e funcionalidades aplicados ao Banco XYZ.

Atualmente, as empresas utilizam diversas formas de armazenamento de suas informações críticas e o mercado disponibiliza ferramentas para esse fim. Como a gestão de armazenamento de dados torna-se cada dia mais sofisticado, também se torna mais complicado. A solução para gerir adequadamente um ambiente complexo de armazenamento é abordá-lo estrategicamente.

A ferramenta Tivoli Storage Manager (TSM), desenvolvida pela IBM, disponibiliza esse gerenciamento, além de possibilitar o armazenamento e recuperação das informações. Suas características permitem sua aplicação desde sistemas high-end até ambiente mainframe. O ambiente mainframe se caracteriza por um grande volume de dados, sendo estes processados e gerenciados full time.

A utilização do TSM no processo de backup e recuperação de dados pelo Banco XYZ, conforme demonstrado no trabalho, permite que essa grande quantidade de informação seja armazenada de forma segura, e, sua recuperação, realizada de forma rápida e eficiente, garantindo ao negócio que os dados não serão perdidos em caso de algum desastre.

7.1. Trabalhos Futuros

Como proposta de trabalhos futuros são indicados alguns pontos que podem ser evoluídos.

A fim de dar continuidade ao trabalho, justifica-se estudar a aplicabilidade do TSM em ambientes high-end, que são, em muitas empresas, o suporte para o negócio implantado em mainframe.

Sugere-se aplicar a solução do TSM4VE em servidores que atendam não somente o negócio fim do banco e sim todas as áreas de apoio, como Diretorias, Superintendências, etc. Dessa forma, tanto o backup quanto a recuperação dos dados se dá de forma mais rápida e efetiva, garantindo a disponibilidade e segurança dos dados da empresa.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARROS, E. **Entendendo os Conceitos de Backup - Restore e Recuperação de Desastres**. Rio de Janeiro: Ciência Moderna, 2007.
- CALVANO, I. **Conceitos Básicos de Backup**. 2017. Disponível em: <http://www.linhadecodigo.com.br/artigo/1171/conceitos-basicos-de-backup.aspx>. Acesso em: 21 fev. 2017.
- DELL. **O que é armazenamento?** 2007. Disponível em: <http://www.dell.com/learn/br/pt/brbiz1/storage-solutions-what-is-it>. Acesso em: 15 fev. 2017.
- DIAZ, I. **O que é backup?** Dez. 2011. Disponível em: <http://www.techtudo.com.br/artigos/noticia/2011/12/o-que-e-backup.html>. Acesso em: 21 fev. 2017.
- DORION, P. **IBM Tivoli Storage Manager vs. traditional backup**. Jan. 2007. Disponível em: <http://searchdatabackup.techtarget.com/tip/IBM-Tivoli-Storage-Manager-vs-traditional-backup>. Acesso em: 24 fev. 2017.
- FREIRE, E. S. **Organização e Arquitetura de Computadores**. 2017. Disponível em <http://www.ebah.com.br/content/ABAAAAQe4AH/organizacao-arquitetura-computadores>. Acesso em: 16 fev. 2017.
- IBM Corporation. **Colgate-Palmolive supports continuing business growth with IBM and SAP**. Ago. 2012. Disponível em: http://www-01.ibm.com/software/success/cssdb.nsf/CS/STRD-8X9HCH?OpenDocument&Site=default&cty=en_us. Acesso em: 26 fev. 2017.
- _____. **IBM Tivoli Storage Management Concepts**. RedBooks. 2006. Disponível em: <http://www.redbooks.ibm.com/redbooks/pdfs/sg244877.pdf>. Acesso em: 22 fev. 2017.
- _____. **Introduction to Storage Area Networks and System Networking**. RedBooks. Nov. 2012. Disponível em: <http://www.redbooks.ibm.com/redbooks/pdfs/sg245470.pdf>. Acesso em: 22 fev. 2017.
- _____. **Mainframe concepts**. 2017. Disponível em: http://publib.boulder.ibm.com/infocenter/zos/basics/index.jsp?topic=/com.ibm.zos.zmainframe/zconc_opsyszosintro.htm. Acesso em: 22 fev. 2017.

- _____. **Mainframes photo album**. 2017. Disponível em: http://www-03.ibm.com/ibm/history/exhibits/mainframe/mainframe_album.html. Acesso em: 05 mar. 2017.
- _____. **Piedmont Healthcare reduces data backup time by 86 percent**. Mai. 2013. Disponível em: http://www-01.ibm.com/software/success/cssdb.nsf/CS/RNAE-97YJUF?OpenDocument&Site=default&cty=en_us. Acesso em: 26 fev. 2017.
- _____. **Tivoli Storage Manager: quick reference for backup-archive clients V.4 R.1**. USA: IBM Corporation, 2000.
- _____. **Using the application Program Interface: Tivoli Storage Manager V.4 R.1**. USA: IBM Corporation, 2000.
- LAUNDON, K. C.; LAUDON, J. P. **Gerenciamento de Sistemas de Informação**. Rio de Janeiro. LTC, 2001.
- LOEST, S. R. **Um sistema de backup cooperativo tolerante a intrusões**. Jun. 2009. Disponível em: http://www.ppgia.pucpr.br/lib/exe/fetch.php?media=dissertacoes:2009:2009_sergioloest.pdf. Acesso em: 21 jan. 2017.
- LOPES, S. **O que é um sistema operacional?** Mar. 2008. Disponível em: http://www.oficinadanet.com.br/artigo/851/o_que_e_um_sistema_operacional. Acesso em: 21 jan. 2017.
- MADEIRA, D. **O que são dispositivos de armazenamento?** Jul. 2010. Disponível em: <http://dan-scientia.blogspot.com.br/2010/07/o-que-sao-dispositivos-de-armazenamento.html>. Acesso em: 21 jan. 2017.
- MONTE, A. C.; LOPES, L. F. **A qualidade dos suportes no armazenamento de informações**. Florianópolis: Visual Books, 2004.
- NEDUZIAK, J. **Sistema de Backup utilizando IBM – Tivoli Storage Manager**. Curitiba. Out. 2010. Disponível em: <http://www.ppgia.pucpr.br/~jamhour/RSS/TCCRSS09A/Josmar%20Neduziak%20-%20RSS09A.pdf>. Acesso em: 23 fev. 2017.
- OLIVEIRA, S B. *et al.* **Introdução à Arquitetura de Mainframe e ao Sistema Operacional z/OS**. Rio de Janeiro. Editora Ciência Moderna, 2010.
- RED HAT. **O que é LVM?** 2017. Disponível em: http://web.mit.edu/rhel-doc/3/rhel-sag-pt_br-3/ch-lvm-intro.html. Acesso em: 21 fev. 2017.
- REZENDE, R. **RAID – Parte I. Armazenamento com segurança, performance e baixo custo!** 2012. Disponível em: <http://www.devmedia.com.br/raid-parte-i->

- [armazenamento-com-seguranca-performance-e-baixo-custo/2176](#). Acesso em: 15 fev. 2017.
- RODHES, R. **Managing Data Overflow**. Jul. 2009. Disponível em: <http://www.ibmssystemsmag.com/mainframe/storage/software/Managing-Data-Overflow/?page=1>. Acesso em: 05 mar. 2017.
- ROUSE, M. **IBM TSM (IBM Tivoli Storage Manager)**. Jul. 2013. Disponível em: <http://searchdatabackup.techtarget.com/definition/IBM-TSM-IBM-Tivoli-Storage-Manager>. Acesso em: 24 fev. 2017.
- _____. **MVS (Multiple Virtual Storage)**. Nov. 2005. Disponível em: <http://searchdatacenter.techtarget.com/definition/MVS>. Acesso em: 24 fev. 2017.
- SASSO, J. **TSM na Prática**. Ago. 2010. Disponível em: <http://tsmnapratica.blogspot.com.br>. Acesso em: 23 fev. 2017.
- SETZER, V. W. **Dado, Informação, Conhecimento e Competência**. Dez. 2009. Disponível em: <http://www.ime.usp.br/~vwsetzer/datagrama.html>. Acesso em: 05 mar. 2017.
- SOARES, L. **Meios de Armazenamento**. Fev. 2011. Disponível em: <http://pt.slideshare.net/Louriedson/meios-de-armazenamento>. Acesso em: 15 fev. 2017.
- SOMASUNDARAM, G. *et al.* **Armazenamento e gerenciamento de informações: como armazenar, gerenciar e proteger informações digitais**. Porto Alegre: Bookman, 2010.
- VIANA, G. **O que é um Host?** Fev. 2012. Disponível em: <http://www.techtudo.com.br/artigos/noticia/2012/02/o-que-e-um-host.html>. Acesso em: 15 fev. 2017.
- WAMBEKE, M. **Tivoli Storage Manager for z/OS Media 6.3**. Nov. 2011. Disponível em: <http://mainframe-watch-belgium.blogspot.com.br/2011/11/tivoli-storage-manager-for-zos-media-63.html>. Acesso em: 26 fev. 2017.