

CENTRO UNIVERSITÁRIO UNINORTE
PÓS-GRADUAÇÃO EM GERÊNCIA DE BANCO DE DADOS
SISTEMA DE ARMAZENAMENTO MASSIVO DE DADOS



ARMAZENANDO GRANDES VOLUMES DE DADOS

Prof. Luiz Matos

Rio Branco, AC, setembro de 2019.

AGENDA

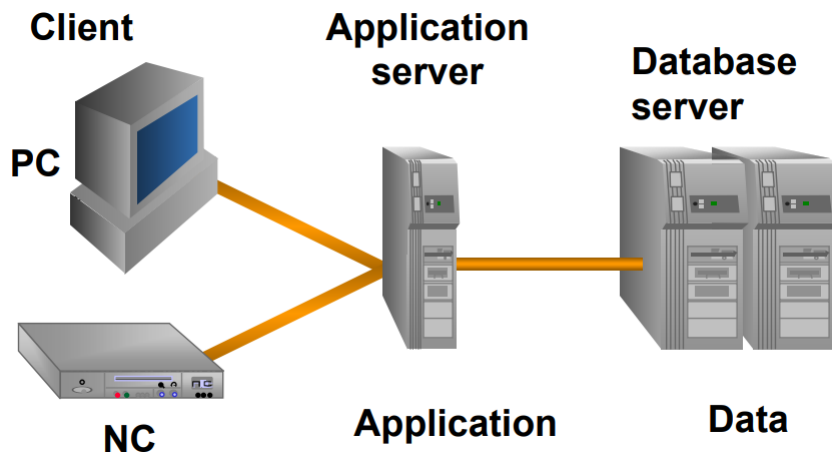
- **Arquiteturas de armazenamento massivo de dados**
- ***Redundant Array of Inexpensive/Independent Disks***
- **Tecnologias de armazenamento**
- **Backup e recuperação**
- **Gerenciamento da infraestrutura de armazenamento**

ARQUITETURAS

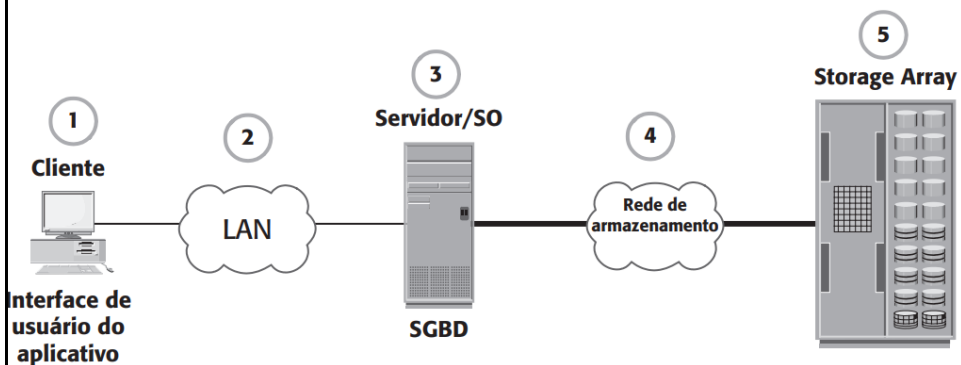
Relembrando . . .

- ❑ A arquitetura de um sistema estabelece a sua **estrutura**, através da:
 - ❑ Identificação dos seus **componentes**;
 - ❑ Definição de suas **funções** e;
 - ❑ Dos **relacionamentos**/interações entre eles.

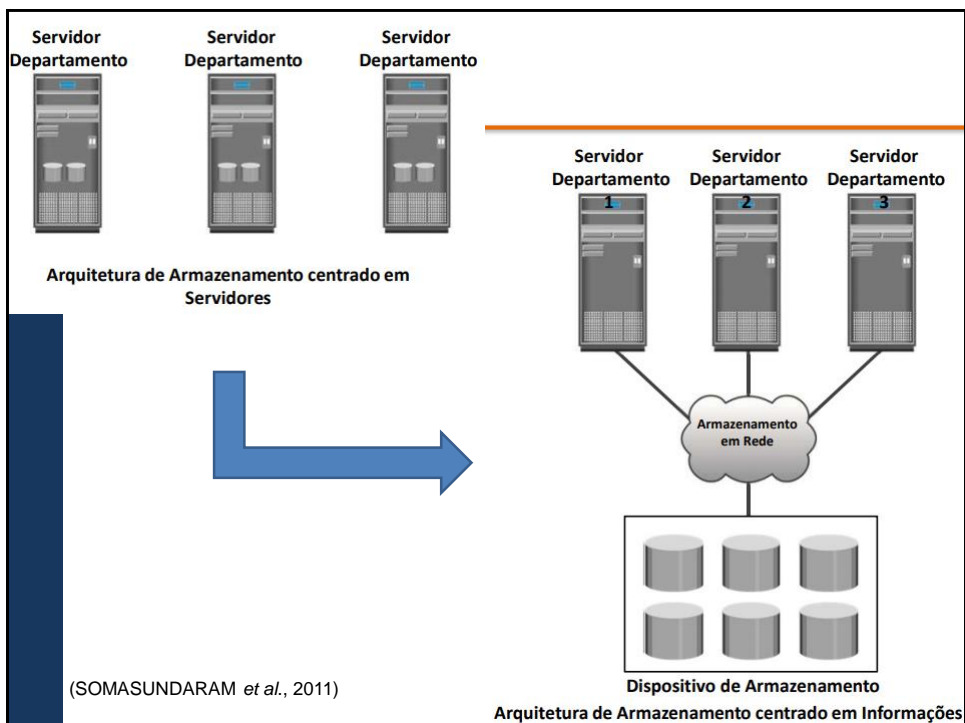
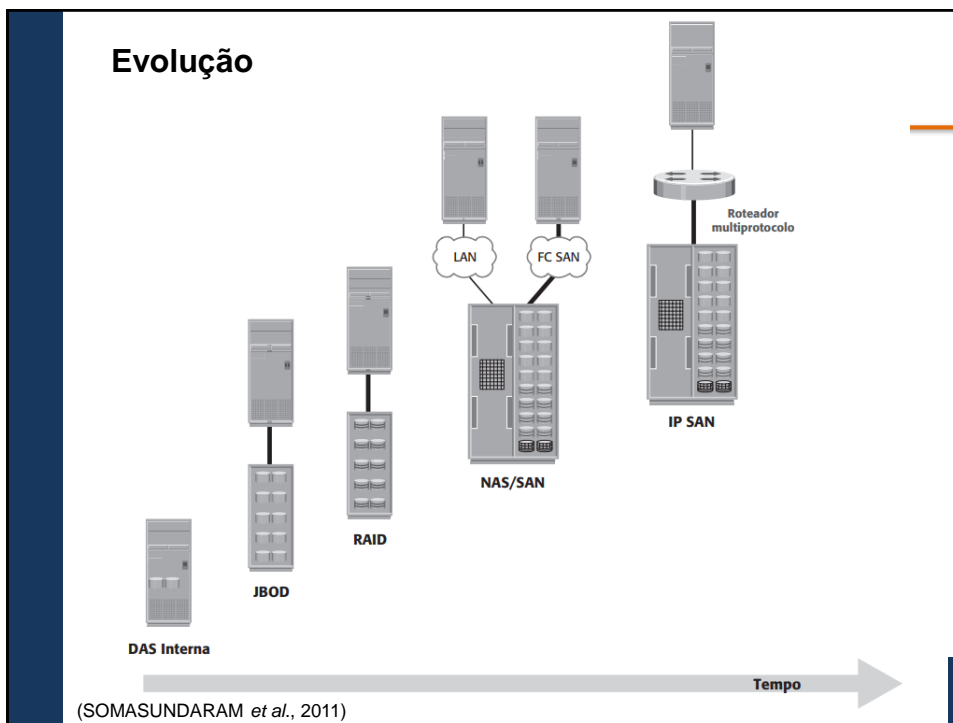
Arquitetura em 3 camadas (3-tier)



Exemplo



(SOMASUNDARAM *et al.*, 2011)



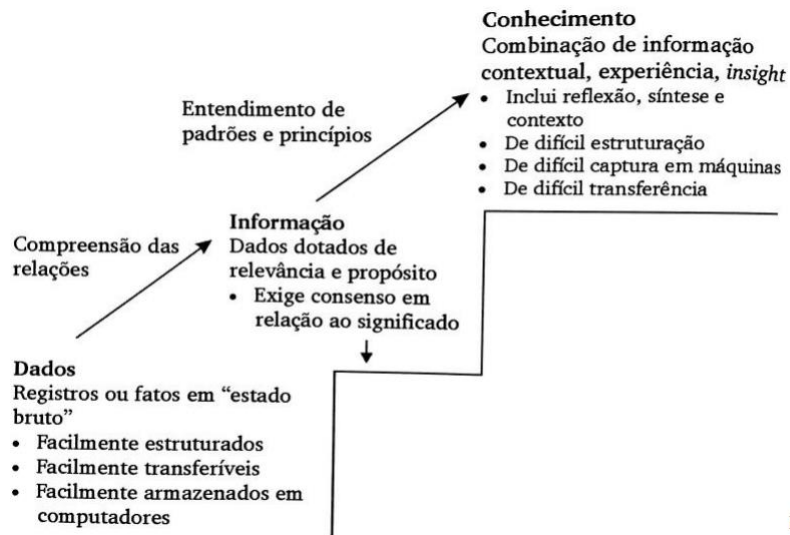
CONTEXTO

- ❑ **INFORMAÇÃO:** principal ativo organizacional!
- ❑ **Aumento das necessidades/quantidade de dados a serem armazenados**
- ❑ **Produção voluntária e involuntária de dados**

CONTEXTO

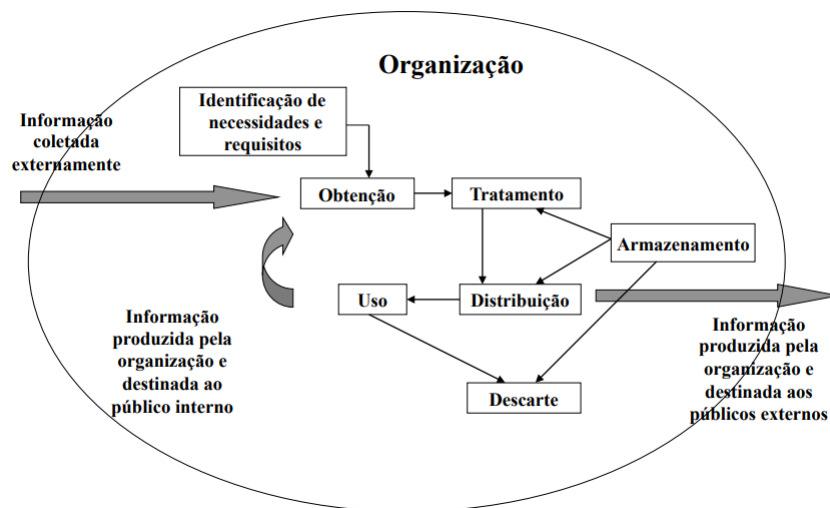
- ✓ **+ capacidade de processamento**
- ✓ **> custos de armazenamento**
- ✓ **redes de comunicação + rápidas (e acessíveis!)**
- ✓ **+ aplicativos e dispositivos inteligentes**

Dados x Informação x Conhecimento

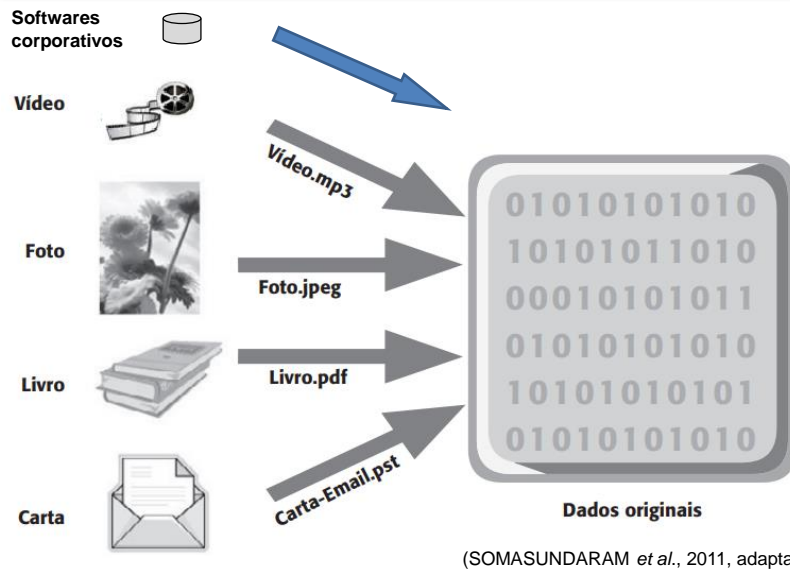


Beal (2008)

Ciclo de vida da informação



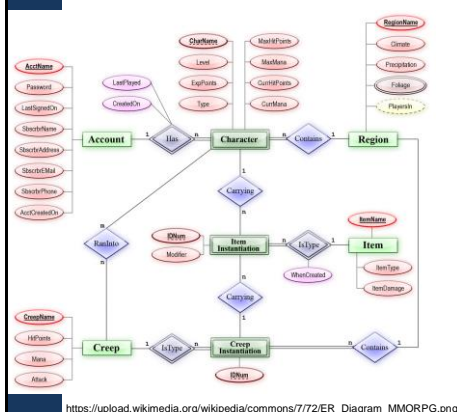
Heterogeneidade de tipos e fontes



Heterogeneidade de tipos e fontes

□ Dados estruturados

- ✓ Seguem um padrão previamente definido



https://upload.wikimedia.org/wikipedia/commons/7/72/ER_Diagram_MMORPG.png

E-mail
Ex: joaoasilva@gmail.com

Senha
Ex: 12345678910

CPF
Ex: 123.456.789-12

Seu nome
Ex: Thiago

Data de nascimento
Ex: 01/01/1999

Sexo
☐ Masculino ☐ Feminino

Telefone
Ex: (99) 99999-9999

☐ Receber notificações por Whatsapp

☐ Desejo receber ofertas por e-mail

CRIAR SEU CADASTRO

Heterogeneidade de tipos e fontes

❑ Dados não estruturados

- ✓ NÃO seguem um padrão previamente definido

Carta de Boas Vindas Caixa de entrada x

Natássia da Silva Nogueira
para eu, Luis, Eufrásia, Wanderson +
Boa tarde Prezado Professor Luiz,

seguem anexas algumas informações para auxiliá-la no módulo de pós-graduação desse fim de semana

- Carta de boas vindas
- Modelo de plano de aula
- Modelo da folha de atividades
- Modelo de slide (caso seja necessário)

Estamos a disposição para auxiliar no que for necessário.

Cordialmente,

Natassia Nogueira
Supervisora Acadêmica de Pós-Graduação FGV
UNINORTE – União Educacional do Norte
☎ (68) 3302-7088 ☎ Cel. (68) 98401-6136
✉ natassia.nogueira@uninorteac.com.br
🌐 www.uninorteac.com.br
CENTRO UNIVERSITÁRIO
uninorte



Jogo de Banho Isabelle 5 Peças - Casa & Conforto
★★★★★ (14)
3 ofertas a partir de:
R\$ 69,99
1x de R\$ 23,33 sem juros



Smart TV LED 43" Samsung 43J5290 Full HD com Conversor...
★★★★★ (38)
29 ofertas a partir de:
R\$ 1.299,00
10x de R\$ 129,90 sem juros

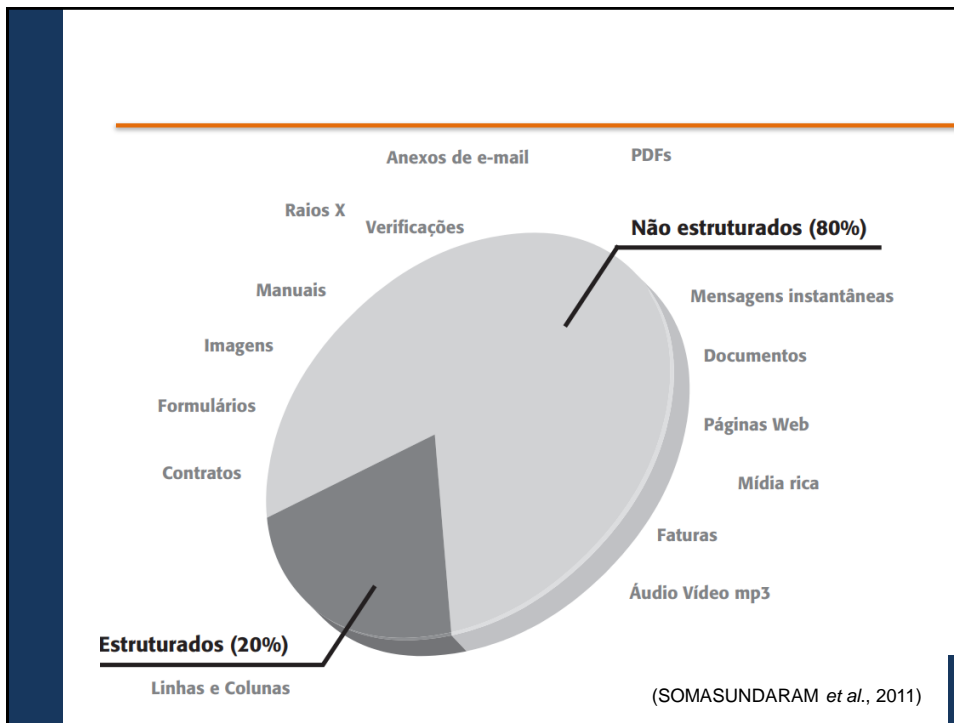
Heterogeneidade de tipos e fontes

❑ Dados semiestruturados

- ✓ Representação intermediária de modelo/estrutura

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<receita nome="pão" tempo_de_preparo="5 minutos" tempo_de_cozimento="1 hora">
  <titulo>Pão simples</titulo>
  <ingredientes>
    <ingrediente quantidade="3" unidade="xícaras">Farinha de Trigo</ingrediente>
    <ingrediente quantidade="7" unidade="gramas">Fermento</ingrediente>
    <ingrediente quantidade="1,5" unidade="xícaras" estado="morna">Água</ingrediente>
    <ingrediente quantidade="1" unidade="colheres de chá">Sal</ingrediente> </ingredientes>
  <instrucoes>
    <passo>Misture todos os ingredientes, e dissolva bem.</passo>
    <passo>Cubra com um pano e deixe por uma hora em um local morno.</passo>
    <passo>Misture novamente, coloque numa bandeja e asse num forno.</passo>
  </instrucoes>
</receita>
```

```
{ "Alunos": [
  { "nome": "Juvenal Antunes", "notas": [ 8, 9, 5 ] },
  { "nome": "Francisco Xavier", "notas": [ 8, 10, 7 ] },
  { "nome": "Maria Antonieta", "notas": [ 10, 10, 9 ] }
]}
```

Desafios

- ☐ Armazenamento, proteção e otimização
- ☐ Tomada de decisão
- ☐ Vantagem competitiva
- ☐ Novos negócios

Big Data

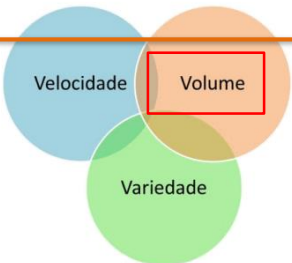
*“Termo utilizado para descrever
grandes volumes de dados”.*
(IBM)

Big Data



Bit	0 ou 1	Menor unidade de dados
Byte	8 bits	Um caractere
Kilobyte	~1.000 bytes	Uma página de livro
Megabyte	~1.000.000 de bytes	1/5 da obra de Shakespeare
Gigabyte	1.000.000.000 de bytes	1 hora de vídeo

Big Data



Terabyte	~1.000.000.000.000 de bytes	385TB = Biblioteca do Congresso Americano
Petabyte	~1.000.000.000.000.000 de bytes	1,5PB = todas as músicas
Exabyte	~1.000.000.000.000.000.000 de bytes	2,5EB = produzimos por dia
Zettabyte	~1.000.000.000.000.000.000.000 de bytes	1,8ZB = produzimos em um ano

Big Data




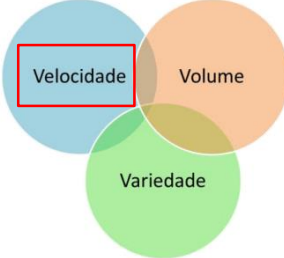
~ 40 PB	Google
~ 60 PB	Smartphones e tablets
~20 PB	Emails
Redes sociais, documentos, transações etc.	

Novo supercomputador mais poderoso do mundo entra em operação nos EUA

Por LEONARDO MÜLLER, @leomuller - 01º PRODUTO - 08 JUN 2018 - 13:05

COMPARTILHAR


2.245 compartilhamentos

Google inaugura centro de computação na nuvem em São Paulo

Por Stephanie Kohn | 19 de Setembro de 2017 às 13h19

TUDO SOBRE



Google

"Agora somos vizinhos e estamos ao seu lado". Foi com esta frase que Joe Kava, VP Global de Data Center do Google, anunciou a chegada do data center Google Cloud ao Brasil. Esta é a primeira locação da plataforma na América Latina e promete melhoria de 80% a 95% na latência para três países da região: Brasil, Chile e Argentina.

Além de mais agilidade no acesso aos dados armazenados, o lançamento da região de nuvem em São Paulo traz ainda a possibilidade de clientes brasileiros pagarem suas faturas em Reais. A conversão para a moeda local tem como objetivo atender às empresas brasileiras que tinham algum impedimento em comprar fora do país.

Big Data

- ✓ **4.4 billion** internet users (June 2019)
- ✓ **300 million** new social media users each year (550/minute)
- ✓ **474,000 Tweets** PER MINUTE
- ✓ Users are **watching 4,333,560 videos** every minute
- ✓ **300 hours** of video are uploaded to YouTube every minute
- ✓ Instagram users upload over **100 million photos and videos** everyday and **69,444 million** posts/minute

<https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/>

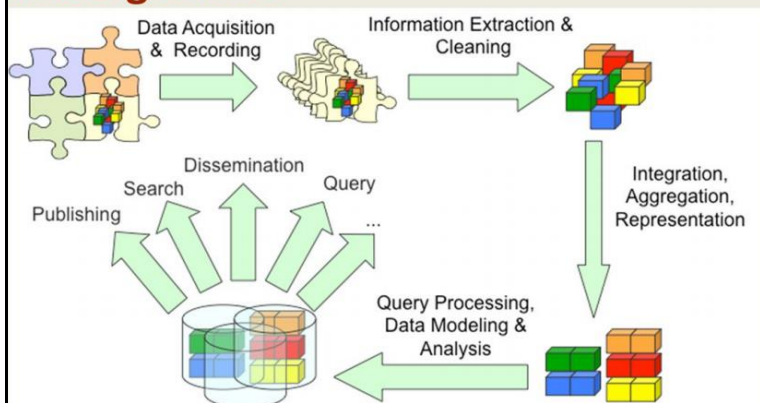
Big Data

- ✓ Every minute on Facebook:
 - **510,000 comments** are posted
 - **293,000 statuses** are updated
 - **136,000 photos** are uploaded

- ✓ 3.5 billion Google searches each minute of everyday
(40,000 searches/second)

<https://blog.microfocus.com/how-much-data-is-created-on-the-internet-each-day/>

It is Big Data !



The Big Data Analysis Pipeline
 H. V. Jagadish – ACM SIGMOD Blog - 05/06/2012
 Challenges & Opportunities w/ Big Data – Online report

Slide do Prof. Altigran (UFAM)

Onde aplicar?

- ☐ Tomada de Decisão
- ☐ Inteligência de Negócios
- ☐ Comércio Eletrônico
- ☐ Economia
- ☐ Saúde
- ☐ Clima
- ☐ Educação
- ☐ Biodiversidade
- ☐ Etc. etc. etc.

Conceitos fundamentais



❑ Data center

- Computadores + sistemas de armazenamento + dispositivos de rede + alimentação (energia) + controles de ambiente (ar, incêndio, acesso)

❑ Virtualização

- Abstração do hardware permitindo que vários sistemas operacionais sejam executados de maneira simultânea em um único computador

❑ Computação em nuvem

- dd

❑ Gerenciador de volumes lógicos (LVM)

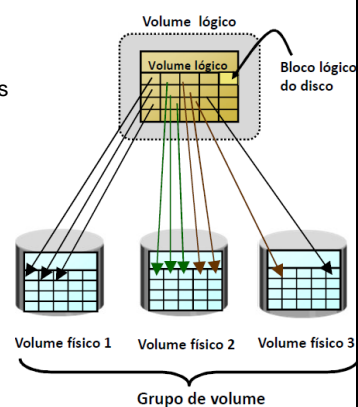
- Converte a visão física do armazenamento para uma visão lógica, mapeando os blocos de dados de cada visão.

➤ Particionamento

- 1 disco físico em unidades lógicas menores

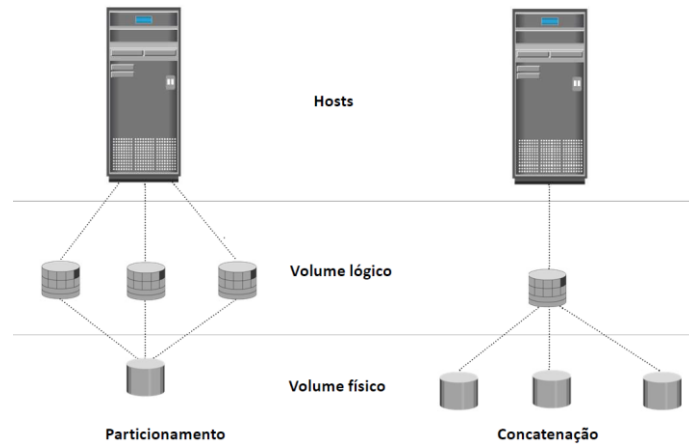
➤ Concatenação

- N discos físicos em volume virtual maior



(SOMASUNDARAM *et al.*, 2011)

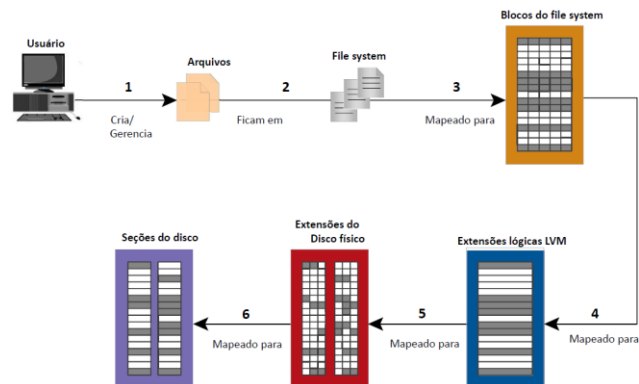
❑ Gerenciador de volumes lógicos (LVM)



(SOMASUNDARAM *et al.*, 2011)

❑ Sistema de arquivos

- Estrutura lógica e rotinas de software que controlam o acesso aos arquivos.

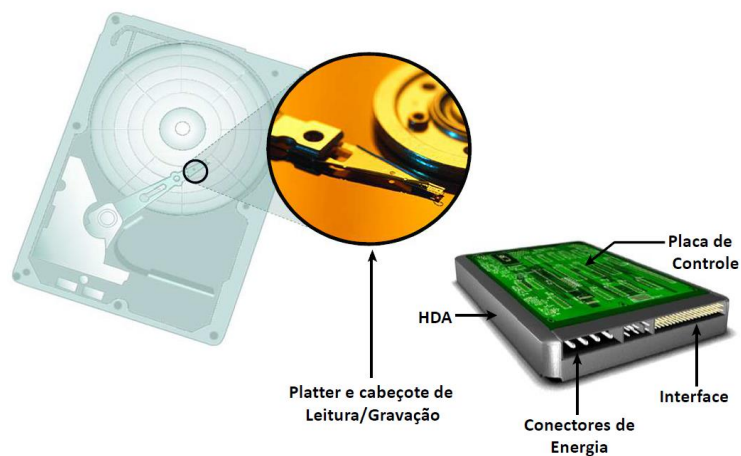


(SOMASUNDARAM *et al.*, 2011)

❑ Dispositivos de armazenamento

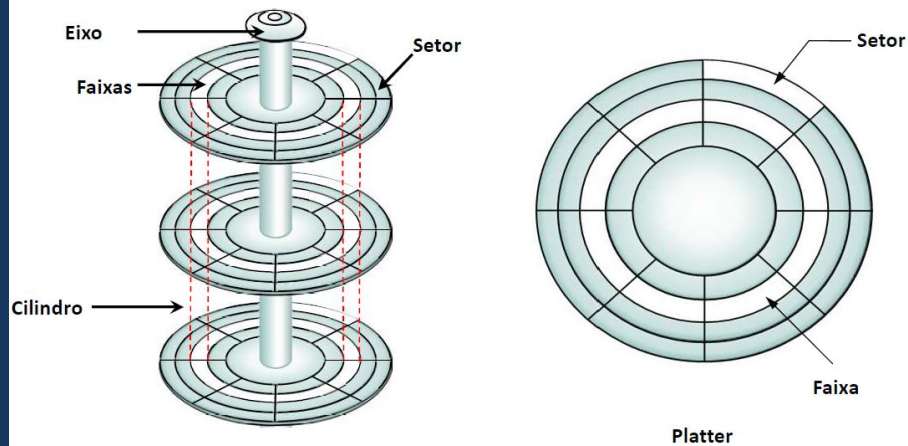
- Fita magnética (baixo custo, acesso sequencial, desgaste)
- Discos ópticos (capacidade e velocidade limitados)
- Disco rígido (grande capacidade de armazenamento, acesso aleatório)
- Flash drive (alto desempenho e baixo consumo de energia)

❑ Estrutura física do disco rígido



(SOMASUNDARAM *et al.*, 2011)

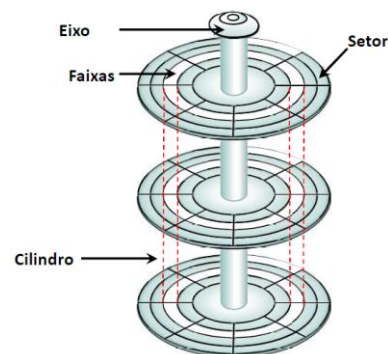
□ Estrutura física do disco rígido



(SOMASUNDARAM *et al.*, 2011)

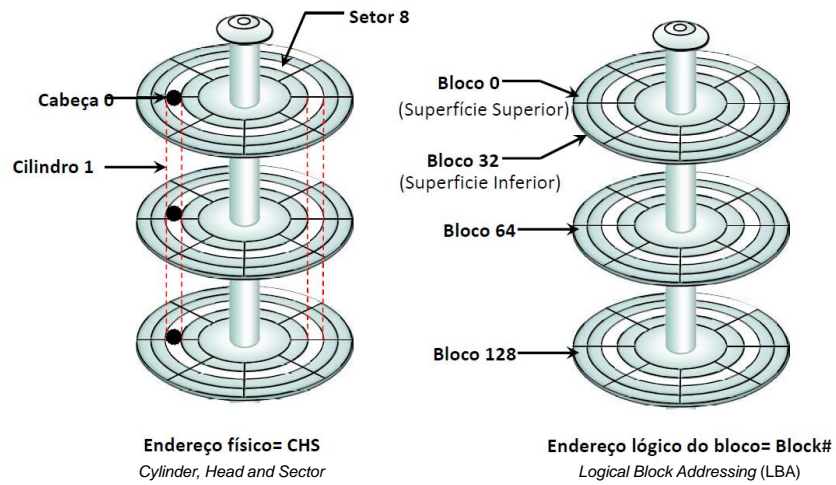
□ Estrutura física do disco rígido

- **Faixas:** anéis concêntricos no platter
- **Setor:** menor unidade de armazenamento individual endereçável
- **Cilindro:** conjunto de faixas idênticas em ambas superfícies de cada platter



(SOMASUNDARAM *et al.*, 2011)

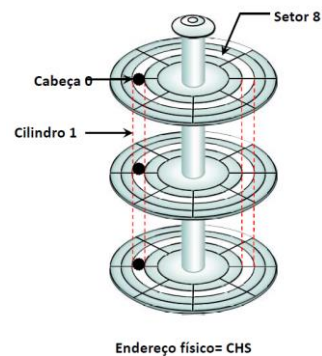
□ Estrutura física do disco rígido



(SOMASUNDARAM *et al.*, 2011)

□ Estrutura física do disco rígido

- 8 setores por faixa
- 6 cabeçotes
- 4 cilindros
- Blocos lógicos = $8 \times 6 \times 4 = 192$



(SOMASUNDARAM *et al.*, 2011)

