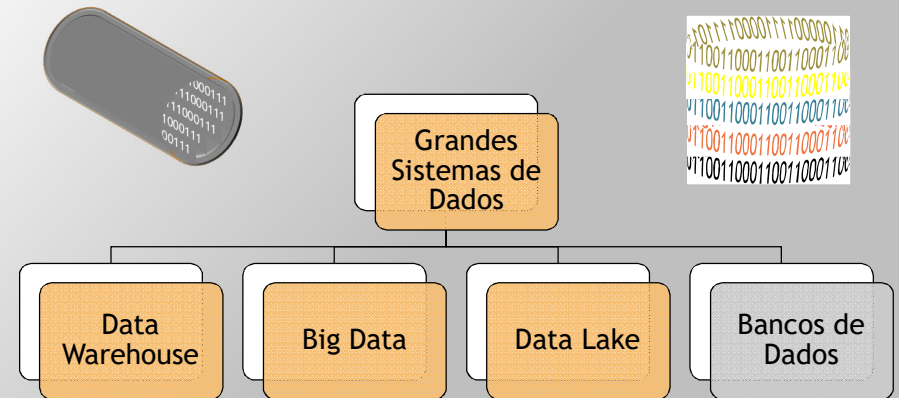


TSW017 GRANDES SISTEMAS DE DADOS

Prof. Dr. Jorge Rady

1

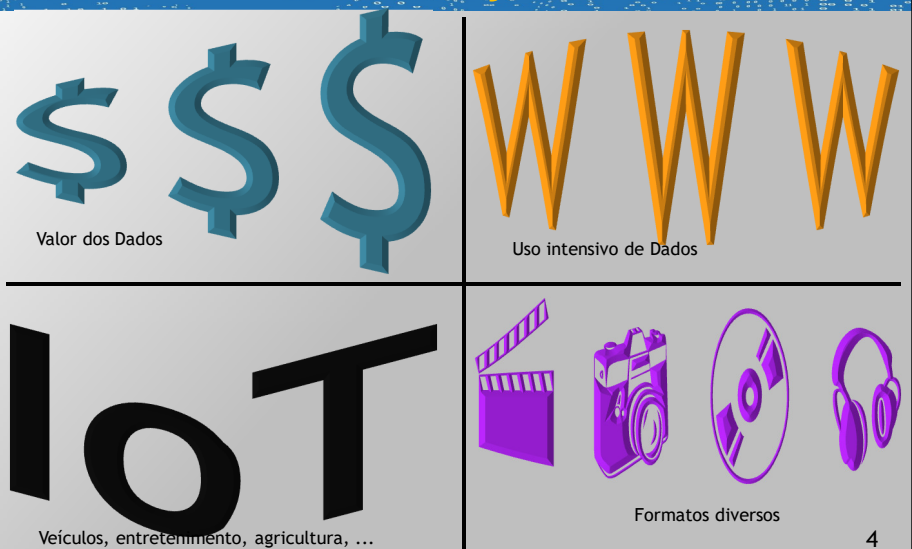
Contextualização



2

Grandes Sistemas de Dados Motivações

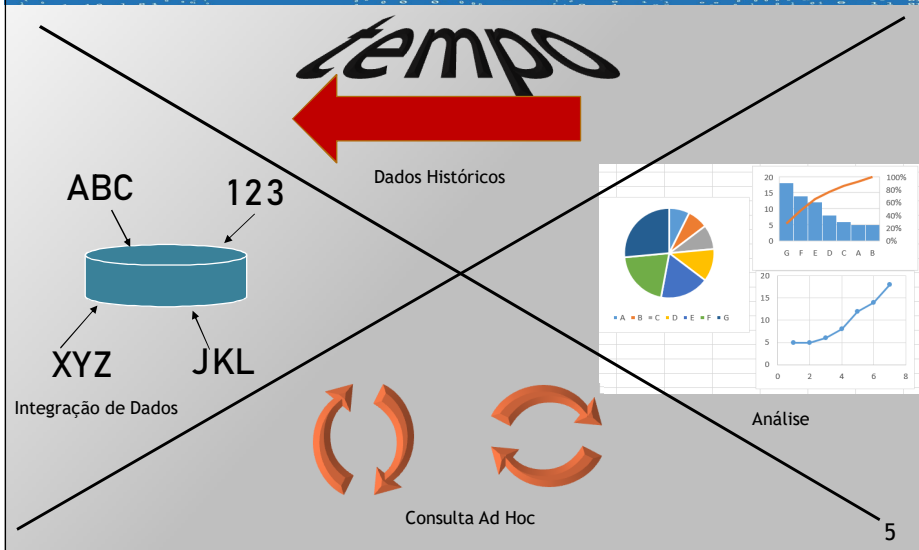
Grandes Sistemas de Dados - Motivações



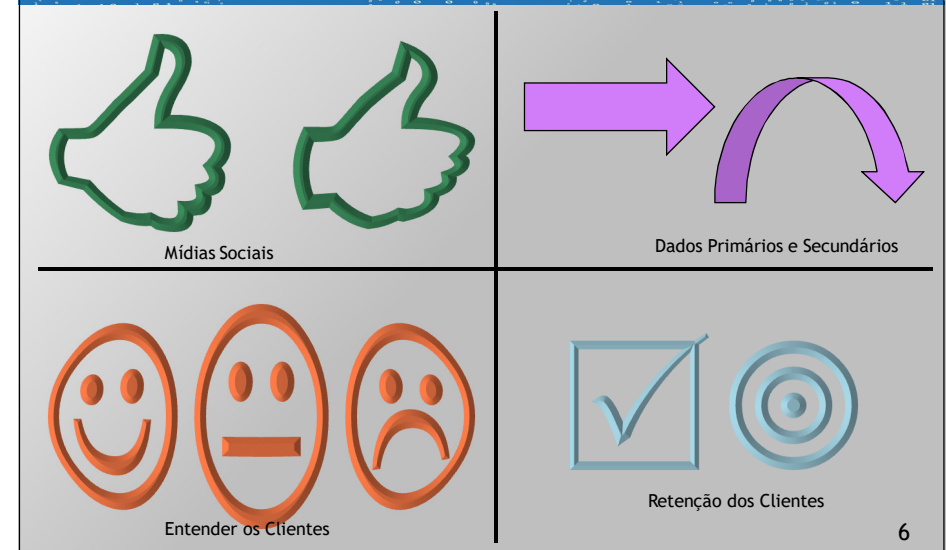
3

4

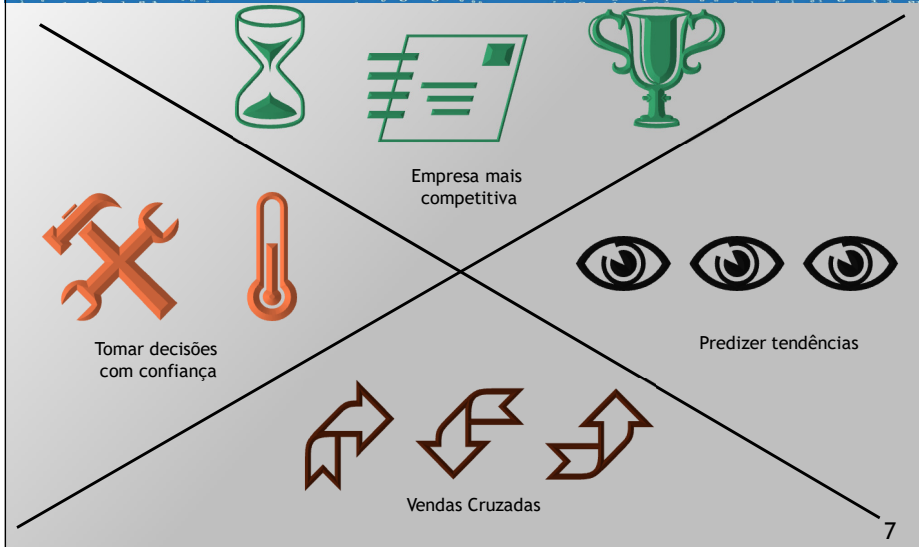
Grandes Sistemas de Dados - Motivações



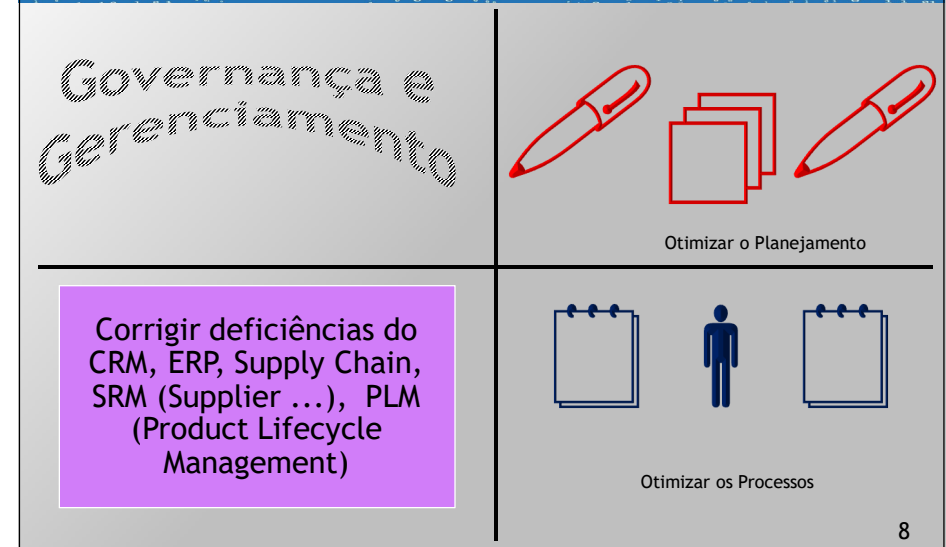
Grandes Sistemas de Dados - Motivações



Grandes Sistemas de Dados - Motivações



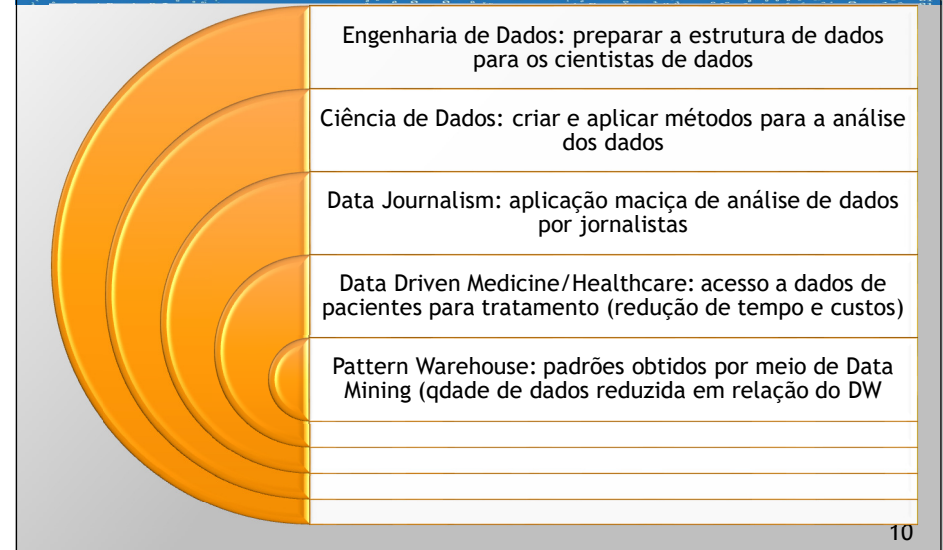
Grandes Sistemas de Dados - Motivações



Grandes Sistemas de Dados - Motivações



Grandes Sistemas de Dados - Motivações



Grandes Sistemas de Dados - Motivações

❖ Data Gravity: Dave McCrory

Data Gravity is a theory around which data has mass. As data (mass) accumulates, it begins to have gravity. This Data Gravity pulls services and applications closer to the data. This attraction (gravitational force) is caused by the need for services and applications to have higher bandwidth and/or lower latency access to the data.

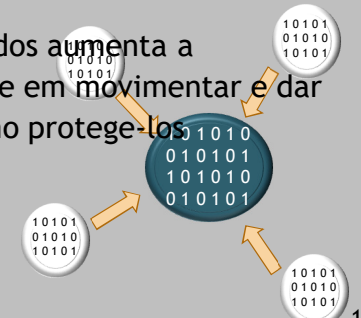
❖ **Massa de Dados** = Volume (tamanho do conjunto de dados) x Densidade (taxa de compressão)

❖ **Massa da Aplicação** = Volume (qtde de memória x qtde de disco) x Densidade (taxas de compressão x taxa de utilização de CPU)

Grandes Sistemas de Dados - Motivações

❖ Princípios de Data Gravity:

- Muitos dados atraem mais dados e por consequência atraem mais pessoas
- Grandes corpos de dados têm mais “massa” e maior atração
- Com maior quantidade de dados aumenta a responsabilidade e dificuldade em movimentar e dar acesso a tais dados, bem como protege-los



Grandes Sistemas de Dados - Definições

13

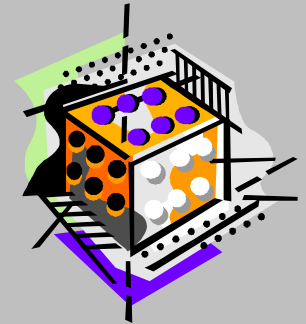
Grandes Sistemas de Dados - Definições

❖ Sistema de Base de Dados

❖ Sistema computadorizado de armazenamento de coleções de registros/fichas

❖ Funções básicas:

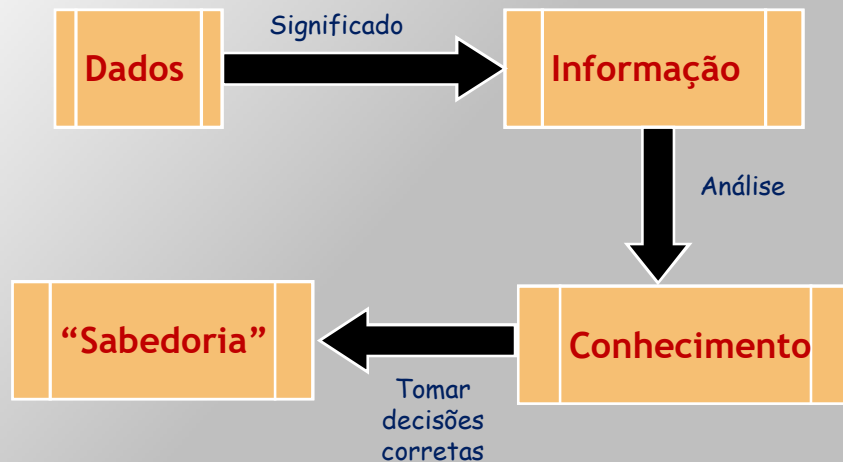
armazenar/recuperar/
atualizar/excluir dados



Dados x **Informações**

14

Dados, Informação, . . .



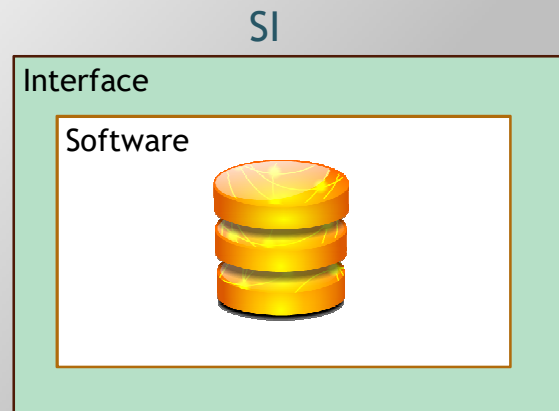
15

Grandes Sistemas de Dados - Definições

Sistemas de Informação

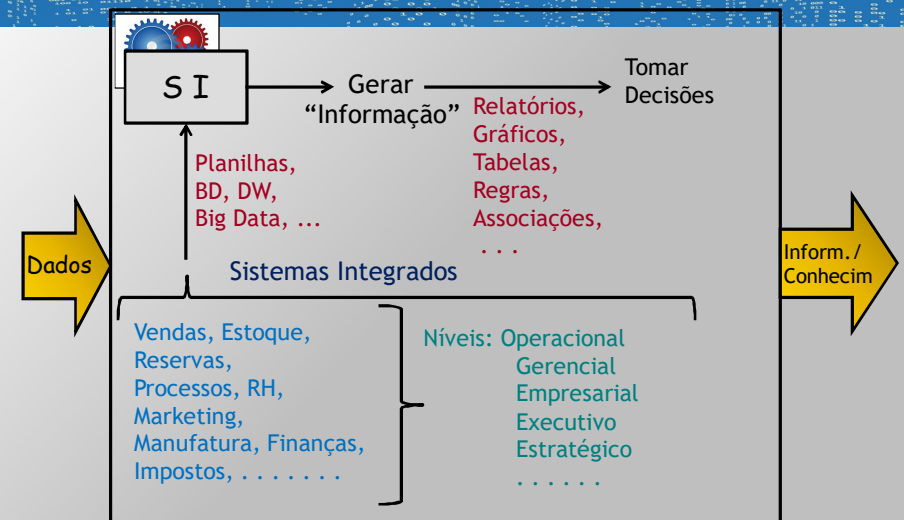
16

Sistemas de Informação



17

Sistemas de Informação



18

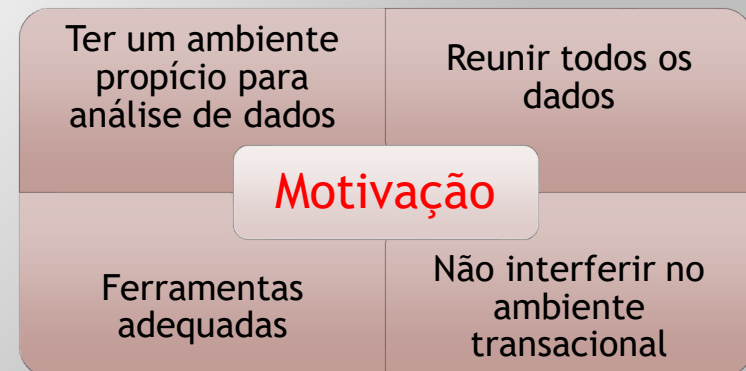
Sistemas de Informação

Definição: Conjunto de componentes computacionais inter-relacionados que coletam, processam, armazenam e distribuem informação para apoiar o processo de tomada de decisões e o controle de uma organização

Contêm informações sobre pessoas, lugares e coisas significativas à organização e ao ambiente ao seu redor

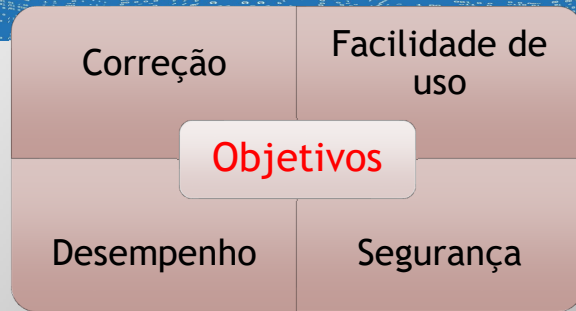
19

Sistemas de Informação



20

Sistemas de Informação



❖ Implementação

- SGBD / Ferramentas DW/Big Data/Data Lake
- Linguagem/ambiente de programação
- Interface Web

21

Sistemas de Informação



Características desejadas de um SI

- Flexibilidade
- Suportar modelos analíticos
- Manter trilha de alternativas e consequências
- Refletir entendimento dos grupos da empresa
- Sensível aos procedimentos operacionais

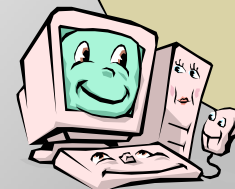
22

Sistemas de Bancos de Dados Aplicações Convencionais

23

Aplicações Convencionais / Tradicionais

Transações Bancárias,
Bolsas de Valores,
Vendas,
Reservas de Passagens,
Controle de Estoque,
Folha de Pagamento,
...



24

Aplicações Convencionais / Tradicionais

Dados uniformes

Grande número de registros “pequenos”

Dados alfanuméricos

Transações curtas

Modelagem consagrada

Softwares maduros

25

Bancos de Dados Relacionais

ORACLE® 1979

IBM DB2 1983

PostgreSQL 1987

Microsoft SQL Server 1988

MySQL® 1995

26

Bancos de Dados Orientados a Objetos

27

BD Orientados a Objetos

- ❖ Pós-Relacional
- ❖ Dados mais complexos \Rightarrow documentos, imagens, voz, vídeo - Internet/ Intranets
- ❖ Utilização da Orientação a Objetos
 - Diagramas UML
 - Modelagem OO
 - SGBDOO

VERSANT

OBJECTIVITY/DB®

GEMSTONE S™

InterSystems®
CACHE

28

Bancos de Dados Pós-Relacionais



29

BD Pós-Relacionais

❖ NoSQL ou NewSQL

➤ Chave-Valor



➤ Documentos



➤ Colunas



➤ Grafos



30

Grandes Sistemas de dDdos Características Básicas

31

GSD - Características básicas

Grande volume de dados



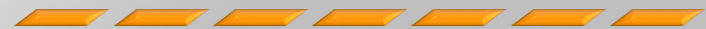
Sistemas para análise de dados



Tipos variados de dados



Conjunção de dados de várias origens



Isolar o ambiente transacional



Necessidade de novas equipes de desenvolvimento



32

GSD - Características Básicas

Segurança de dados é indispensável

Dados históricos

Grande disponibilidade

Excelente desempenho

Dados não padronizados

33

GSD - Características Básicas

Dados

Processamento

Armazenamento

Controle Geral

Decisões Estratégicas

34

Data Warehouse e Data Mining

35

Data Warehouse

Explorar
Dados



Obter
Informações



Tomar
Decisões
Acertadas

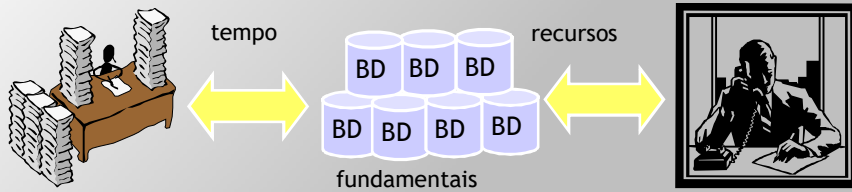
Obter Melhor Desempenho nos Negócios

Principal Origem dos Dados: BDs Operacionais

Análise dos Dados: Necessidade de Ferramentas
de Suporte à Decisão

36

Data Warehouse

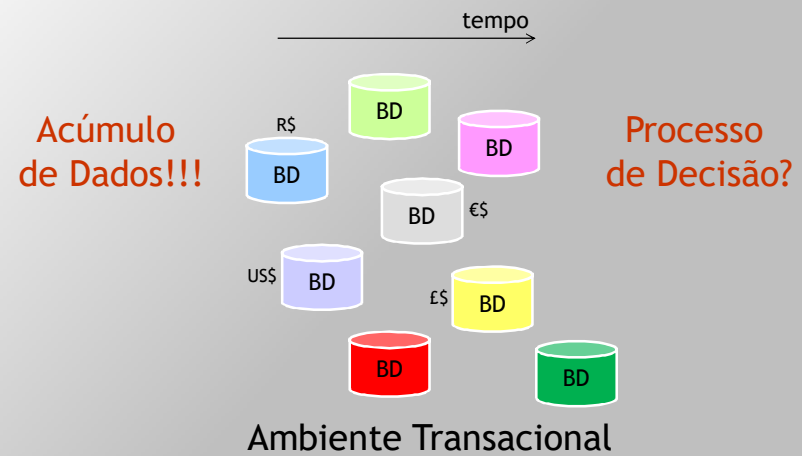


❖ Tipos de Dados

- Registros tradicionais
- Documentos completos
- Planilhas
- Figuras, gráficos, desenhos técnicos, . . .

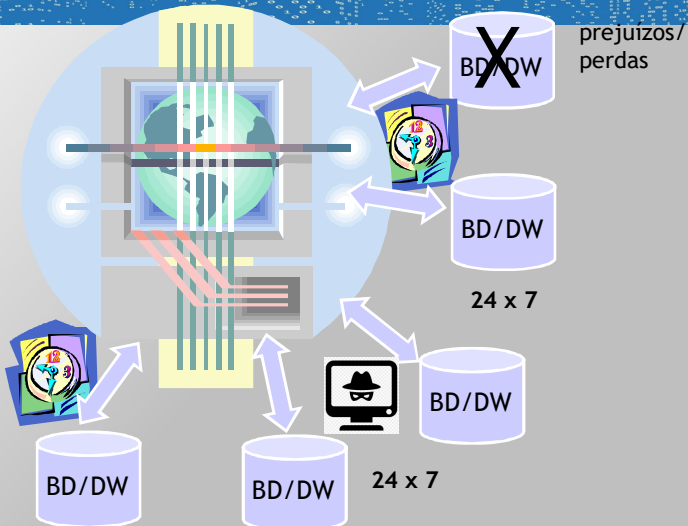
37

Motivação - fatores internos



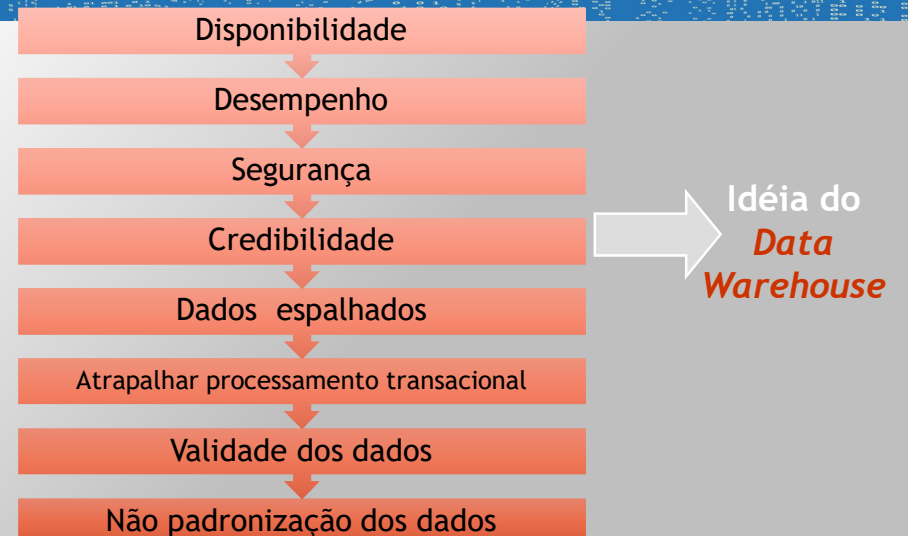
38

Motivação



39

Motivação



40

Data Warehouse

Por que DW?

Ambiente
distinto

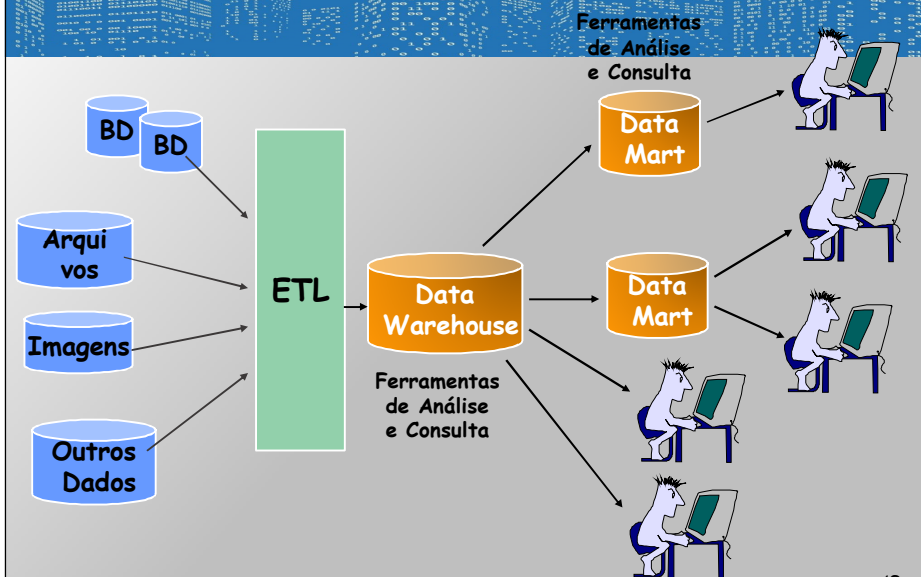
Dados
históricos

Decisões:
levar em
conta todas
informações
disponíveis

Rapidez

41

Data Warehouse



42

Data Warehouse

Suporte à tomada de
Decisões

Tendências de consumo
Aumento/redução da produção
Regulação de estoques
Campanhas publicitárias de
incentivo ao consumo
Eficiência do pessoal
Simulação de alterações
Estratégias de venda
...



43

Data Warehouse

Unificar/
Converter
Dados?

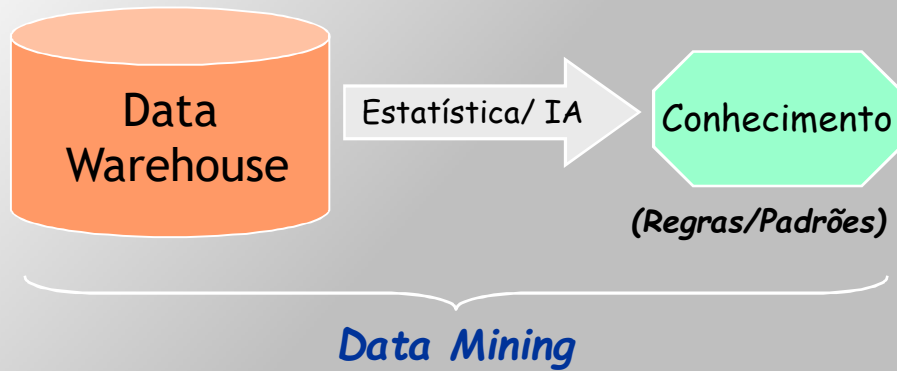
Sumarizar
Dados?
Quais?

Questões

Ferramentas OLAP

44

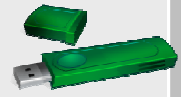
Data Mining



45

Data Warehouse

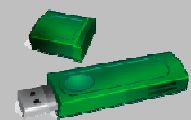
- Pentaho 
- SQL Server BI Edition 
- Power BI 
- Oracle Advanced Analytics/OLAP 
- IBM InfoSphere 
- Microstrategy 
- Tableau 
- Qlik 
- Talend 
-



46

Data Mining

- ❖ Weka, 
- ❖ Rapid Miner, 
- ❖ SQL Server Analytics Service 
- ❖ IBM Predictive Analysis 

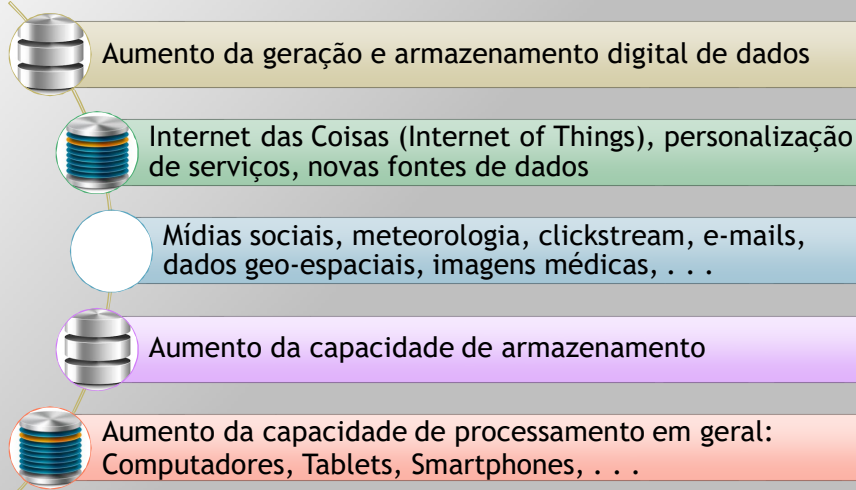


47

Big Data

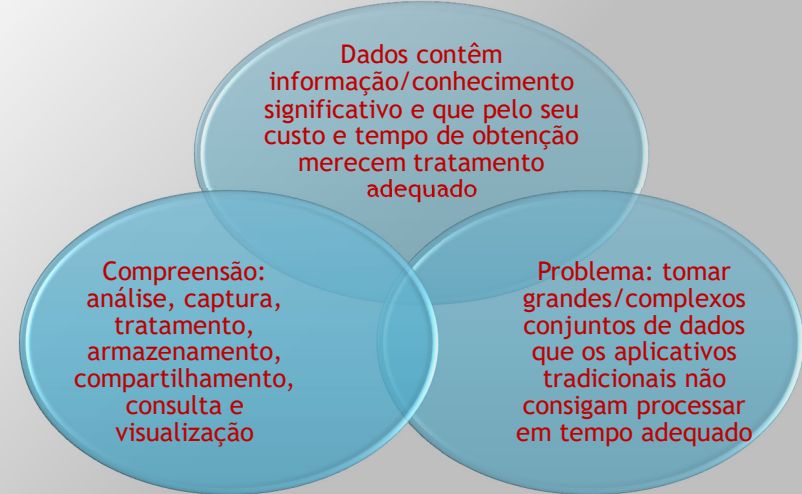
48

Big Data - Motivação



49

Big Data - Motivação



50

Big Data - Características

- ❖ Nova geração de tecnologias e arquiteturas que visam extrair valor a partir de bases de dados muito grandes, com grande variedade de origens, com grande velocidade de captura e análise
- ❖ “Big” no Big Data não é somente o tamanho da base de dados
- ❖ Não se refere a nenhum patamar de volume de dados
- ❖ Os dados são “big” quando seu tamanho se torna elemento ativo do problema

51

Data Lake

52

Data Lake

Armazenar dados sem se preocupar como vai ser seu uso

Requisito: armazenar também os metadados

Quando vai analisar os dados → necessário adotar algum esquema

53

Data Lake

Dados estruturados, semiestruturados e não estruturados, todos armazenados na forma “nativa”

“Ingestão” de dados em qualquer formato

Disponibilidade imediata dos dados

O trabalho fica para o momento da análise

Tipos de dados: Web, Sensores, Logs, Redes Sociais, Imagens, ...

54

Data Lake - Etapas

1. Ingestão de dados - entrada de dados

2. Seleção dos dados a serem analisados

3. Aplicação de ferramenta analítica de dados - saída

4. Utilização dos resultados da análise

55

Grandes Sistemas de Dados Áreas de Aplicação

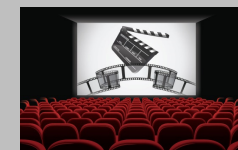
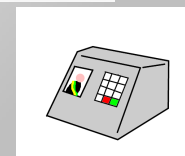
56

Áreas de Aplicação



57

Áreas de Aplicação



58

Grandes Sistemas de Dados Qualidade de Dados

59

GSD - Qualidade de Dados

Impactos nos lucros
(DW: tomada de decisões estratégicas)

Exemplos: ataques a locais errados, perda de equipamentos, questões médicas

QD: pode ser melhorada pelo incentivo ao uso dos dados (analogia com fenômeno biológico da atrofia - Ken Orr)

Comparação da má QD com a presença de colesterol no sangue (Bill Inmon)

60

GSD - Qualidade de Dados

Baixa QD: principais causas

- Digitação inicial
- Degradação → obsolescência
- Uso incorreto: não conhecimento do significado/má interpretação
- Reestruturações frequentes: troca de BD, migração, ...
- Baixo nível de monitoramento

61

Grandes Sistemas de Dados Segurança da Informação

62

GSD Segurança da Informação

Norma ABNT NBR ISO/IEC 27002 Tecnologia da Informação - Técnicas de Segurança - Código de Prática para a Gestão da Segurança da Informação

Lei Geral de Proteção de Dados Pessoais (LGPD) de 2018

Acessos não autorizados, alteração ou destruição intencional de dados

Controle de acesso físico e lógico

Estabelecer política de acesso adequada (controle de senhas, etc.)

Cuidado com ameaças internas de furto de dados

63

Governança e Gerenciamento de Dados

64

Governança de Dados

- Definição: organização de pessoas, processos e tecnologia para habilitar uma organização a alavancar os dados como recurso
- Visa obter maior taxa de sucesso, entregar mais valor e reduzir tempo de entrega de projetos
- Prover as condições para um gerenciamento eficiente das atividades de TI
- Priorizar e direcionar os objetivos
- Monitorar o ambiente de TI

65

Gerenciamento de Dados

Colocar em prática as linhas mestras da Governança

Aplicar controles e processos para alcançar os objetivos estratégicos estabelecidos na Governança

Gerenciamento

Vai de ações estratégicas até ações operacionais

Tratar a informação como um recurso muito valioso

66

Governança de Dados - Focos

Política, Normas e Estratégia

- Identificar responsáveis e estabelecer direitos de decisão

Qualidade de Dados

- Definir objetivos
- Monitorar a qualidade

Privacidade, Conformidade e Segurança

- Proteção de Dados
- Avaliar riscos e definir controles para seu gerenciamento

Planejamento Estratégico

- Determinar necessidade dos dados
- Avaliar o nível de maturidade do gerenciamento dos dados

67

Governança de Dados

Métricas

- Valor dos dados
- Custo do gerenciamento de dados
- Número de profissionais envolvidos
- Cumprimento dos objetivos
- Maturidade do processo de gerenciamento dos dados

68

Atores do Sistema

Recursos Externos e Restrições



69

Funções Gerenciais

❖ Interpessoais

- Símbolo
- Líder
- União

❖ Informacionais

- Centralizar
- Disseminar
- Porta voz

❖ Decisão

- Organizar
- Tratar distúrbios
- Alocar recursos
- Negociar



CIO - Chief Information Officer

- Gerente senior responsável pelos SI

70