

# TCC EM SISTEMAS DA INFORMAÇÃO

Aula 5- Eixo temático 4 – Gestão da Informação em grandes volumes de dados



**Estácio**

## Objetivo desta segunda aula

- apresentaremos o quarto eixo temático do curso:
  - Gestão da Informação em grandes volumes de dados



## Conteúdo Programático desta aula

- Identificar o quarto eixo temático da área de TI.
- Reconhecer as principais características desse eixo temático a fim de produzir seu trabalho de conclusão de curso.



## Grande Volume de Dados

Para entendermos esta área de pesquisa, temos primeiro que verificar o que aconteceu com a área da tecnologia da informação nestes últimos anos.

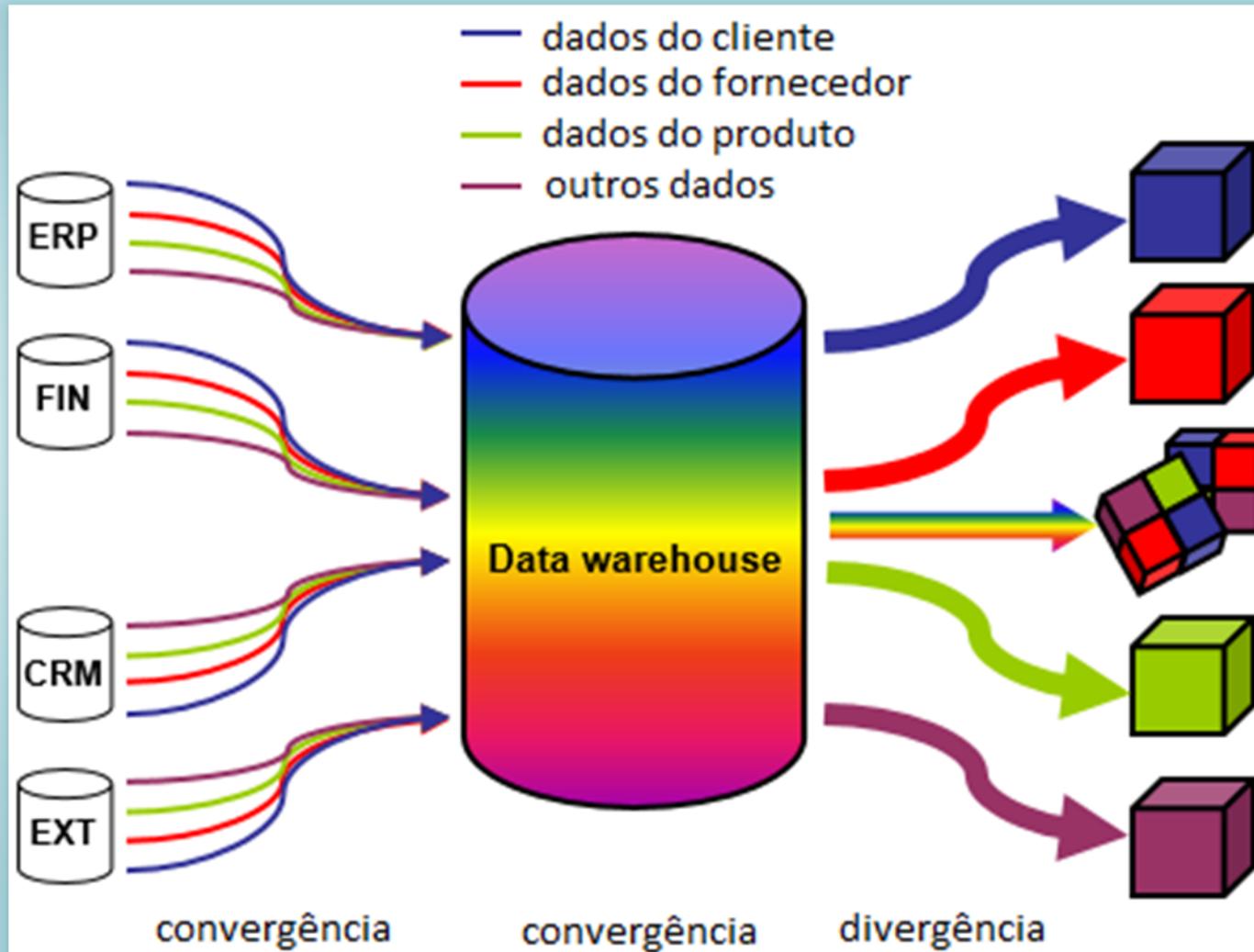
Com a solidificação dos grandes Data Warehouse, a informática teve um novo problema que é tratar este grande volume de dados. Vamos entender este problema.



## Data warehouse

É um banco de dados que não sofre atualização. É usado para geração de relatórios e análise de dados. É um grande repositório de dados onde tudo que ocorre dentro de uma empresa é armazenado e os dados não são descartados, principalmente os dados históricos.

## Data warehouse



## Data Warehouse

O Data Warehouse criou uma nova cultura no armazenamento de dados, pois ao se ter poder computacional de se armazenar todas as informações e de se poder tratar esta imensidão de dados, a computação passou para um novo nível, o nível da descoberta de padrões de comportamento.

O Data Mining ou garimpagem de dados é a área de pesquisa que busca padrões nestes grandes latifúndios de dados. Só que as pesquisas avançaram para um nível muito mais acima, onde os dados são armazenados em todas as operações.

## Datamining

Com este armazenamento ou registro de tudo que é feito na Internet ou nos sistemas, temos grandes bases de dados para podermos determinar novos padrões de comportamento dos indivíduos de nossa sociedade. Este é o grande desafio do Big Data.



## Big Data

- É um conceito na tecnologia da informação onde busca definir o armazenamento de grandes massas de dados, com maior velocidade. Este conceito está suportado em 5V's, que são: velocidade, volume, variedade, veracidade e valor.



## Big Data - suas características

### Velocidade

De acordo com o Gartner group, a velocidade define o tempo em que os dados estão sendo produzidos e também quão rápido estes dados podem ser tratados para enviar a resposta. O grande desafio para a maioria das organizações é o tempo de reação devido a velocidade como os dados crescem.

## Big Data - suas características

### Variedade

A diversidade de formatos que os dados se encontram hoje em dia é também um grande desafio. Áudio, vídeo, sistemas OLAP, arquivos texto, e-mail, coleta de dados através de sensores e medidores, dados do mercado financeiro, transações financeiras, banco de dados tradicionais são exemplos das diversas formas de coleta de dados. Foi estimado que 80% dos dados de uma empresa é não numérico. Só que todos estes dados precisam ser usados para a geração de padrões para a tomada de decisão.

## Big Data - suas características

### Volume

Como estamos armazenando todos os dados, e não mais sofrendo atualizações, quer dizer, armazenando os dados históricos, o volume de dados aumenta sempre. Além disso, outros fatores também auxiliam para este aumento. Redes sociais, aumento de dados de sensores que coletam dados. Como hoje em dia, o custo de armazenamento é decrescente, outras questões surgem, como a de se avaliar a relevância dos dados armazenados. Se realmente eles deveriam ser armazenados.

## Aplicações e objetivos

- A utilização do Big Data busca resolver problemas socioeconômicos, melhoria na qualidade de vida, ou a utilização sustentável do ambiente em que vivemos.
- Um ótimo exemplo sobre o tratamento de dados é retratado no filme “Moneyball” (O homem que virou o jogo), interpretado pelo ator Brad Pitt. Neste filme, o galã de Hollywood interpreta um gerente de um time de baseball, que contrata um “Nerd” que desenvolveu um software que analisa a performance de todos os jogadores da liga americana.

## Aplicações e objetivos



## Big Data - Aplicações e objetivos

- Hoje, o armazenamento de dados cresce exponencialmente, visto que a captura é feita em tempo real através de diversos dispositivos. Segundo dados fornecidos pela IBM, em 2008 foram gerados cerca de 2,5 quintilhões de bytes todos os dias e que cerca de 90% de todos os dados armazenados no mundo foram gerados nos últimos dois anos.

**Efetuar o tratamento destes dados é o grande desafio dos dias de hoje.**

## Desafios do Grande Volume de Dados



## Problemas que precisam ser resolvidos nesta área

- Identificar os dados através de Modelagem Computacional para reduzir a massa de dados
- Como utilizar processamento paralelo com recursos heterogêneos para aumentar o poder computacional (GRID)
- Estudar os sistemas que possuem larga escala de heterogeneidades, através de modelos e mecanismos de conciliação e integração
- Estudos de infra-estruturas adaptáveis e inteligentes
- Utilização de Recuperação da informação, através de fatores como localização do usuário, perfil de interesse, etc

## Problemas que precisam ser resolvidos nesta área

- Estudo da geração de descritores de conteúdo, multimodal e algoritmos para extração e indexação desses descritores
- Como trabalhar com a heterogeneidade na aquisição dos dados
- Estudo dos novos tipos de interfaces
- Como lidar com a confiabilidade dos dados e tratar a propriedade intelectual

## Business Intelligence (BI)



## Business Intelligence

- Esta nova área da TI veio para dar dinamismo e competitividade aos executivos das empresas, visto que através da análise de grandes massas de dados, os sistemas inteligentes das empresas municiam os gestores das corporações com informações pertinentes para a tomada de decisão. Estas informações saem da decodificação de padrões armazenados nas grandes bases de dados coletados e que sem o poder computacional de hoje, não seria possível perceber tais associações.

## Business Intelligence



- Este termo definido pelo Gartner Group, descreve as habilidades das organizações para armazenar e acessar os dados, explorando informações, analisando-as e desenvolvendo percepções e entendimentos sobre os assuntos que auxiliam no processo decisório dos gestores das empresas.

## Inteligência Empresarial

- A Inteligência Empresarial é acumulada pelas empresas durante anos, e estas informações são armazenadas, para que esta inteligência ganhe sustentação na sua vantagem competitiva. Geralmente, os coletores de BI coletam suas informações dentro das empresas.
- Nosso papel nesta área é desenvolver sistemas inteligentes capazes de identificar padrões nestes grandes repositórios de dados para que seja gerado relatórios úteis para a tomada de decisão.

## Fechamento

Nesta aula vimos:

- a) Data Warehouse
- b) Data Mining
- c) Conceito de Big Data
- d) As aplicações e desafios em Big Data
- e) Business Intelligence / Inteligência Empresarial

Fim

