



Fundamentos de Business Intelligence e Análise de Dados

BLOCO: B.I. E ANÁLISE DE DADOS

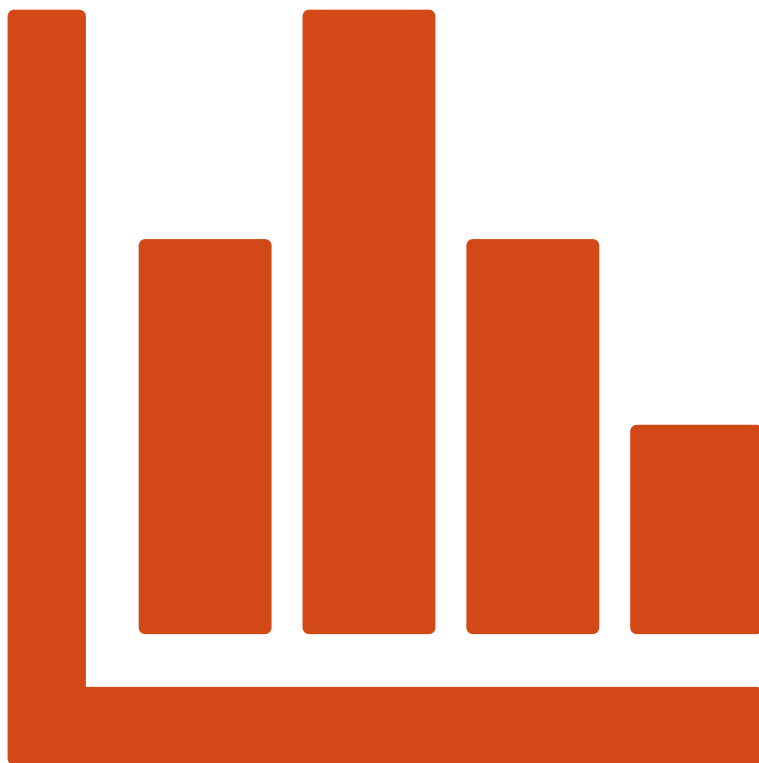
PROF. RODRIGO EIRAS, M.SC.

[ETAPA 1] AULA 2 – COMPREENDER O VALOR DA ANÁLISE DE DADOS PARA AS ORGANIZAÇÕES



Na aula anterior...

- Falamos da importância do Business Intelligence atualmente nas empresas
- Falamos da tomada de decisão
- Entendendo como um dado é transformado até se tornar uma informação e posteriormente conhecimento
- Entendemos ainda que correlação não implica causalidade
- Vimos alguns exemplo de B.I.



O Valor da Análise de Dados

Compreender o valor da análise de dados para as organizações



- Compreender o conceito de análise de dados e sua aplicação em bases de dados é uma necessidade primária para qualquer enfoque empresarial.
- Volumes de dados sendo armazenados por si só não teriam significância se não aplicados a uma análise, a uma pergunta, a um questionamento.
- Imagine uma empresa de telefonia que armazena as chamadas de todos os seus clientes para todos os respectivos destinatários.
- Do que valeria tais armazenamentos se não tivéssemos perguntas sobre o mesmo dataset?

Compreender o valor da análise de dados para as organizações

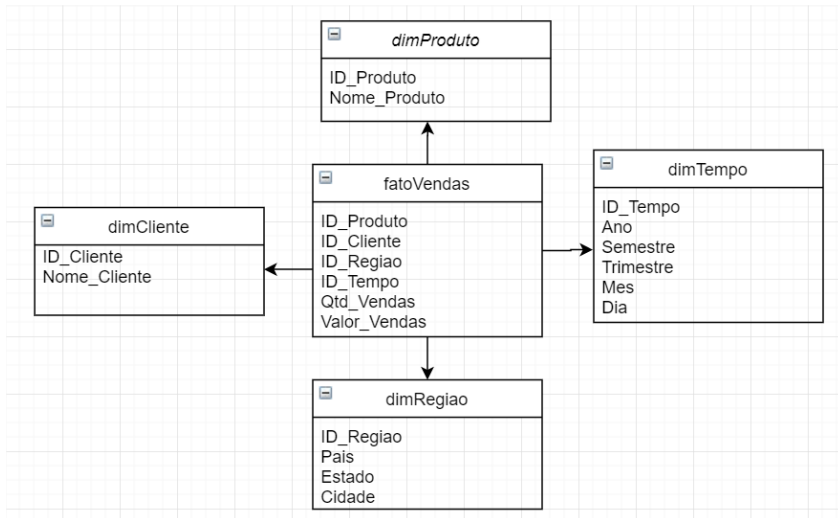
Cientes	Número	Resultado
Cliente 1	2121457852	VERDADEIRO
Cliente 2	2121457852	VERDADEIRO
Cliente 3	2121457852	VERDADEIRO
Cliente 4	2121457852	FALSO
Cliente 5	2121457852	VERDADEIRO
Cliente 6	2121457852	VERDADEIRO
Cliente 7	2121457852	VERDADEIRO
Cliente 8	2121457852	VERDADEIRO
Cliente 9	2121457852	FALSO
Cliente 10	2121457852	FALSO
Cliente 11	2121457852	FALSO
Cliente 12	2121457852	FALSO

- Uma pergunta básica:
 - Um cliente ligou para um determinado número?
- Essa pergunta teria um simples resposta: “SIM” ou “NÃO”. Ou seja, você estaria, através do consumo em uma base de dados, realizando uma consulta através de uma linguagem de programação.
- Essa resposta poderia ter um impacto importantíssimo para um critério de auditoria na organização, por exemplo.
- Assim, você verá com detalhes, como tais informações agregaram valor organizacional.



Dados transacionais e informações

Dados transacionais e informações



- Eles são caracterizados pela realização de transações com curta duração e pequeno volume de dados manipulados.
- O foco dos mesmos é “escrever” informações.
- A transferência gera uma transação que tem um valor, uma origem, limite de tempo, etc.
- São sistemas que dão garantia de integridade, ordem temporal de cada uma das transações.
- Um dos principais requisitos dos sistemas transacionais é a performance, ou seja, é necessário que a transação ocorra no momento em que foi requerida.



Big Data

The image shows a perspective view of a data center aisle. On the right side, there are tall, black server racks. The floor is made of light-colored square tiles, and the ceiling has a grid pattern with recessed lighting. The text 'Big Data' is written in a large, white, sans-serif font across the middle of the image. There are two thin white horizontal lines, one above and one below the text. On the rightmost rack, the text 'Blue Gene supercomputer' is visible vertically.

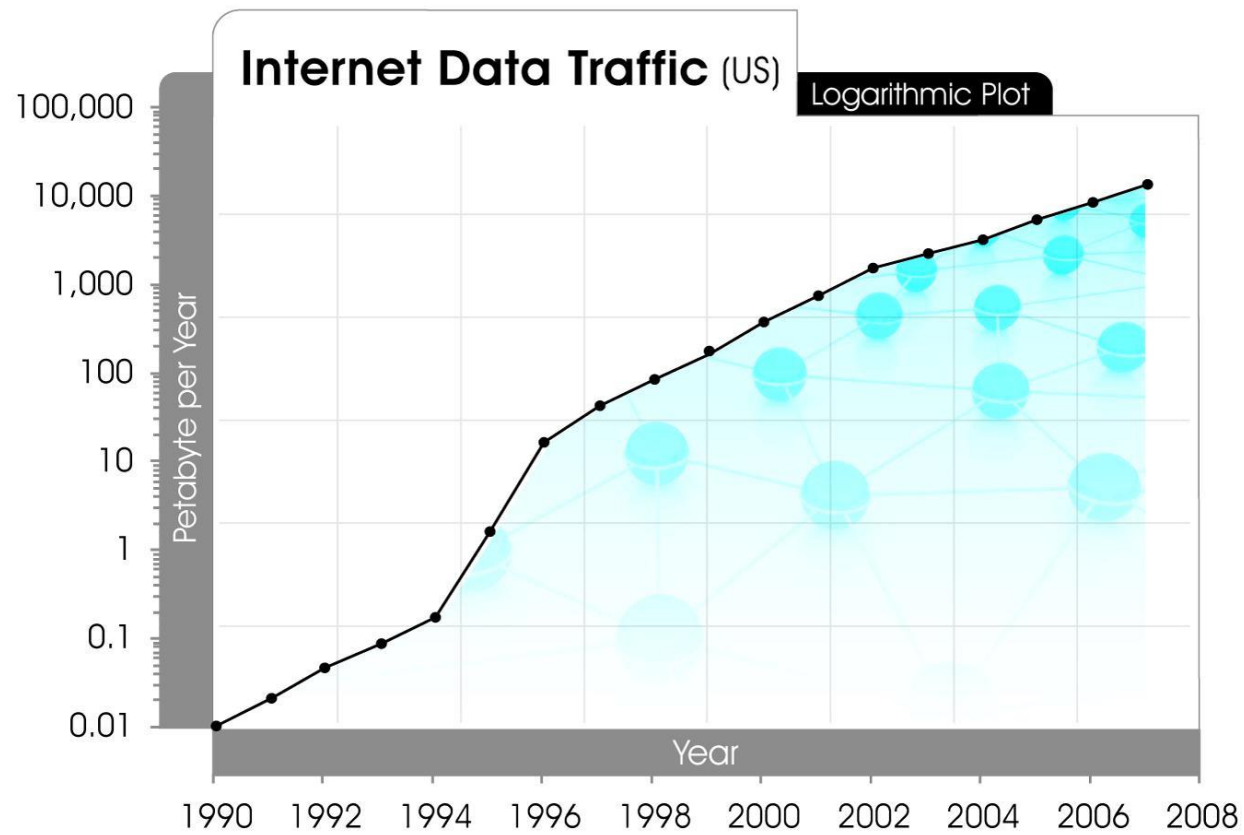
BIG DATA



Big Data?

- Alguma vez você já pensou na quantidade de dados que você já gerou?
- Pense nas informações que você já forneceu para bancos, empresas de cartão de crédito, companhias aéreas e, principalmente, para mídias sociais (fotos, curtidas e compartilhamentos no Facebook, por exemplo).
- Foi desse grande volume de dados gerados por todos nós que surgiu o que popularmente chamamos de Big Data.

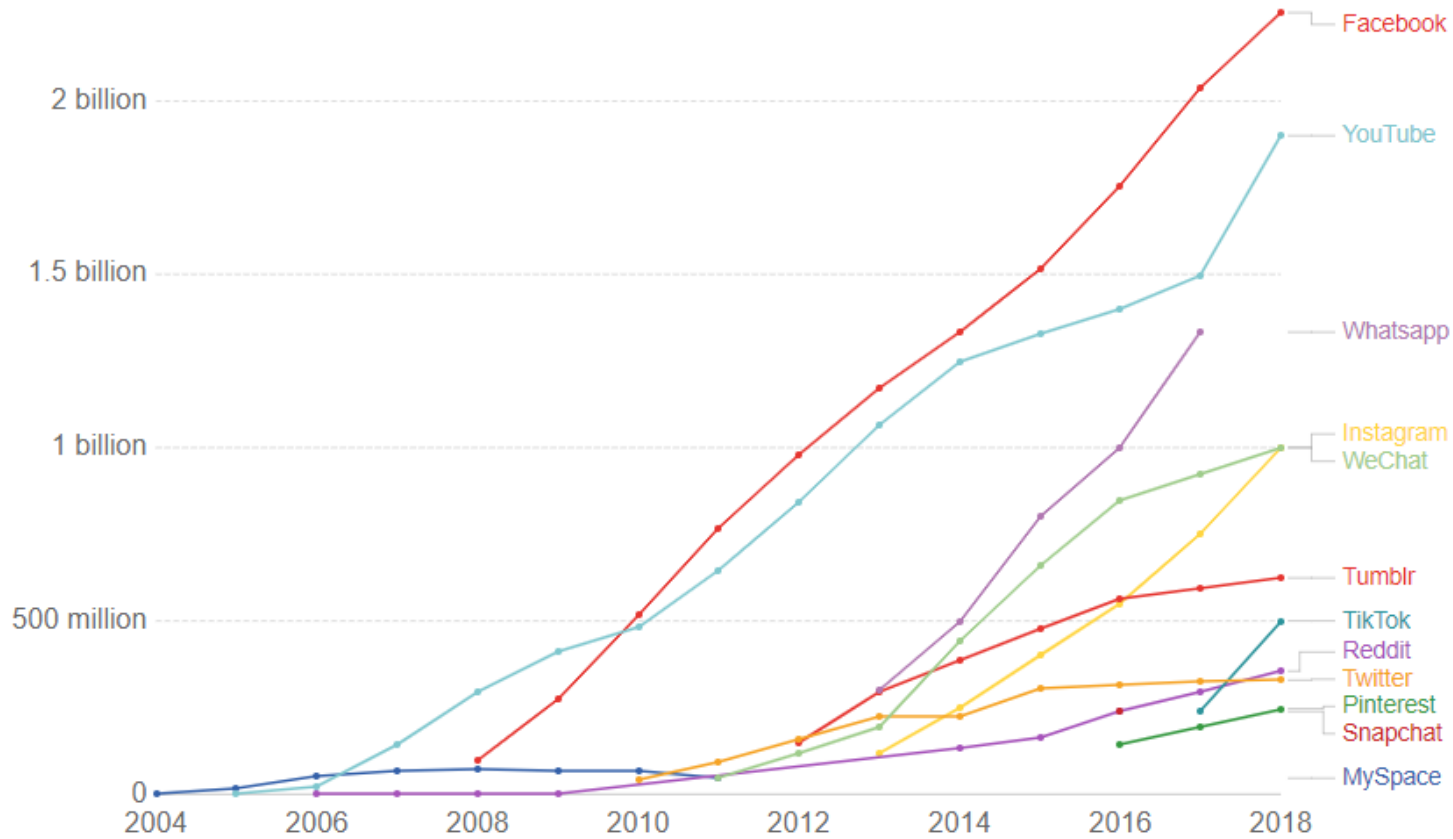




Big Data

Number of people using social media platforms

Estimates correspond to monthly active users (MAUs). Facebook, for example, measures MAUs as users that have logged in during the past 30 days. See source for more details.



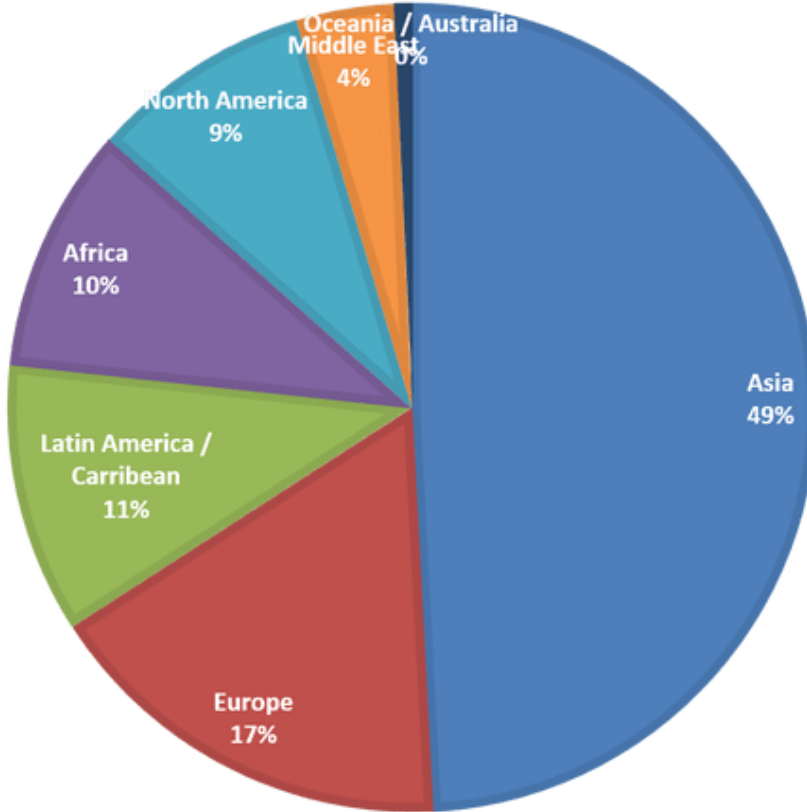
Source: Statista and TNW (2019)

CC BY

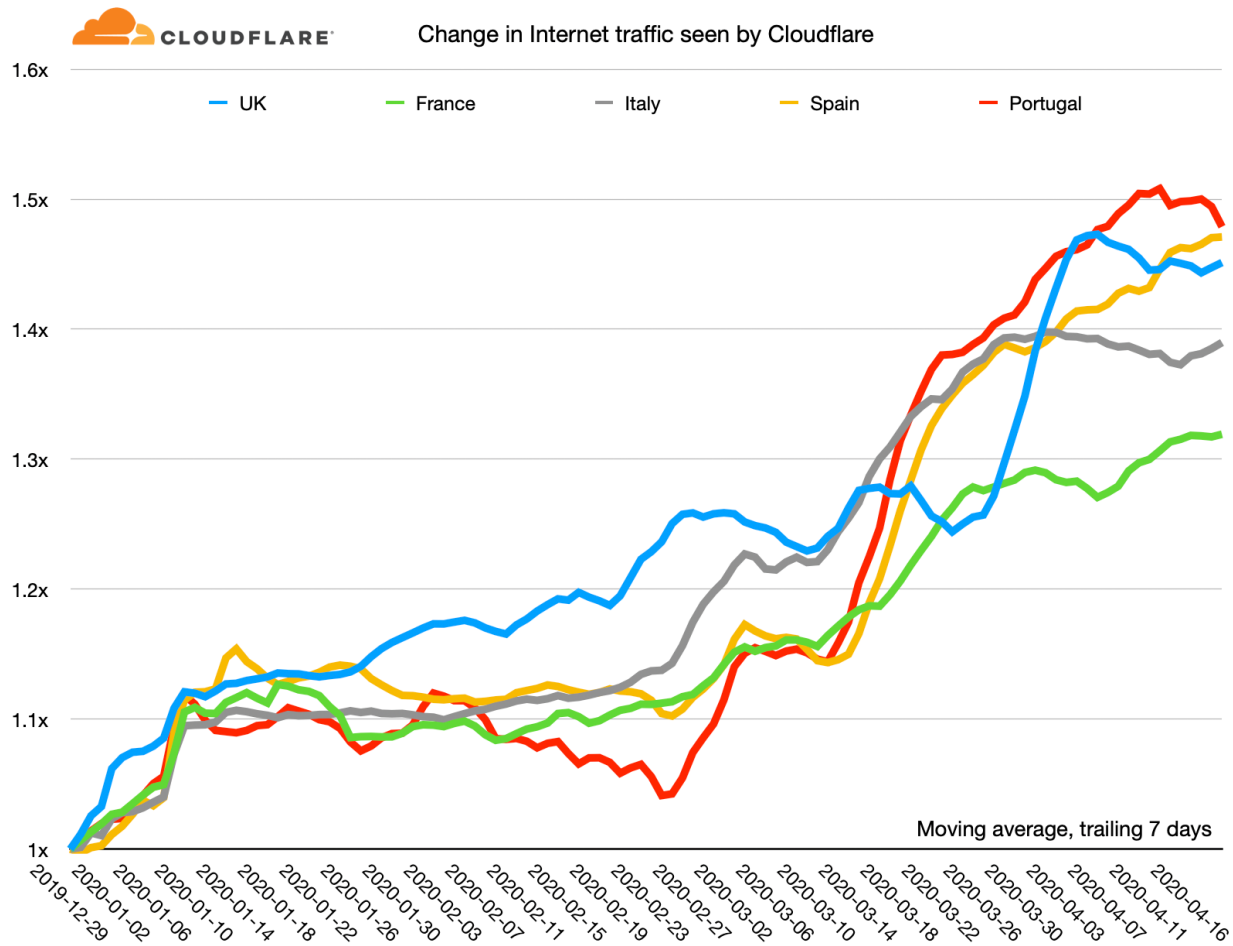
Big Data

GLOBAL INTERNET USER DISTRIBUTION (%)

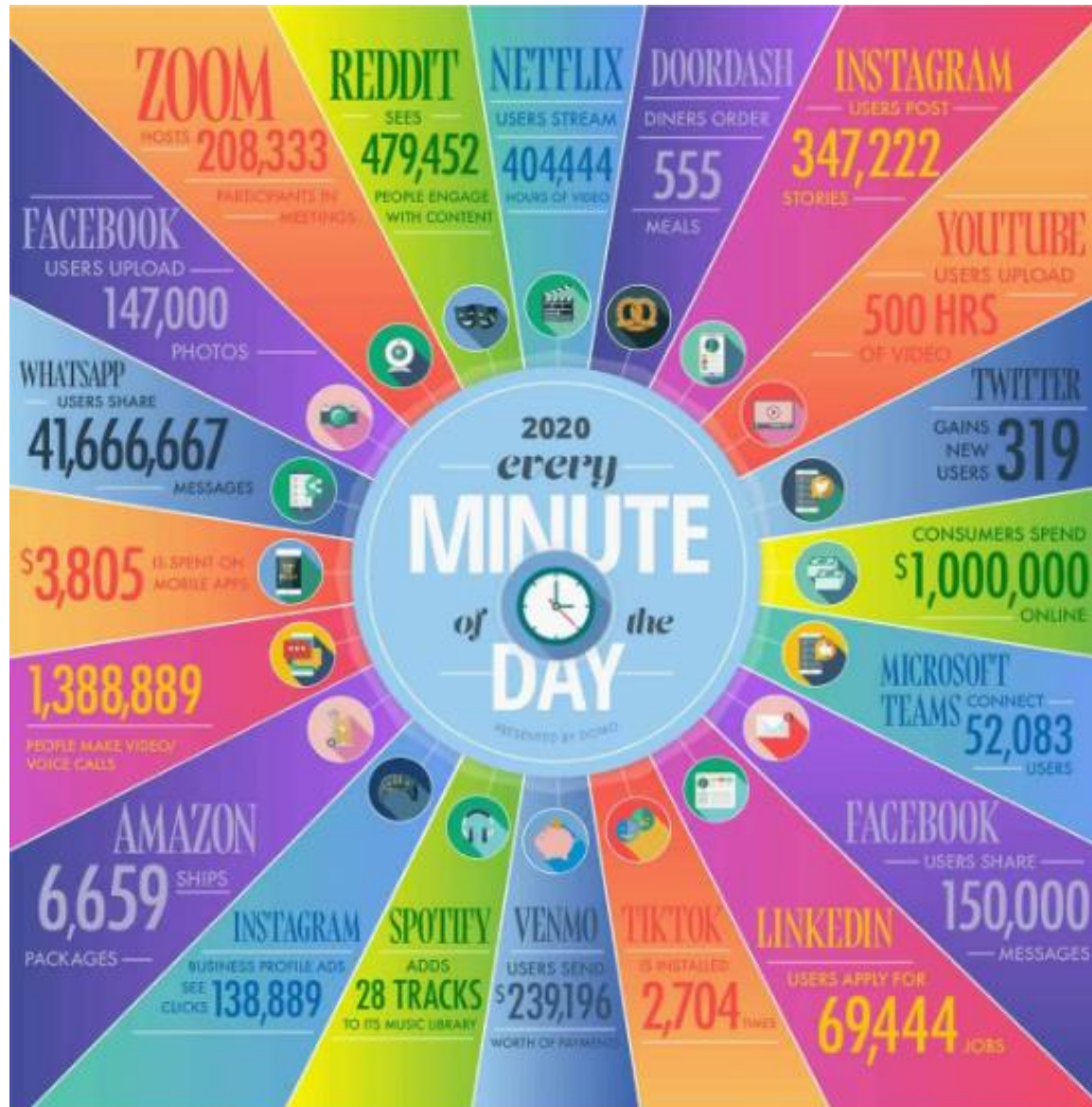
■ Asia ■ Europe ■ Latin America / Carribean ■ Africa ■ North America ■ Middle East ■ Oceania / Australia



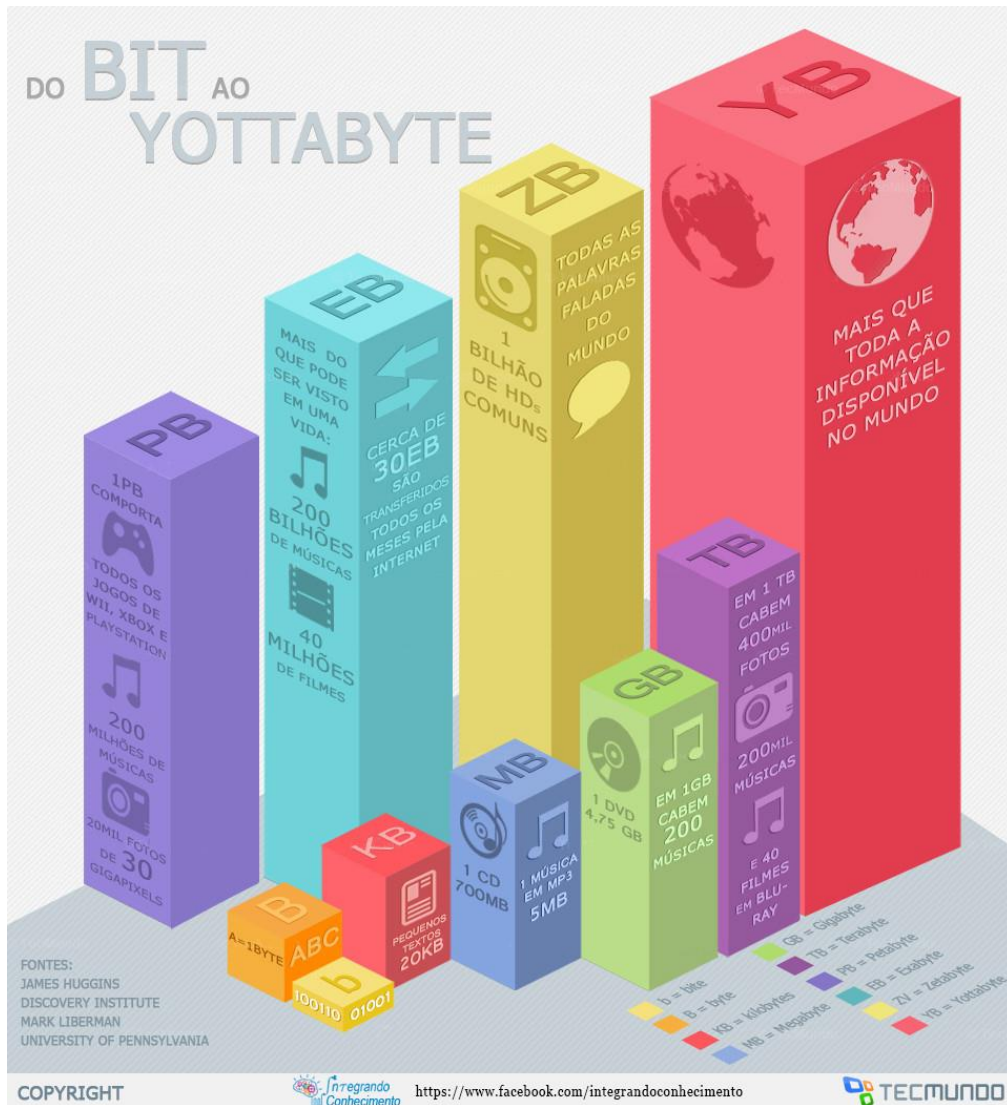
Big Data



Big Data & CoViD-19



Big Data & CoViD-19



Big Data – Como funciona

- A maioria desses dados foram gerados recentemente
- Nos últimos 3 anos, produzimos mais dados que em todo o resto da história da humanidade
- Na verdade, cerca de 90% foram gerados nesses últimos anos
- Você sabia que a cada dia, são gerados 2,5 quintilhões de bytes?
 - Para simplificar, um quintilhão é 1 seguido de 18 zeros

Big Data – Como funciona



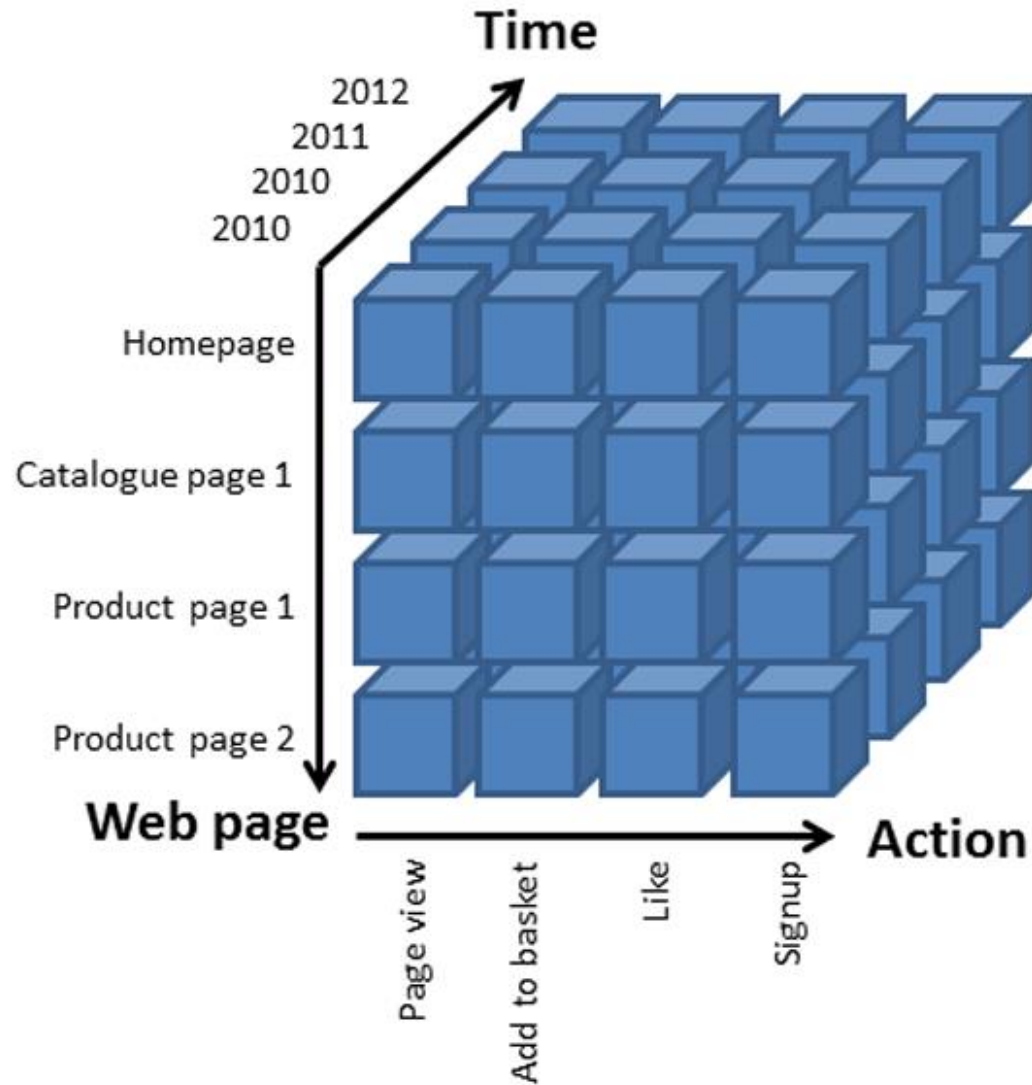
- No entanto, não se trata apenas de mídias sociais
- Por exemplo, diariamente, empresas geram inúmeras e diversas informações em um sistema ERP
- Máquinas equipadas com sensores, dados de geolocalização
- Dependendo do modo como esses dados são captados, seu formato vai variar muito, por exemplo, dados de mídias sociais tendem a ser não estruturados, muitos podem estar em formato de vídeo, comentários e por aí vai.

Sistemas transacionais - OLTP

What Is an OLTP Database?



- Sistemas transacionais, também conhecidos como sintéticos ou ainda OLTP – *Online Transactional Processing* - são aqueles que, como o nome sugere, baseiam-se em transações.
- Exemplos deste tipo de sistemas:
 - Sistemas Contábeis;
 - Aplicações de Cadastro;
 - Sistemas de Compra, Estoque, Inventário;
 - ERPs, CRMs.
- Os sistemas transacionais se caracterizam pela alta taxa de atualização, grande volumes de dados e acessos pontuais, ou seja, pesquisas cujo resultado seja de pequeno volume (até milhares de linhas, mas preferencialmente menos).

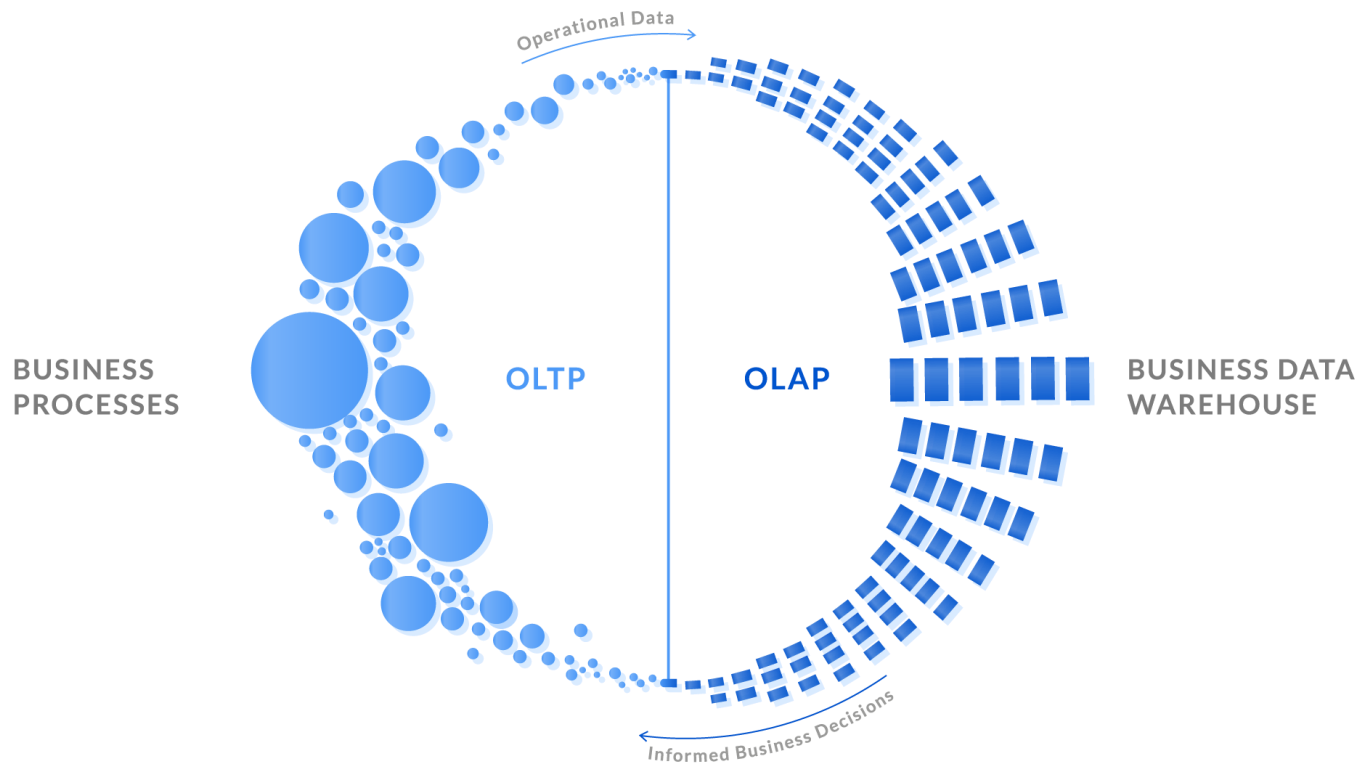


Sistemas analíticos - OLAP

- Já os sistemas analíticos, ou OLAP – Online Analytical Processing – se caracterizam por fornecer subsídio para tomadas de decisão
- A partir de análises realizadas sobre bases de dados históricas, por vezes com milhões de registros a serem totalizados

Característica	Sistemas Transacionais(OLTP)	Sistemas Analíticos(OLAP)
Atualizações	Mais freqüentes	Menos freqüentes
Tipo de Informação	Detalhes	Agrupamento
Quantidade de Dados	Poucos	Muitos
Precisão	Dados atuais	Dados históricos
Complexidade	Baixa	Alta
Consistência	Microscópica	Global
Exemplos	CRM, ERP, Supply Chain	MIS, DSS, EIS
Terminologia	Linhas e Colunas	Dimensões, Medidas e Fatos

OLAP vs OLTP



OLAP vs OLTP

OLAP vs OLTP – Resumindo!

Transacionais:

- A realização de atualizações frequentemente, mantendo os dados atuais;
- Informação detalhada com a maior granularidade possível (consistência microscópica);
- Pesquisas pontuais, portanto de baixa complexidade, relacionado ao negócio (do ponto de vista técnico, a pesquisa pode ser bem elaborada).

Analíticos:

- Uma vez que os dados são históricos, as atualizações não precisam ser tão frequentes. Por exemplo, numa comparação entre a produtividade de três filiais de uma empresa para um determinado produto nos últimos quatro meses, por mês, o dia de hoje ou mesmo ontem não é, em geral, de grande representatividade;
- As análises geralmente agrupam informações, sendo tais agrupamentos mais importantes neste contexto do que os dados detalhados. No exemplo do item anterior, o importante é a produção conjunta mensal, e não a produção de uma unidade particular do produto analisado.

Big Data – continuando...



- Muito se fala de Big Data como uma tecnologia revolucionária, principalmente para os negócios.
- Apesar de alguns exagerarem um pouco, de fato o Big Data tem afetado a maneira que gestores e empresas atuam.
- E se afeta os negócios, merece atenção.
- Big Data é um termo que surgiu devido à grande quantidade de dados gerados por sistemas internos de empresas, redes sociais, e muito mais.
- Todos esses dados gerados, quando analisados podem fornecer informações jamais imaginadas.

Big Data – os 5v's



- Para entender o que é Big Data, é preciso entender o conceito dos 5v's Big Data e como essa estrutura pode auxiliar sua empresa a gerar insights.
- Antes de mais nada, você sabe o que é Big Data?

Big Data Analytics – Impactos nas empresas





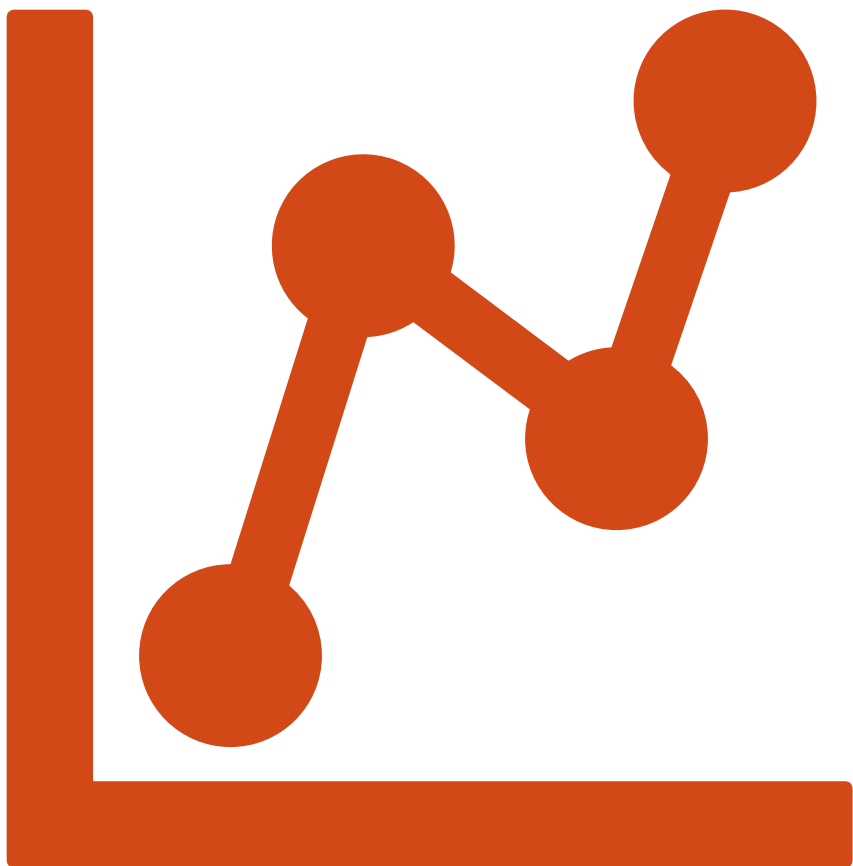
Big Data & Internet das Coisas (IoT)

Sem Big Data não há internet das coisas

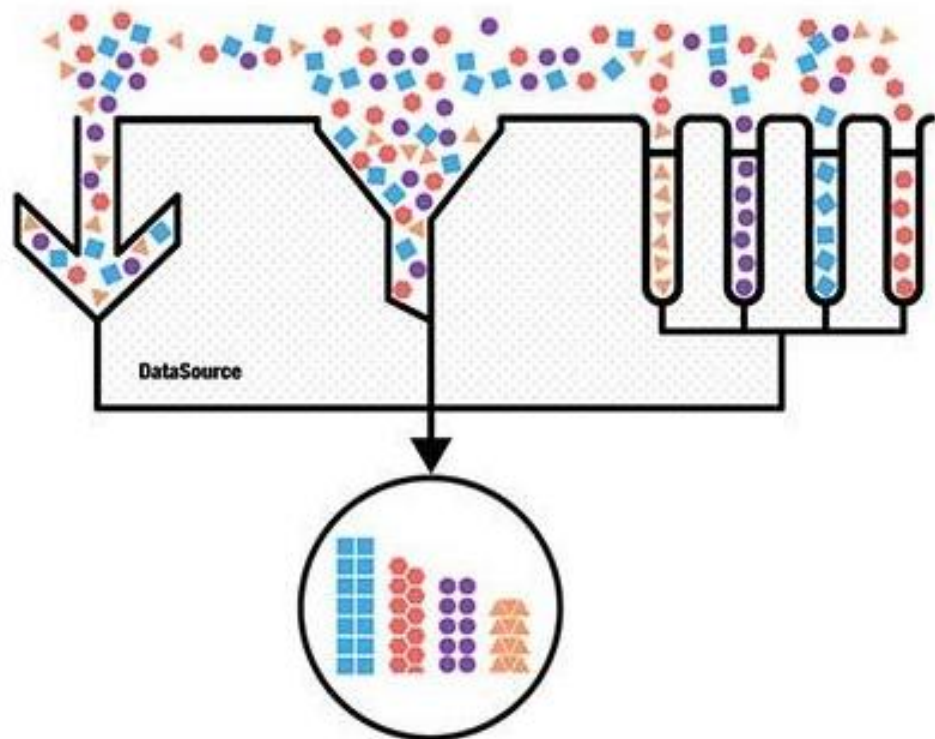


Big Data & Internet das Coisas (IoT)

<https://youtu.be/ZSdUopp4umg>



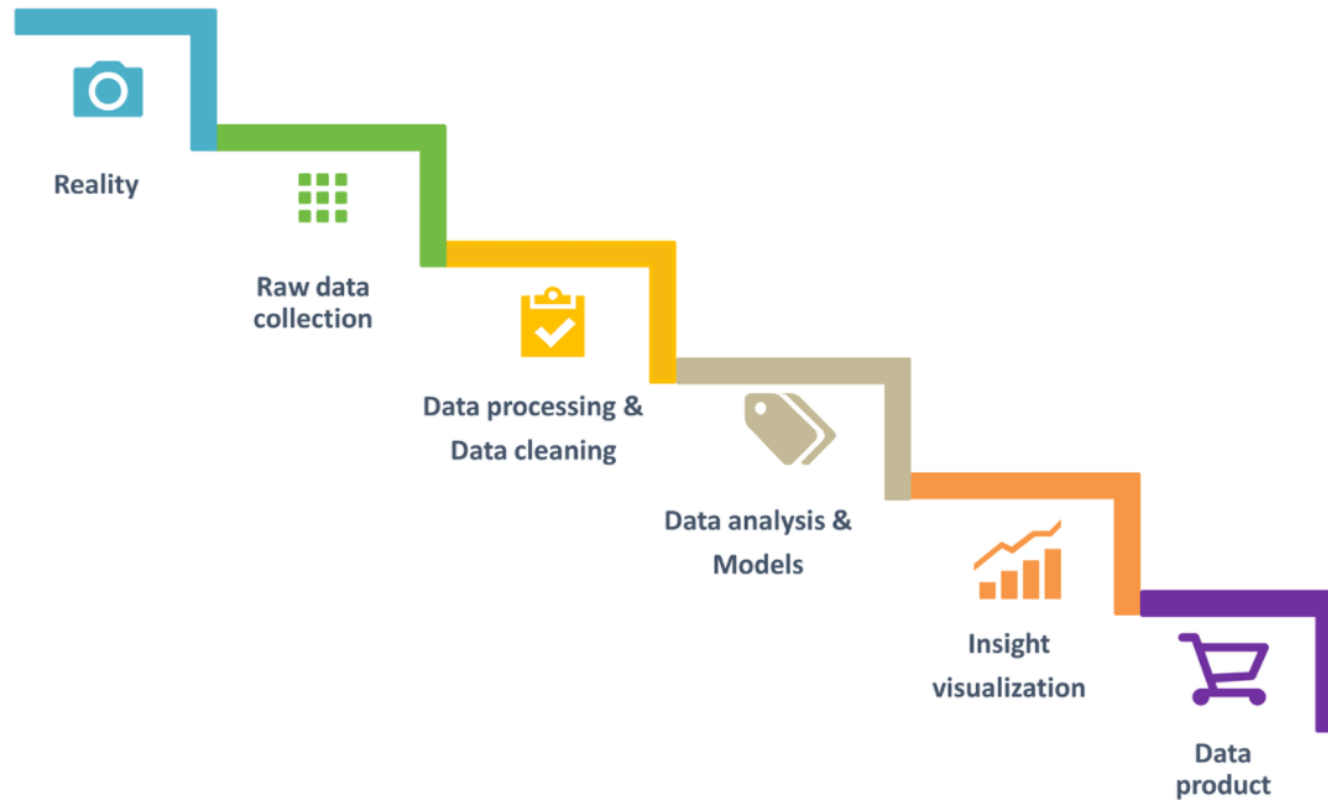
Conceito de Análise de Dados



Compreender o conceito de análise de dados

- Como sabemos, as fontes de dados são as mais diversas possíveis.
- Estamos sempre citando banco de dados quando consideramos nossos conteúdos, mas é importante identificar o conceito de análise de dados antes do conceito de banco de dados!

The process of data analysis



Compreender o conceito de análise de dados

- A análise de dados é o processo pelo qual se dá ORDEM, ESTRUTURA e SIGNIFICADO aos dados
- Consiste, basicamente, na transformação dos dados coletados em conclusões e/ou ligações úteis, seguindo os propósitos:
 - Descrever e resumir dados;
 - Identificar relações e diferenças entre as variáveis;
 - Comparar variáveis;
 - Fazer previsões.

Enfoque da análise de dados

◦ QUALITATIVO

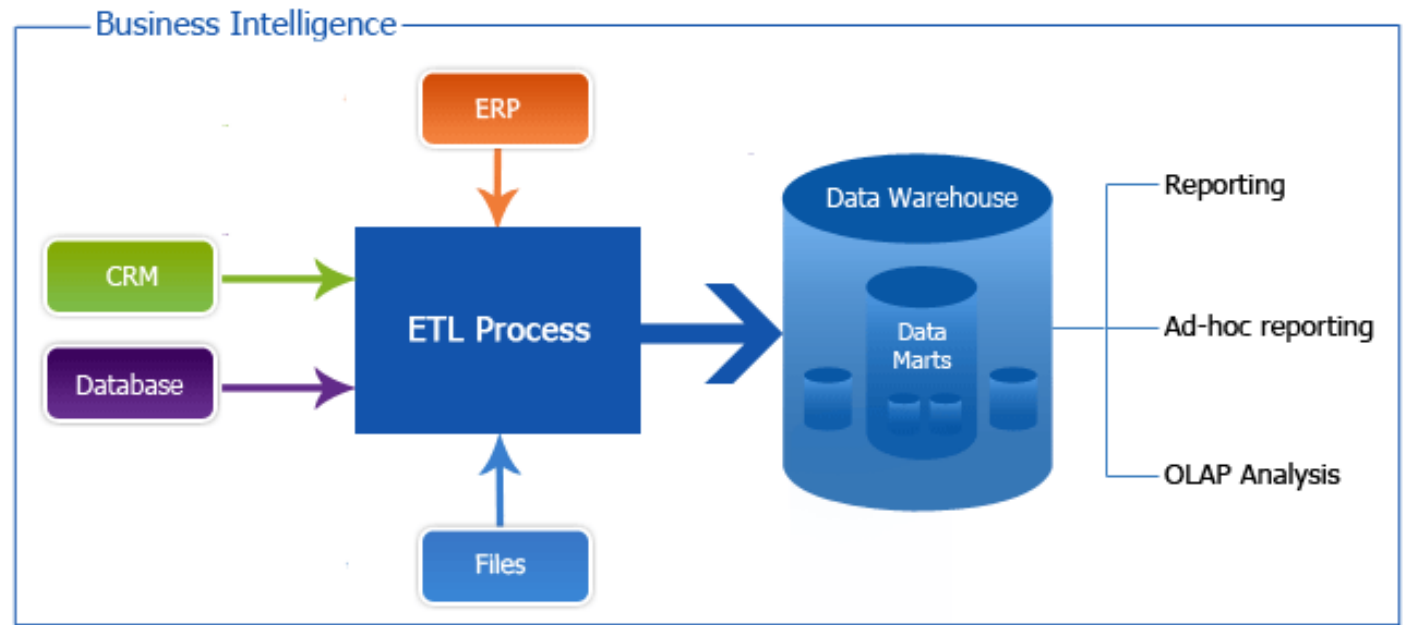
- Ajuda a definir hipóteses;
- Estuda as motivações;
- Fornece informações detalhadas a perguntas ou problemas sobre um projeto ou atividade;
- Indaga porque determinado fato ou problema está ocorrendo;
- É indutivo;
- É exploratório;
- Permite conhecer tendências, comportamentos, atitudes, etc;
- Não permite inferir os resultados para toda uma população.

◦ QUANTITATIVO

- Estuda ações ou intervenções;
- É dedutivo;
- Fornece dados para provar hipóteses;
- É conclusivo;
- Mede o nível das intervenções, tendências, atividades, etc;
- Produz informações quantificáveis sobre a magnitude de um problema, mas não fornece informações sobre o motivo do fato estar ocorrendo;
- É possível inferir os resultados para toda uma população.

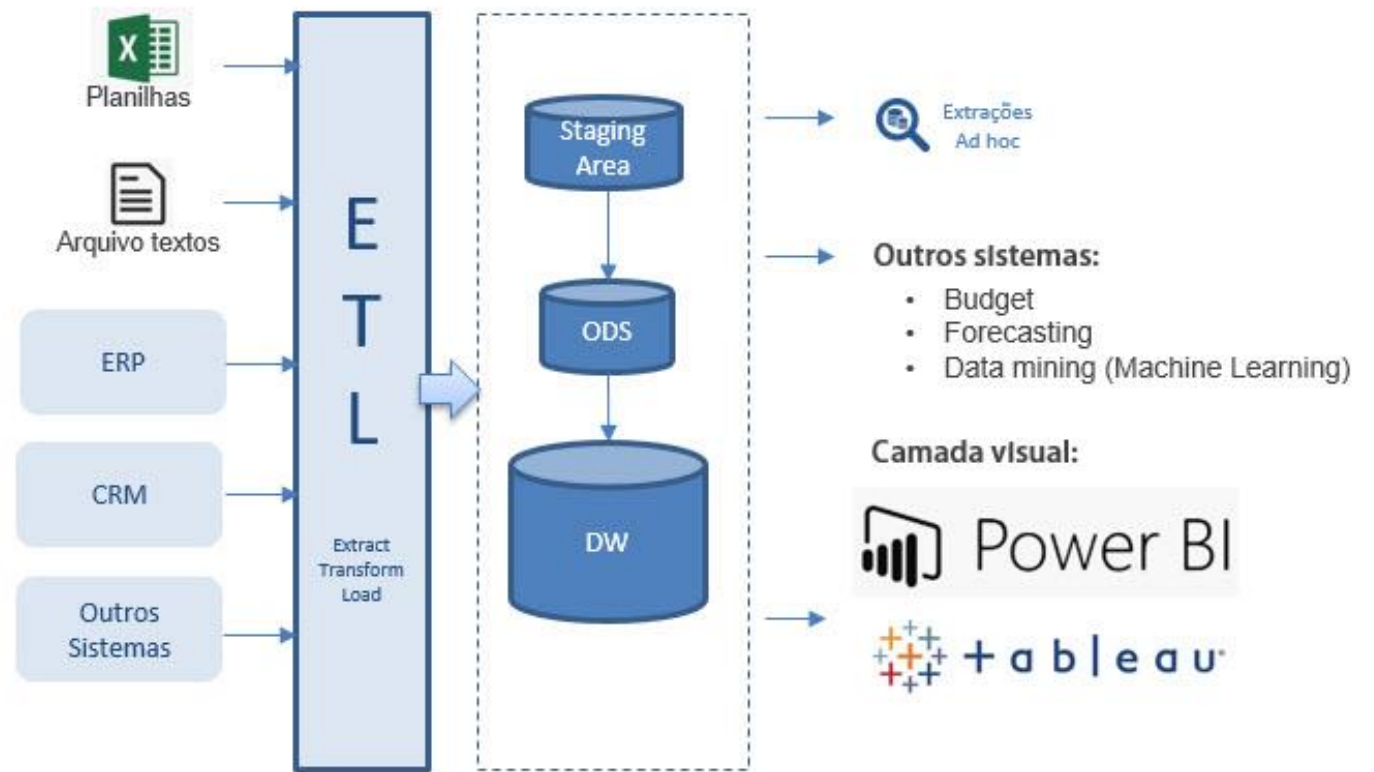
Enfoque da análise de dados

- A análise de dados consolidou no mercado conceitos e tecnologias específicas, como BI (Business Intelligence) e DW (Data Warehouse).
- Como padrão, estas duas utilizam dados transacionais para construir suas bases de dados onde usuários irão realizar suas consultas.

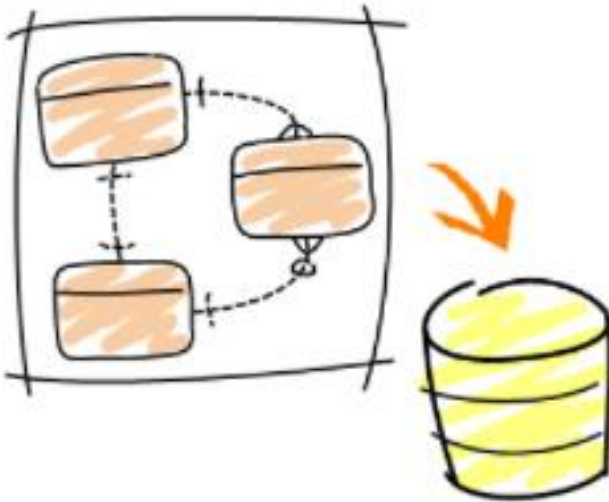


Enfoque da análise de dados

- Processos de ETL são utilizados para transformar bases transacionais em bases analíticas



Big Data Analytics



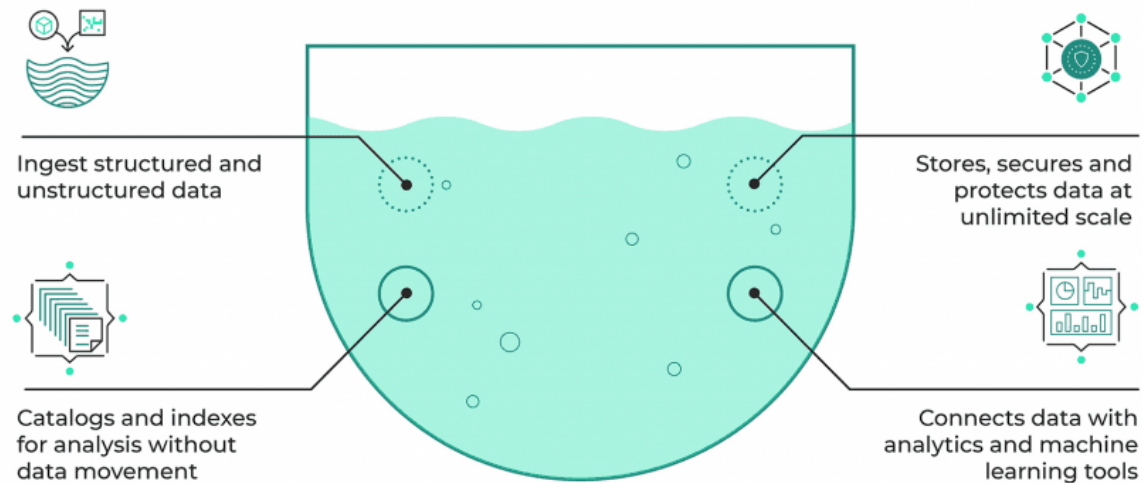
- Nos últimos anos, a área de Bancos de Dados descreveu a Internet como um grande "banco de dados" onde se tem inúmeros dados, porém, são dados semi-estruturados ou não estruturados.
- Mais recentemente, surgiu o conceito de Big Data, o qual consiste basicamente na coleta e no armazenamento de imensos volumes de dados existentes na Internet relacionados a um domínio de problema qualquer.
- Sendo assim, Big Data pode ser visto como uma solução importante para auxiliar na tomada de decisão, complementando os sistemas de Data Warehouse que muitas empresas já possuem.
- Um Data Warehouse é construído e utilizado para armazenar informações relativas às atividades de uma organização em bancos de dados, de forma consolidada



Big Data Analytics

- Por serem soluções com conceitos um pouco diferentes, a integração das informações neles contidas, torna-se complexas.
- Por exemplo, do lado do Data Warehouse, tem-se a lista de todos os clientes, as cidades de uma empresa, os produtos adquiridos nas últimas compras, entre outros dados transacionais.
- Pelo lado do Big Data, se tem um conjunto de dados de inúmeras pessoas, de inúmeras cidades e de inúmeros produtos, e para cada tipo de dado, mais dados adjacentes e não estruturados.
- Nova terminologia: DATA LAKE

Data Lake Features



Data Lakes

- O data lake é um repositório que centraliza e armazena todos os tipos de dados gerados pela e para a empresa.
- Eles são depositados ali ainda em estado bruto, sem o processamento e análise e até mesmo sem uma governança.
- A ideia é manter na organização dados que podem ser estrategicamente úteis, mesmo que eles, na realidade, não sejam requeridos em nenhum momento posterior.
- O data lake seria — em alguns casos já é — o local de armazenamento dessas informações.

DATA LAKE

vs

DATA WAREHOUSE

Data



unstructured

Users



Data Scientists,
Data Analysts

Use cases



Stream Processing,
Machine Learning,
Real time analysis

Raw

Data Lakes contain unstructured, semi structured and structured data with minimal processing. It can be used to contain unconventional data such as log and sensor data

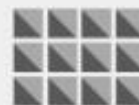
Large

Data Lakes contain vast amounts of data in the order of petabytes. Since the data can be in any form or size, large amounts of unstructured data can be stored indefinitely and can be transformed when in use only

Undefined

Data in data lakes can be used for a wide variety of applications, such as Machine Learning, Streaming analytics, and AI

Data



Structured

Users



Business Analysts

Use cases



Batch Processing,
BI. Reporting

Refined

Data Warehouses contain highly structured data that is cleaned, pre-processed and refined. This data is stored for very specific use cases such as BI.

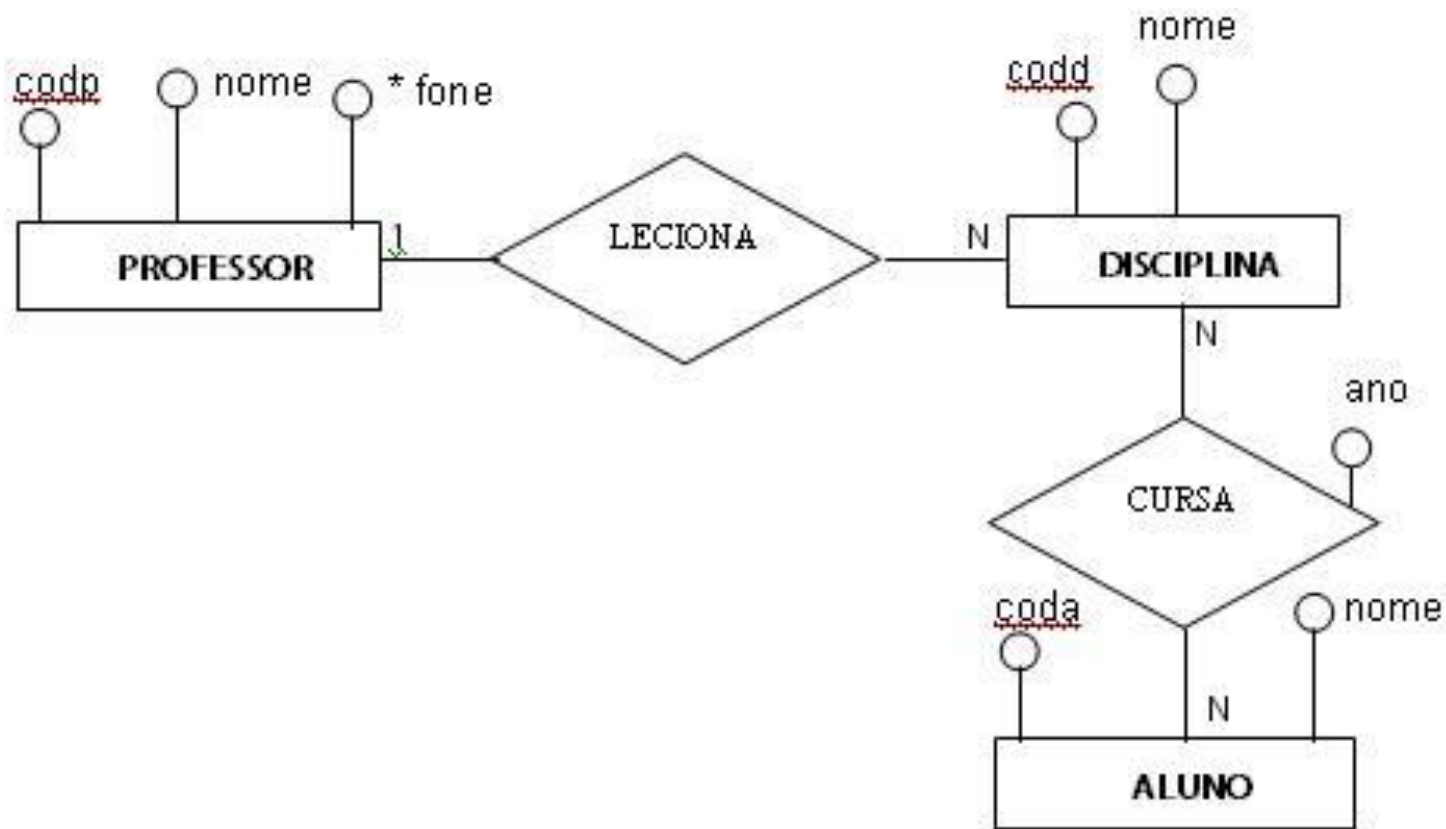
Smaller

Data Warehouses contain less data in the order of terabytes. In order to maintain data cleanliness and health of the warehouse, Data must be processed before ingestion and periodic purging of data is necessary

Relational

Data Warehouses contain historic and relational data, such as transaction systems, operations etc

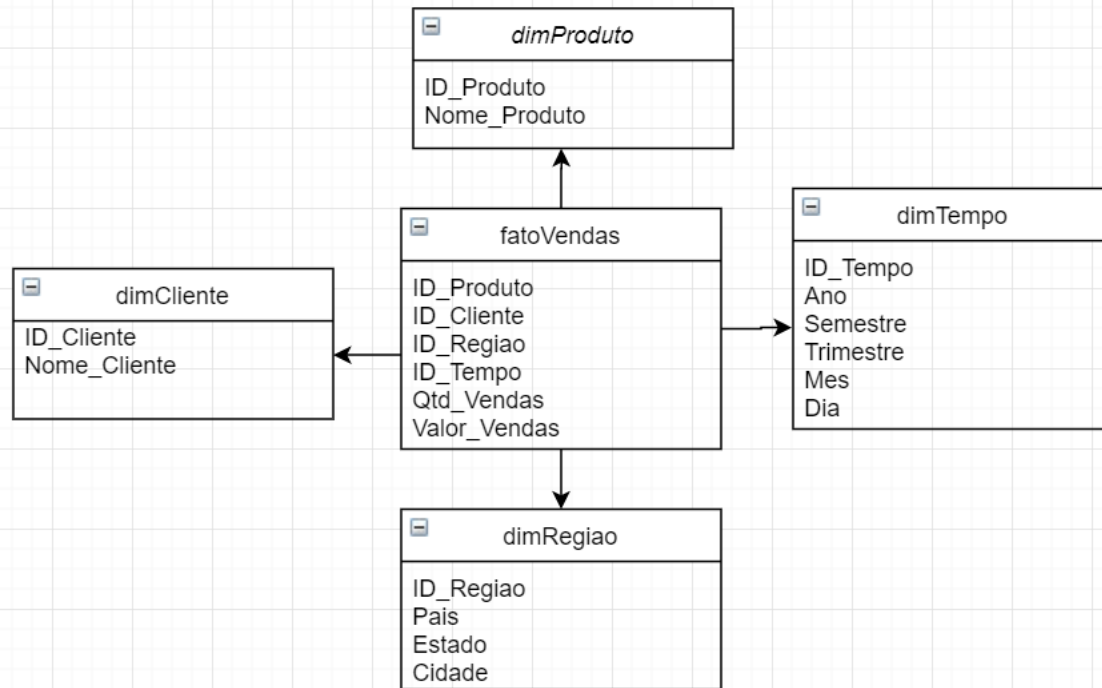




Resumindo: de novo! (transacional)

- O modelo de dados transacional é o projetado para atender a sistemas transacionais, ou seja, os que focam nos dados provenientes de interações de um sistema informatizado com suas fontes de dados.
- Por exemplo, quando se finalizam as compras no supermercado.
- Ao receber a fatura, recebe-se a lista como nomes dos produtos e seus preços, o total, etc.
- Esta fatura diz respeito somente àquela compra. Não há histórico de suas outras compras.
- A modelagem transacional prima por armazenar e recuperar os dados, normalmente poucos, referentes a numerosas transações.

Resumindo: de novo! (informacional)



- Já a modelagem informacional, também chamada de dimensional, visa a um objetivo oposto ao da transacional, ou seja, atende à necessidade de tomada de decisão baseada em grandes volumes de dados coletados das bases de dados transacionais.
- O histórico é essencial.
- Com isso, o banco de dados informacional, ou data warehouse, armazena uma imensa quantidade de dados.

Na próxima aula...

Vamos falar mais um pouco sobre o valor do BI para as organizações e os riscos que eventualmente cercam esses tipos de processos!

