



Fundamentos de Business Intelligence e Análise de Dados

BLOCO: B.I. E ANÁLISE DE DADOS

PROF. RODRIGO EIRAS, M.SC.

[ETAPA 5] AULAS 1 E 2 – BIG DATA ANALYTICS



Na aula anterior...

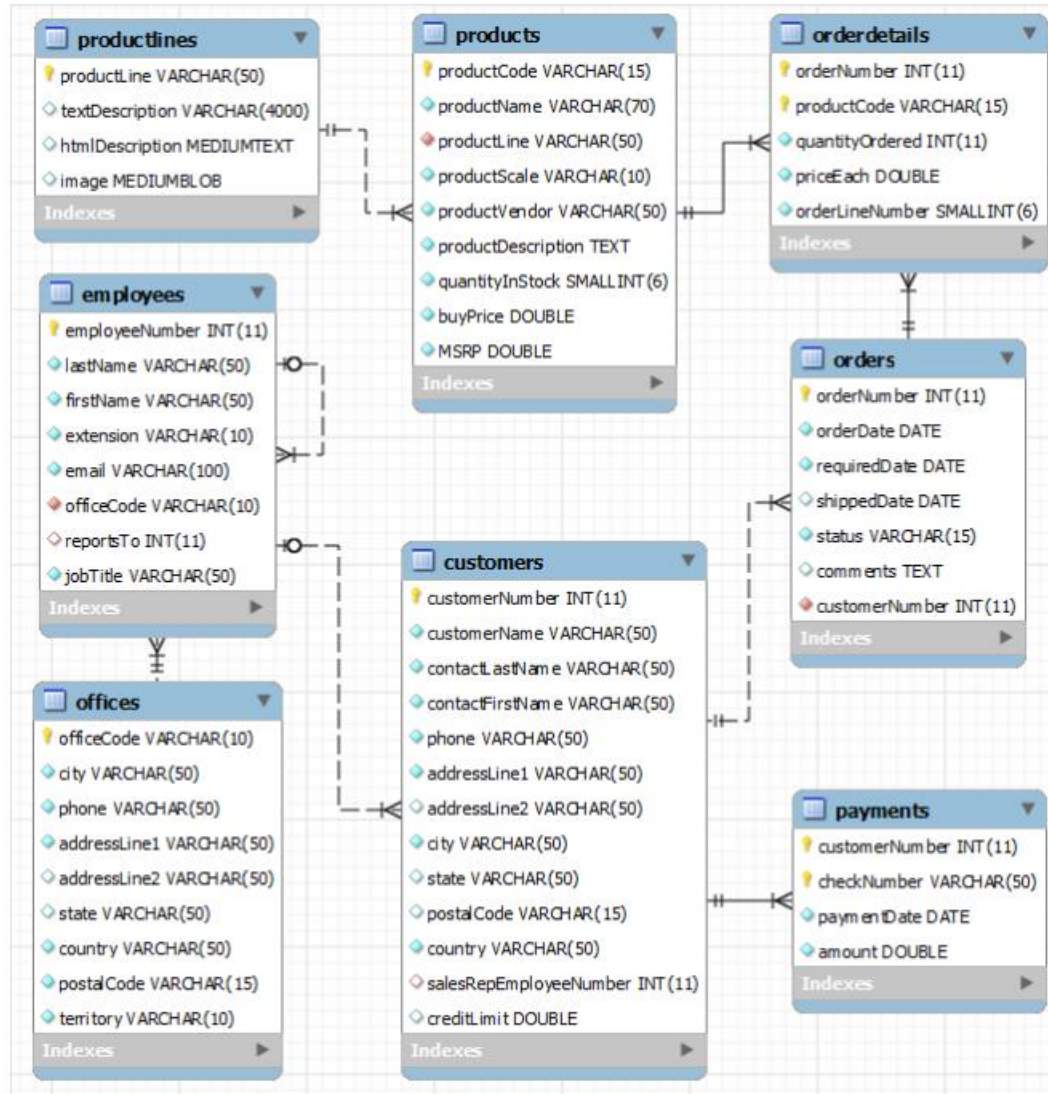
- Fizemos exercícios práticos iniciais com a linguagem SQL



Agenda

- Correção exercícios da Etapa 4
- Compreender os paradigmas de análise de dados existentes
- Compreender o panorama do mercado
- Mais exercícios (intermediários) usando SQL (Base IMDb)

Esquema para consulta



- Key: (Part of) Primary Key
- Filled Diamond: NOT NULL
- Not filled Diamond: NULL
- Red colored: (Part of) Foreign key
- Blue lined Diamond: Simple attribute (no key)

Can be combined for example:

- is a Red colored Key so it's a Primary Key which is also a Foreign Key
- is a Yellow (non Red) Key so it's only a Primary Key
- is a blue lined filled diamond so it's a NOT NULL simple attribute
- is a red colored filled diamond so it's a NOT NULL Foreign Key
- is a blue lined not filled diamond so it's a simple attribute which can be NULL
- is a red colored not filled diamond so it's a Foreign Key which can be NULL

Exercícios

Acesso o phpMyAdmin disponível em: <https://vmi578219.contaboserver.net/phpMyAdmin>

- Usuário e senha fornecido pelo professor via chat do Zoom.
- A base de dados a realizar consultas chama-se: classicmodels

Executar as consultas abaixo:

- Listar o endereço, cidade, estado e país de todos os “offices” (escritório)
- Listar o primeiro nome, último nome e também o cargo dos empregados
- Listar o primeiro nome, ultimo nome e cargo de empregados mas somente que trabalhem em NYC

Exercícios

Executar as consultas abaixo:

- Listar todos os produtos cujos preços estão entre 90 e 100
 - DICA: Pode usar função BETWEEN ou operadores lógicos maior que e menor que
- Listar todos os produtos cujos os preços NÃO estão entre 20 e 100
- Listar o nome e país de todos os clientes que possuem um consultor de venda (coluna: salesrepemployeenumber)
 - DICA: Precisa usar a função NULL
 - Consegue mostrar o número (nome) do consultor de venda em uma coluna?

Exercícios

Executar as consultas abaixo:

- Listar o primeiro nome e sobrenome de empregados e clientes na mesma tabela.
 - DICA: Precisa usar a função UNION

DESAFIOS:

- Listar o nome completo de clientes e empregados em uma única coluna.
 - DICA: Precisa usar as funções UNION e CONCAT

Alexander Semenov
Allen Nelson

- Como podemos repetir a consulta acima porém de forma ordenada?
 - DICA: Precisa usar a função ORDER BY

- Como separar os clientes e empregados com uma coluna de identificação?

- EXEMPLO:

Adrian Huxley	Customer
Akiko Shimamura	Customer
Alejandra Camino	Customer

Compreender os paradigmas de análise de dados existentes

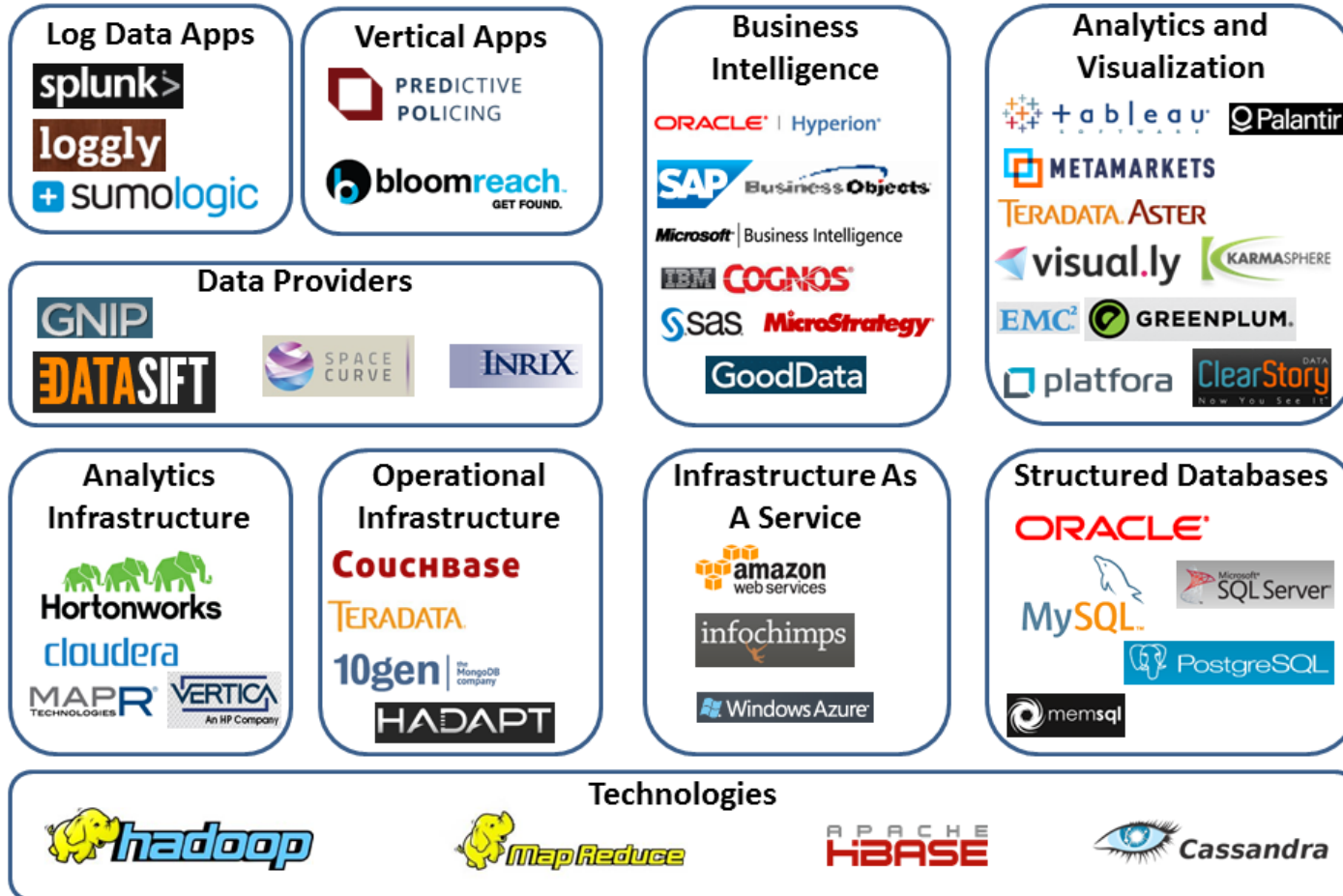
- Um dos temas mais relevantes para os profissionais de TI na atualidade é o BigData Analytics.
- BigData é muito mais que um grande volume de dados e, Analytics, é muito mais que estatística aplicada.
- Por isso, tais conceitos, tanto BigData, quanto Analytics, nem sempre são bem definidos, motivo pelo qual muitos de nós, erroneamente, nos apegamos a paradigmas tradicionais existentes e conceituamos as novas tecnologias com o “olhar” do passado.

O valor de Big Data vem da sua dinâmica, provocada pela análise dos dados ou seja, dos dados falarem por si mesmo.

Cezar Taurion



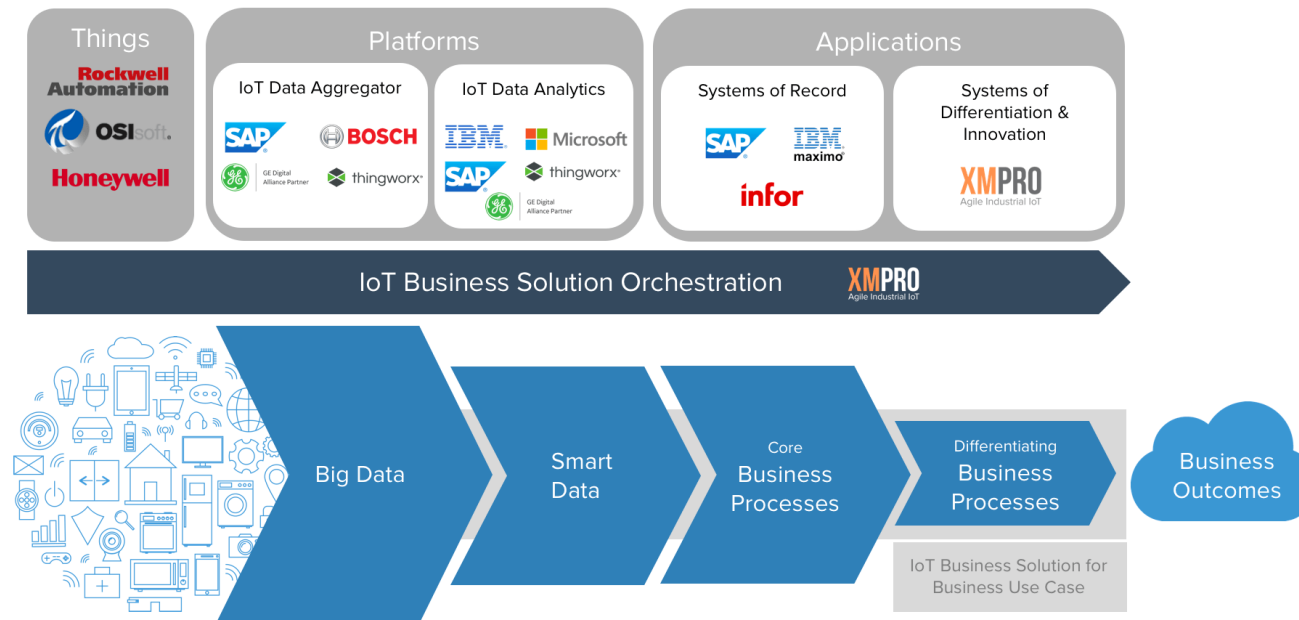
Big Data Landscape



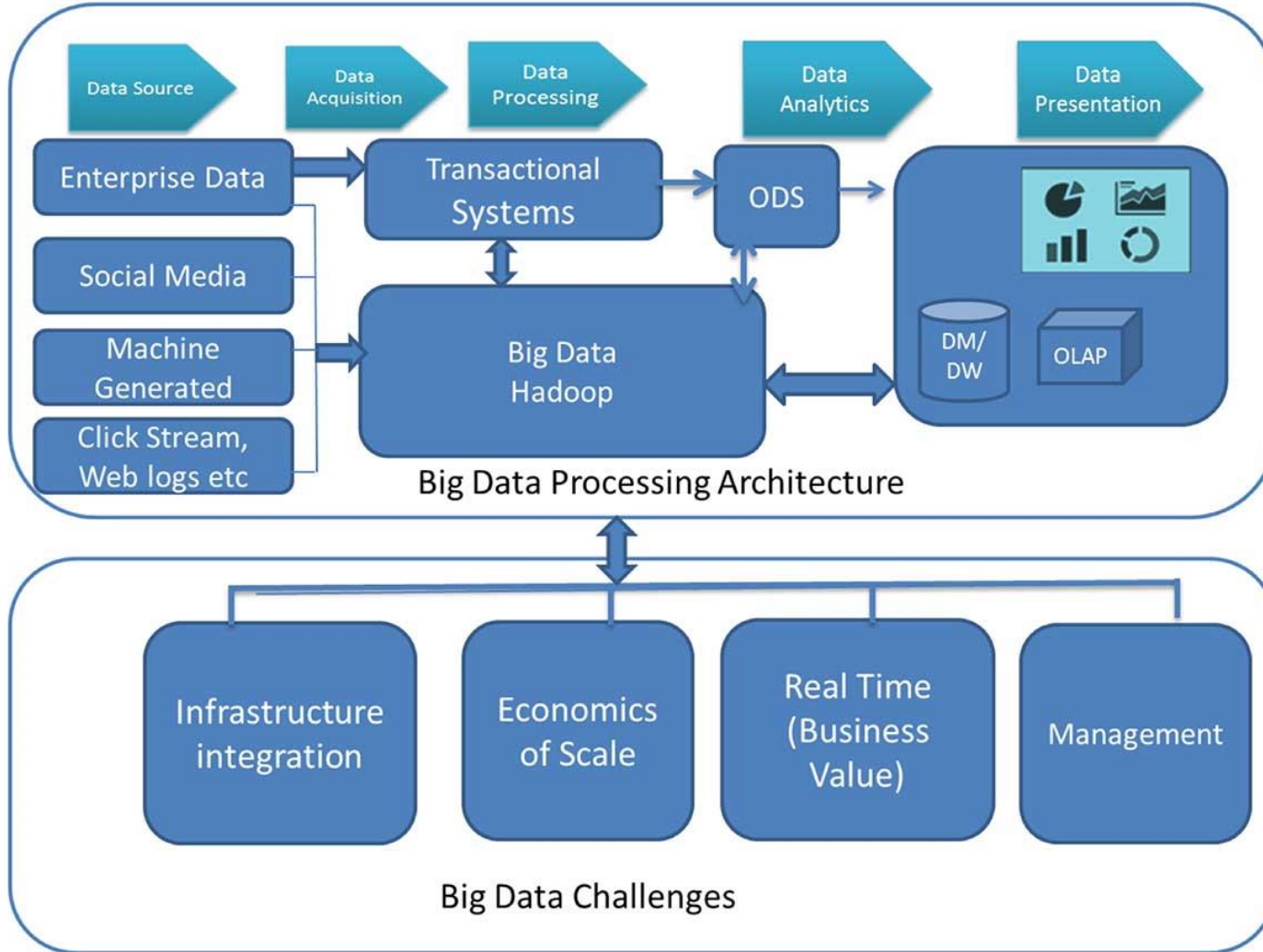
Compreender os paradigmas de análise de dados existentes

- BigData não é apenas uma nova forma de armazenar dados.
- Grandes frustrações são geradas nas empresas quando tentam simplificar uma ideia de BigData por simplesmente comprar tecnologias do fornecedor X ou Y.

Compreender os paradigmas de análise de dados existentes



- Big Data Analytics traz consigo mudanças na maneira de pensar os dados.
- Por exemplo, a relação deste conceito com VOLUME DE DADOS.
- Ao sair do pensamento baseado na escassez para abundância de dados você deve pensar diferente.



Compreender os paradigmas de análise de dados existentes

- Por limitação e/ou dificuldade tecnológica você pode ser induzido a construir um modelo mental de escassez de dados.
- Com isso, você pode simplificar uma série de práticas, como análises estatísticas por amostragem.
- A partir de uma pequena amostra de dados, você poderá extrapolar para um cenário mais amplo.

I. A acurácia dos dados e, conseqüentemente, dos resultados está intimamente relacionada a amostragem. Pense em uma pesquisa de opinião baseada em uma amostragem aleatória de usuários de telefones fixos: *se a coleta for feita no horário de trabalho, as respostas não representarão a opinião das pessoas que trabalham fora do horário de trabalho – podem ter respostas diferentes das que atenderão o telefone no horário de trabalho (as pessoas que estarão em casa).*

II. Um outro ponto é termos uma amostra de dados pequena, como realizado hoje em dia, não terá representatividade estatística. Um exemplo? As pesquisas de intenção de voto. Geralmente pesquisam-se em torno de 2.000 pessoas e uma visão geral. Mas se quisermos mais detalhes sobre uma faixa etária em uma determinada região, a amostragem será insuficiente.

Compreender os paradigmas de análise de dados existentes

- Com o tempo os modelos foram refinados, e hoje eles são bastante confiáveis.
- Entretanto, há algumas lacunas:
 - Acurácia
 - Amostragem

Compreender os paradigmas de análise de dados existentes

Tratando de BigData, a ideia é outra.

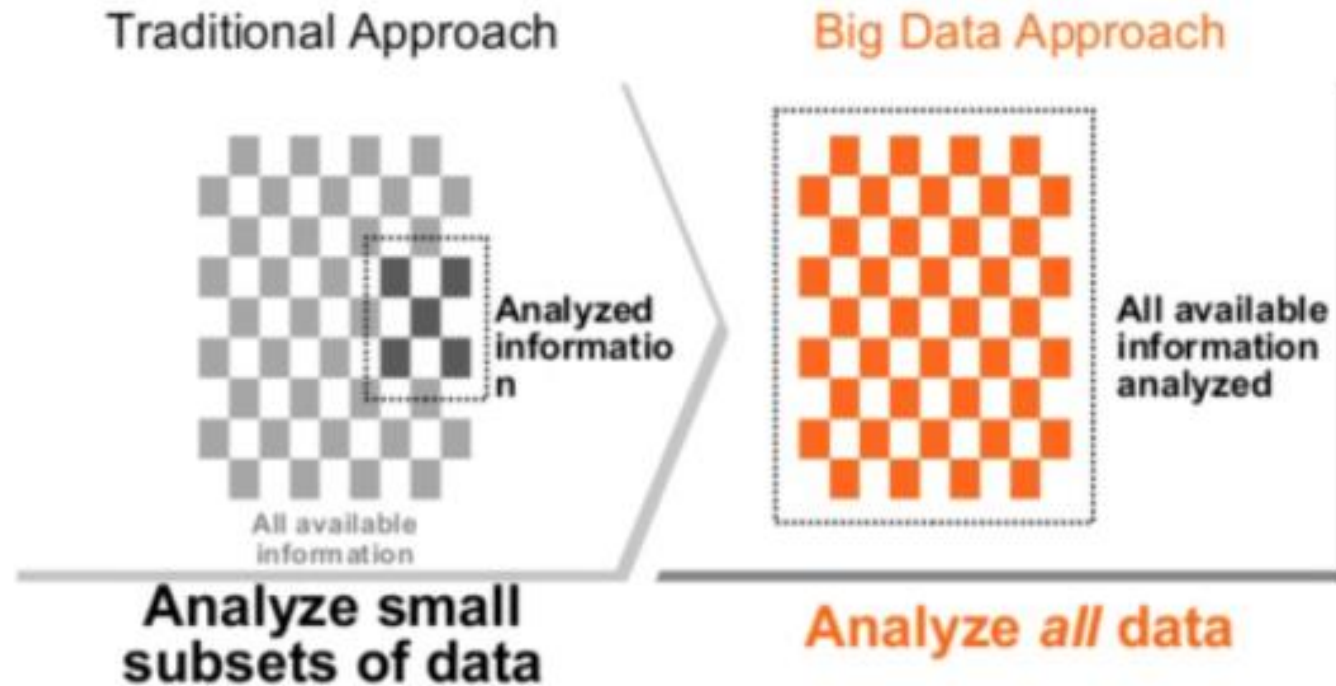
Quando todos os registros estão acessíveis, podemos ter amostras das mais diversas.

Podemos identificar tendências e descobrir correlações não pensadas antes.

Podemos fazer novas perguntas e descer a novos níveis de segmentação.

Assim, quebramos paradigmas, ou seja, aproveitamos oportunidades de fazer perguntas não pensadas antes de analisar os dados.

Compreender os paradigmas de análise de dados existentes



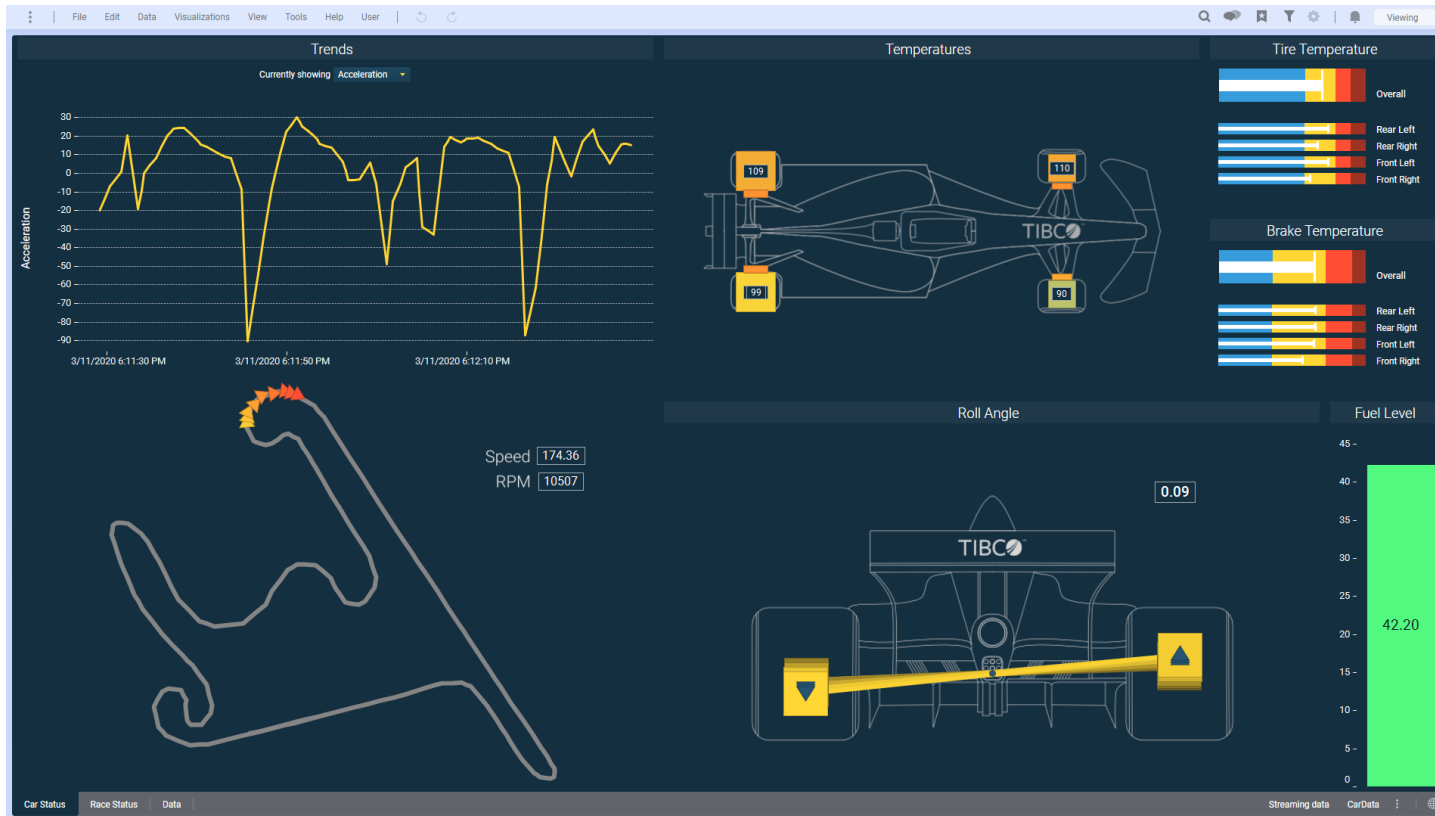
- A relação volume de dados e acurácia está totalmente relacionada, dada a flexibilidade de acesso aos dados e a certeza do melhor método estatístico a ser aplicado.
- A grande mudança de paradigma está em responder as perguntas específicas (que já foram preparadas anteriormente a análise) para entender os dados da maneira como eles foram correlacionados.
- Uma outra interpretação a esta nova abordagem é o desenvolvimento de algoritmos preditivos, que buscam identificar eventos antes que eles aconteçam.



Compreender os paradigmas de análise de dados existentes

- Um exemplo rotineiro na indústria de motores, que coletam dados de sensores, é o volume de dados gerados e o que se pode analisar.
- A ideia básica é explicar a ocorrência de uma determinada falha/quebra de equipamento quando ela ocorre.
- Dada esta nova abordagem (quebra de paradigma) pode-se fazer análises correlacionais para identificar determinados padrões que sinalizam futuros problemas.
- Quanto mais cedo uma provável anormalidade é detectada, mais eficiente é o processo de manutenção.

Compreender os paradigmas de análise de dados existentes



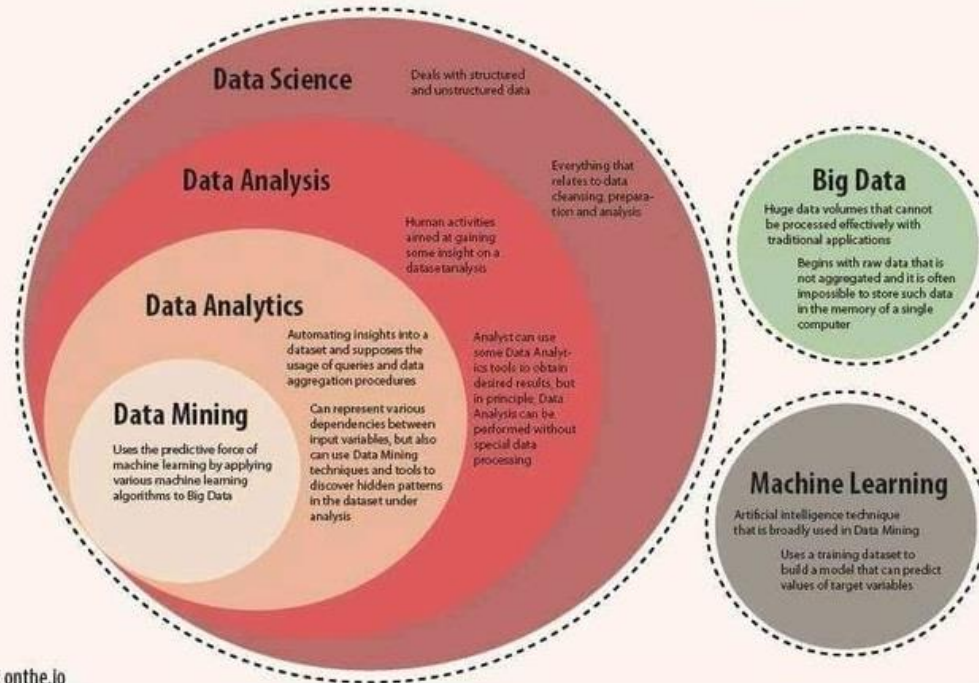
- Na abordagem tradicional podemos reconhecer, por diversos momentos, que um equipamento troca de óleo a cada N meses.
- Assim, saberemos, em média, quando será a próxima troca de óleo.
- Diferente, seria identificar uma próxima troca de óleo correlacionando dados como níveis de temperatura do motor, temperatura externa, corrente do motor, dentre outras.
- O que poderia ocasionar uma troca de óleo a alguns dias antes ou depois dos N meses (abordagem tradicional)



Compreender os paradigmas de análise de dados existentes

- Big Data Analytics é uma mudança de paradigma – a substituição do modelo baseado em intuição (“eu acredito em uma coisa e, assim, explico/provo”) por direcionamento de descobertas, a cada análise dos dados (“eu entendo o comportamento dos dados e suas correlações, assim, explico/provo”).
- Entender a aplicabilidade de um Big Data Analytics vai muito além da escolha por um tecnologia.

What is the difference between Data Science, Data Analysis, Big Data, Data Analytics, Data Mining and Machine Learning?



Source: onthe.io

Compreender os paradigmas de análise de dados existentes

- A primeira abordagem é pensar e identificar as reais oportunidades que esta abordagem pode trazer, ou seja, seu valor.
- Lembre-se sempre que este é um processo evolutivo, de contínuo aprendizado.
- Novas descobertas surgirão a cada iteração, a cada exploração!



Compreender o panorama de mercado

- Muito bem, compreendemos os paradigmas existentes no contexto sobre análise de dados, mas como isso está afetando as empresas?
- Bom, empresas sempre usaram dados para fazer análises internas; o que mudou é que no contexto de Big Data Analytics, os gestores estão munidos com mais dados.
- Além disso, novas tecnologias permitem extrair valor mais facilmente a partir dessas análises.
- Para muitas pessoas, Big Data é um termo que se refere apenas ao acúmulo de um grande volume de dados.



Compreender o panorama de mercado

- Mas, apenas isso não é suficiente para criar valor para seu cliente e muito menos para sustentar alguma vantagem competitiva para sua empresa.
- É preciso saber o que se busca responder (boas perguntas são o primeiro passo para uma estratégia bem-sucedida) e entender como interpretar e usar esses dados.
- Além de compreender os tipos de análises que podem ser usadas, é preciso entender em que investir.



Oportunidades em Big Data Analytics

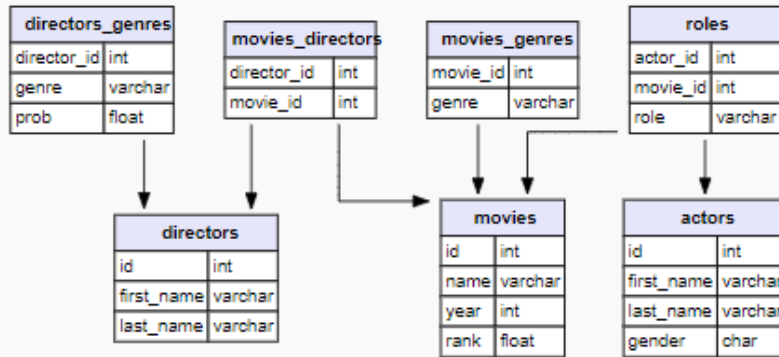
- Identificar relações jamais pensadas:
 - Tendências/padrões por meio da análise de dados em relação a comportamento do usuário ou consumidor.
 - Isso permite identificar tendências de consumo, detectar fraudes (ações que fujam muito do padrão de comportamento), dentre outros;
- Identificar perfis de clientes;
- Desenvolver/otimizar produtos/serviços ou processos:
 - A partir da identificação de tendências e peculiaridades do mercado da organização;
- Otimizar estoques e promoções com análises preditivas mais precisas;
- Construir cenários e simulações mais próximos da realidade antes de agir.

“...A procura por gerentes de projeto com experiência em big data mais do que dobrou (123%) em 2014, segundo a Wanted Analytics, empresa que analisa sites de emprego no mundo todo. Um impulsionador desse mercado é o presidente americano, Barack Obama, que criou uma secretaria de serviços digitais com status de ministério para trabalhar com a quantidade massiva de dados que o governo produz.

Em um evento de computação, Obama fez uma reverência pessoal ao profissional de big data escolhido para chefiar a secretaria, DJ Patil, ex-LinkedIn. “Ajude-nos a construir serviços digitais melhores para o povo americano, ajude-nos a liberar inovações em áreas como saúde e mudança climática”, disse Obama em seu pedido a DJ.

O mais empolgante é que o big data tem efeito sobre profissionais de todos os departamentos. Analistas do banco de investimento UBS usaram vigilância por satélite de 100 estacionamentos do Walmart e recolheram dados sobre o número de carros estacionados em cada um, todos os meses...”

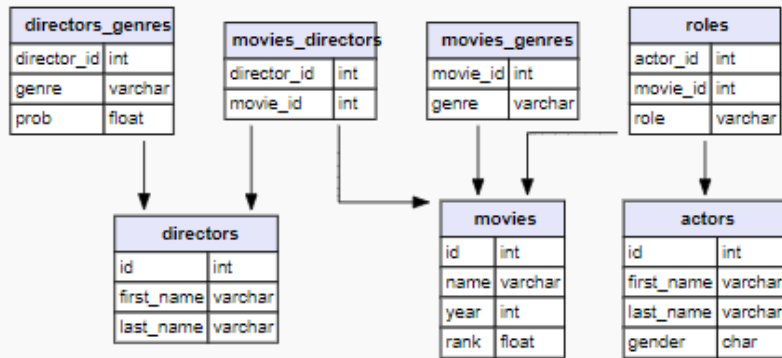
Revista Exame, em 15/05/2015, sobre Big Data



Exercícios de SQL

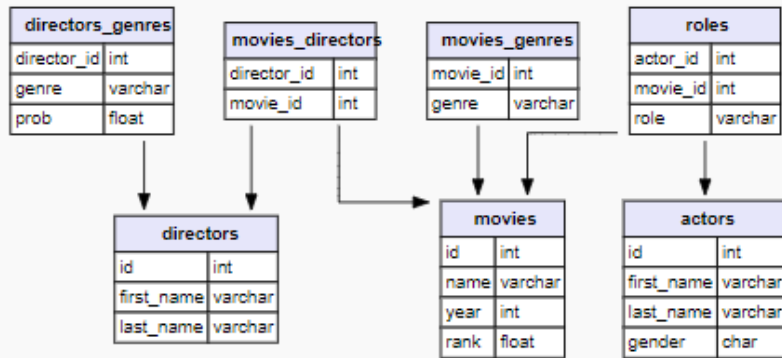
As consultas deverão ser feitas usando a mesma conexão criada no DBViz, com o mesmo usuário e senha da última aula. A base onde as consultas deverão ser feitas: IMDb





Exercícios de SQL

- 1) Quantos filmes existem cadastrados na base de dados?
 - DICA: Função COUNT
- 2) Você consegue exibir somente os filmes que tenham o nome “Central” no título?
 - DICA: Função LIKE
- 3) Liste os filmes cadastrados ordenados pelo rank, maiores “notas” vem primeiro.
 - DICA: Função ORDER BY
- 4) Qual o filme mais velho cadastrado na base que possua uma nota?
 - DICA: Funções ORDER BY e NULL



Exercícios de SQL

5) DESAFIO: Crie uma tabela contendo os nomes de filmes, o ano que foi lançado e os papéis interpretados nos filmes listados pela atriz Fernanda Montenegro

- DICA: INNER JOINS

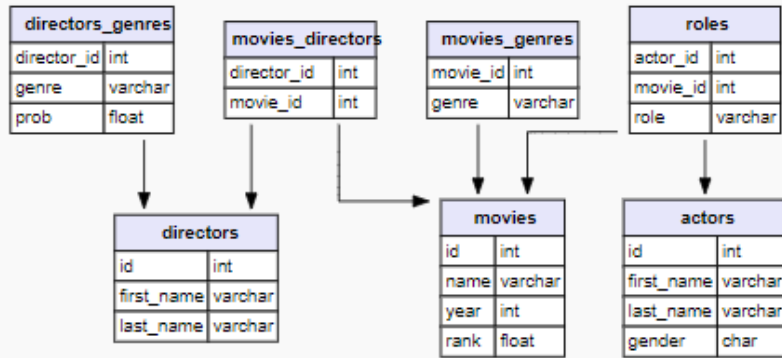
6) Consegue adicionar gênero do filme na tabela criada no exercício 5?

- DICA: INNER JOIN

7) Compare os resultados dos exercícios 5 e 6 e argumente porque existe uma diferença no resultado do número de linhas entre o exercício 5 e o exercício 6.

8) Adicione também o nome do diretor nos filmes em que Fernanda Montenegro atuou.

- DICA: INNER JOIN na tabela do exercício 5 ou 6



Exercícios de SQL

9) Qual a média de nota (rank) dos filmes de José Wilker e Fernanda Montenegro?

- DICA: Funções AVG, INNER JOINS, AND, OR, GROUP BY

10) Qual dos dois atores tem a maior média?

A close-up photograph of a man with dark, curly hair and a prominent mustache. He is smiling broadly, showing his teeth, and giving a thumbs-up gesture with both hands. He is wearing a light blue shirt and a grey jacket. The background is blurred with bokeh lights. The image has a dark overlay with white text.

Na próxima aula...

Começaremos a trabalhar com o Microsoft Power BI.