



DATA BASE

Paulo Ponciano Peral Filho

João Victor B.C Teixeira

Pedro Henrique Albino Leão

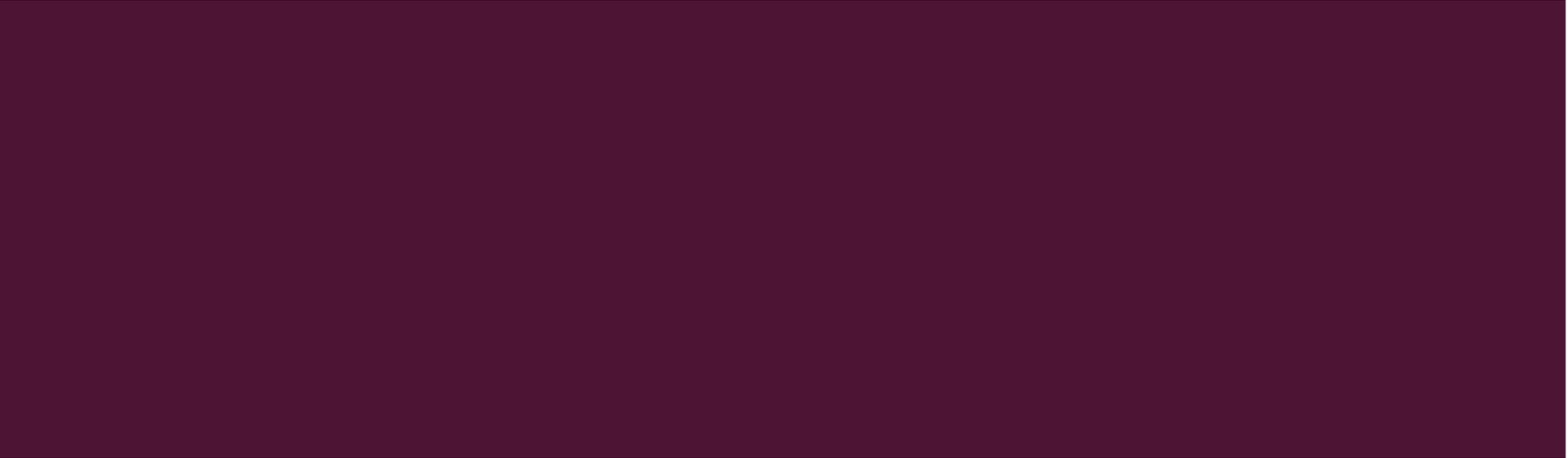
Guilherme Ferreira Santa Cecília

Pedro Henrique Silva Xavier

Gabriel Batista de Oliveira

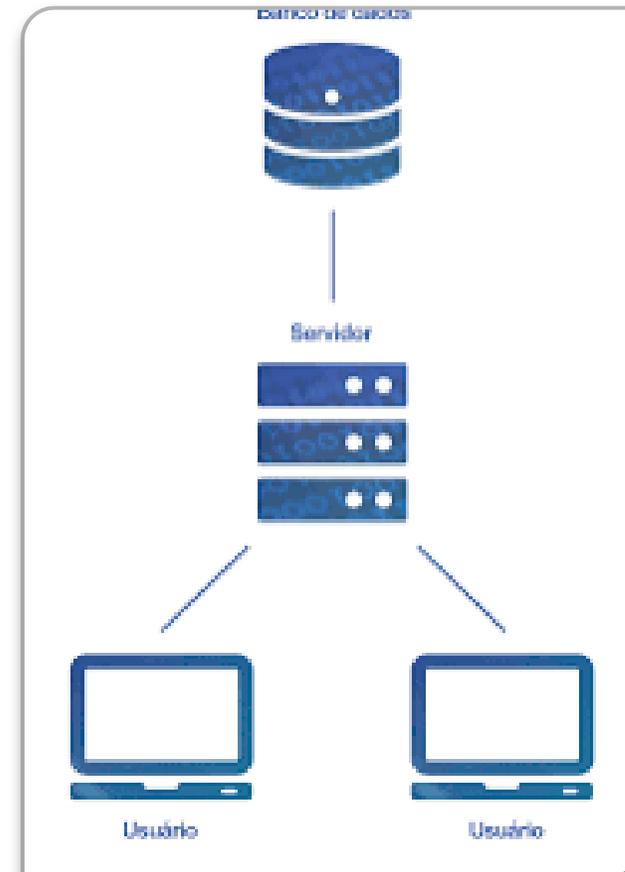


NORMALIZAÇÃO DE BANCO DE DADOS



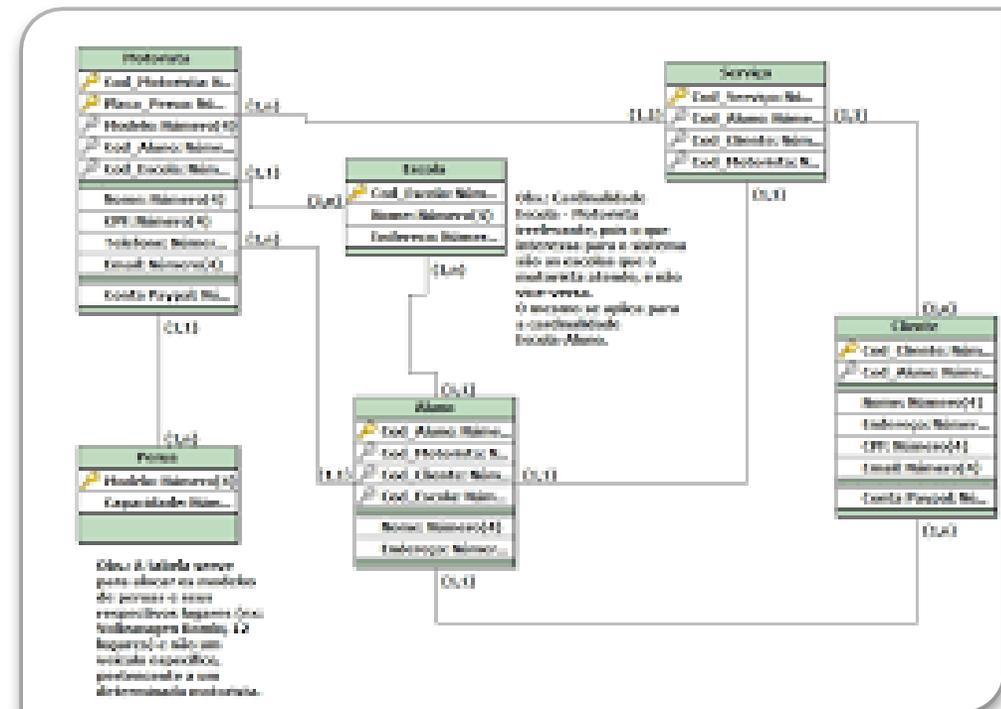
O QUE É NORMALIZAÇÃO?

- Normalização é o processo de modelar o banco de dados de forma eficiente a fim de eliminar redundâncias no banco.



NORMALIZAÇÃO E SEUS BENEFÍCIOS

- Tal procedimento é feito a partir da identificação de uma anomalia em uma relação, decompondo-as em relações melhor estruturadas.



FORMAS NORMAIS

Conjuntos de regras para determinar com qual forma normal o banco é compatível.

- Primeira forma normal
- Segunda forma normal
- Terceira forma normal
- Quarta forma normal

NORMALIZAÇÃO

- *O processo de normalização compreende o uso de um conjunto de regras, chamados de formas normais. Ao analisarmos o banco de dados e verificarmos que ele respeita as regras da primeira forma normal, então podemos dizer que o banco está na “primeira forma normal”. Caso o banco respeite as primeiras três regras, então ele está na “terceira forma normal”. Mesmo existindo mais conjuntos de regras para outros níveis de normalização, a terceira forma normal é considerada o nível mínimo necessário para grande parte das aplicações. [Microsoft 2007]*



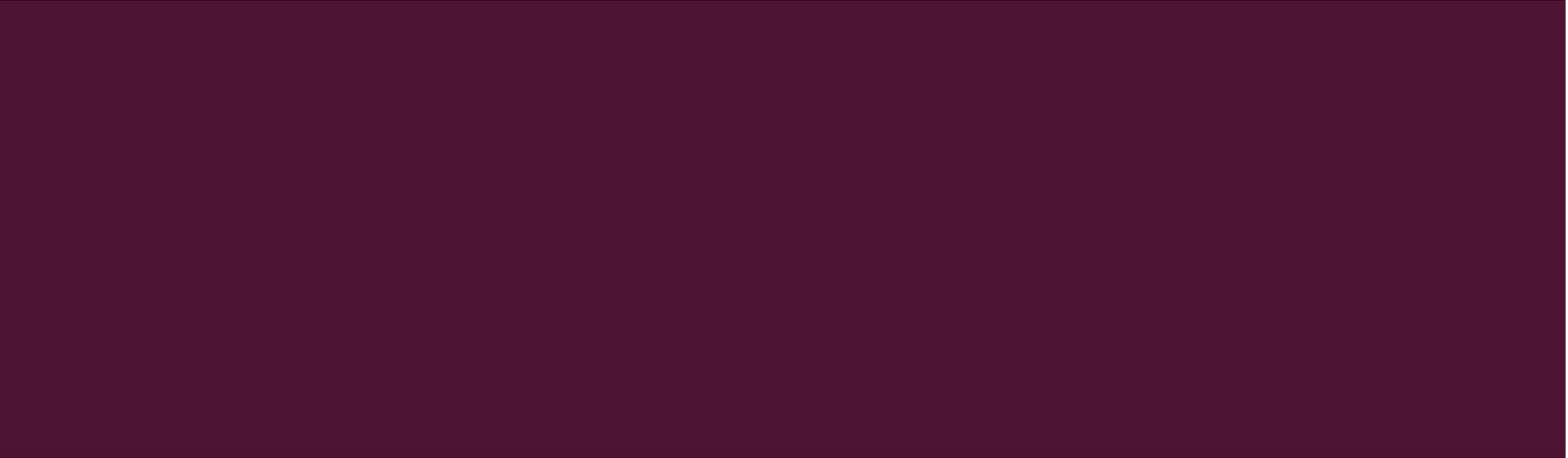
www.dba-pro.com

Apresentador: Josué





SQL - STRUCTURED QUERY LANGUAGE



O QUE É SLQ?

- Linguagem Estruturada de Consulta, ela foi a forma encontrada para que a comunicação com um banco de dados pudesse ser feita de uma maneira descomplicada

COMO FUNCIONA O SQL?

- A linguagem SQL é dividida em quatro agrupamentos de acordo com o tipo de operação a ser executada no banco de dados.

DML

- O DML é a Linguagem de Manipulação de Dados, é o subconjunto mais utilizado da linguagem **SQL**, pois é através da DML que operamos sobre os dados dos bancos de dados com instruções de inserção, atualização, exclusão e consulta de informações.

DCL

- O DCL é a Linguagem de Controle de Dados, é o subconjunto da **SQL** utilizado para controlar o acesso aos dados, basicamente com dois comandos que permite ou bloqueia o acesso de usuários a dados.

DTL

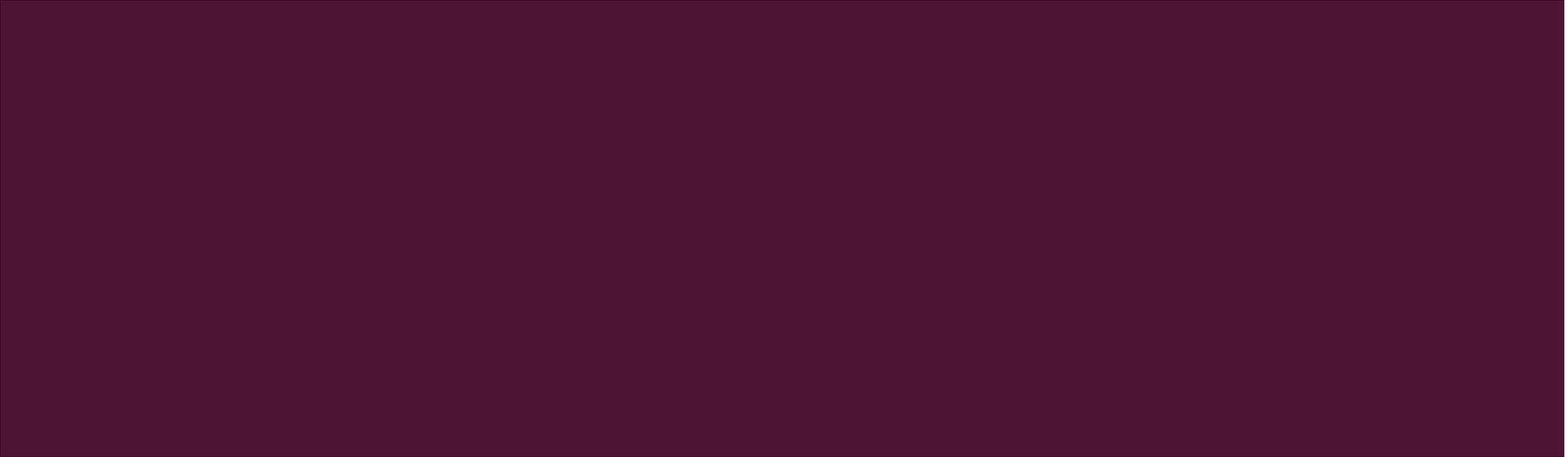
- DTL é a Linguagem de controle de transações sendo um subconjunto da SQL que fornece mecanismos para controlar transações no banco de dados. São três comandos: iniciar uma transação, efetivar as alterações no banco de dados e cancelar as alterações.

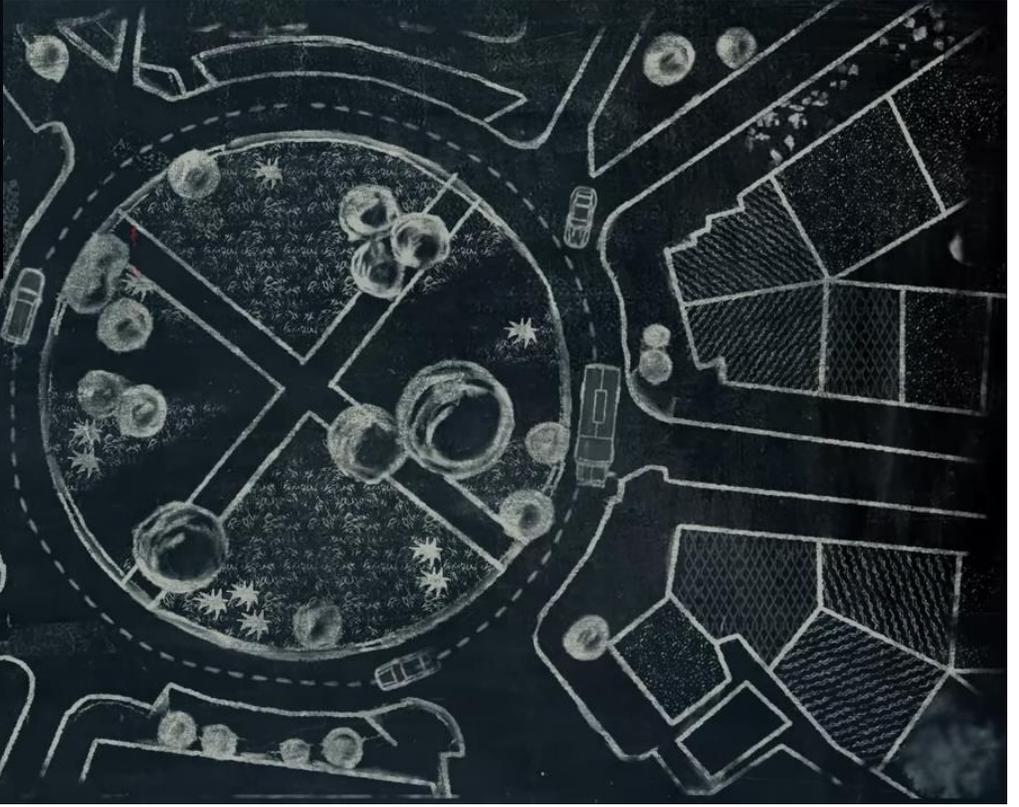
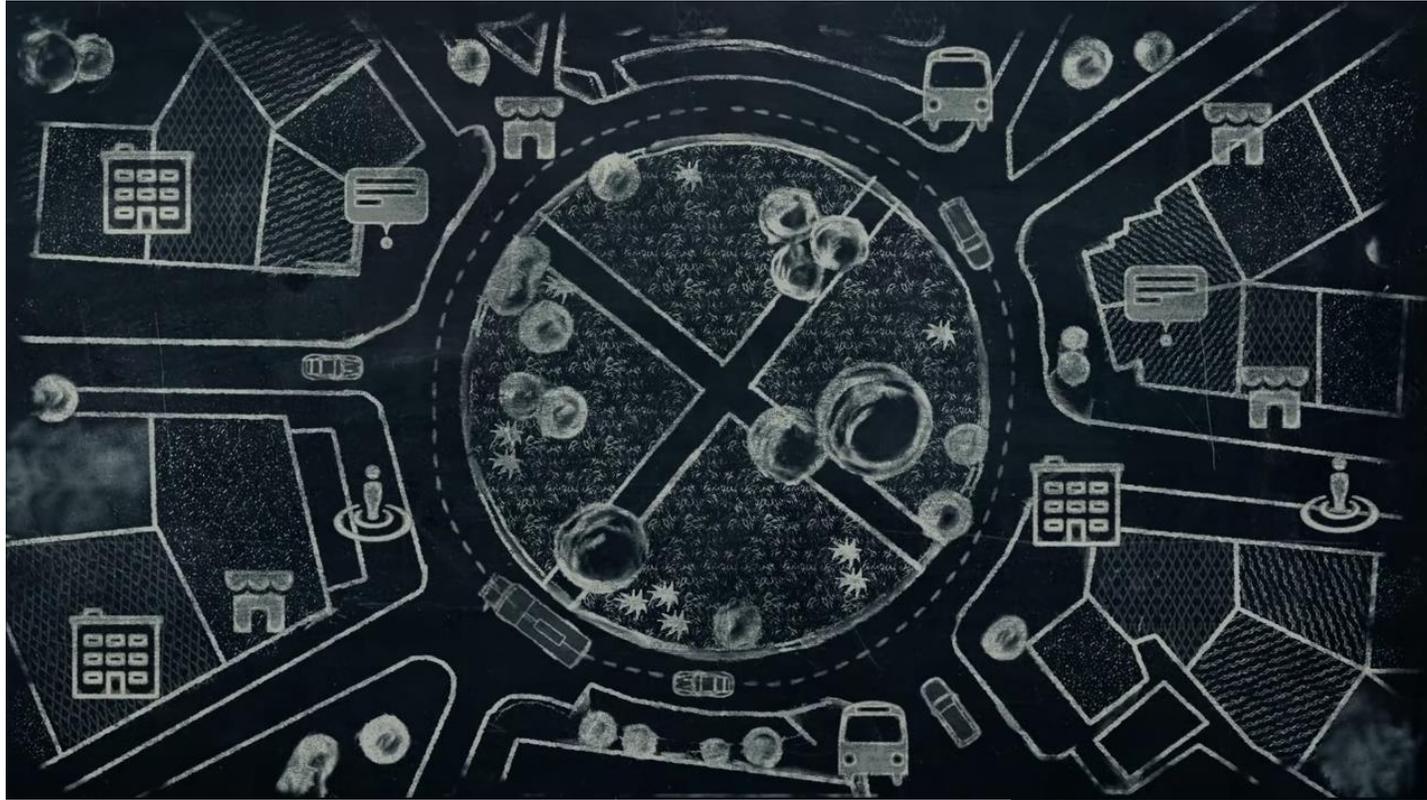
DDL

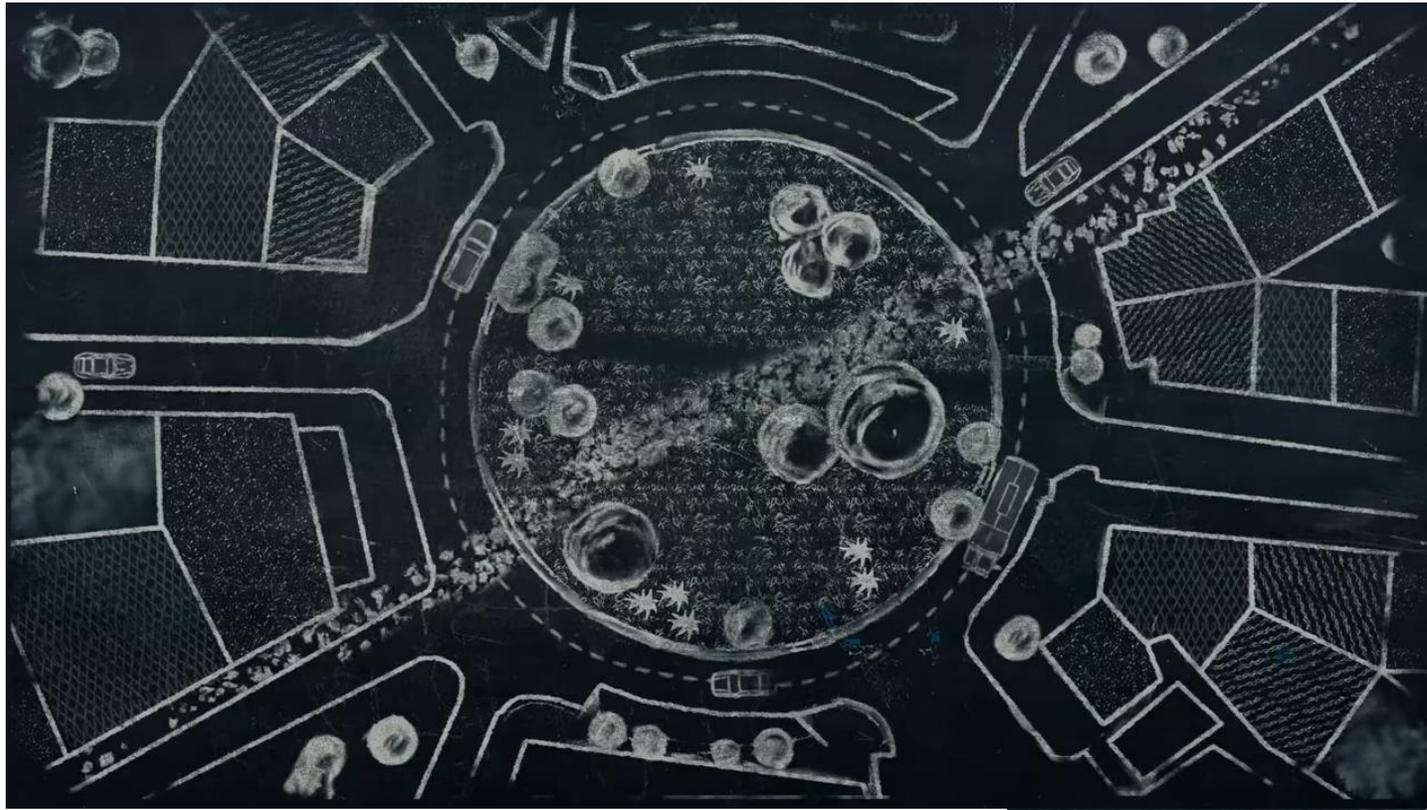
- DDL é a Linguagem de Definição de Dados, é o subconjunto da SQL utilizado para gerenciar a estrutura do banco de dados, com a DDL podemos criar, alterar e remover objetos (tabelas, visões, funções, etc.) no banco de dados



BIG DATA











por que

por que

por que **o céu é azul**

por que **separado**

por que **a propriedade da água de dissolver outras substâncias é importante para a vida**

por que **eu te amo**

por que **a camada de ozônio é importante para a vida na terra**

por que **o filo dos platelmintos recebeu esse nome**

por que **ou porque**

por que **os moluscos têm esse nome**

por que **a concentração de gás carbônico na atmosfera vem aumentando nos últimos tempos**

Pesquisa Google

Estou com sorte

[Denunciar previsões inadequadas](#)



Home

My Channel

Trending

Subscriptions

History

Watch Later

SUBSCRIPTIONS

Add channels

Popular on YouTube

Music

Sports

Gaming

Browse channels

Manage subscriptions

CrazyFrogVEVO Recommended channel for you



Crazy Frog - Axel F

CrazyFrogVEVO
113,914,131 views • 6 years ago



Crazy Frog - Last Christmas

CrazyFrogVEVO
35,174,544 views • 5 years ago

Papiaan Recommended channel for you



Construction Fail Compilation
2015 NEW!

Papiaani
2,524,529 views • 3 months ago



Truck Fail Compilation 2015

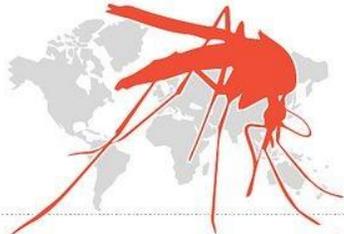
Papiaani
1,382,590 views • 1 month ago

Car Crash Compilation 7 Recommended channel for you



Helping to predict and stop the spread of infectious disease

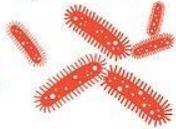
Scientists from IBM are collaborating with John Hopkins University and University of California, San Francisco to combat illness and infectious diseases in real-time with smarter data tools for public health by applying the latest analytic models, computing technology and mathematical skills on an open-source framework.



The rise of **global transportation, trade and climate change** allows insects to easily carry disease organisms like **malaria and dengue fever** across borders, infecting animals as well as humans.

Malaria is responsible for nearly

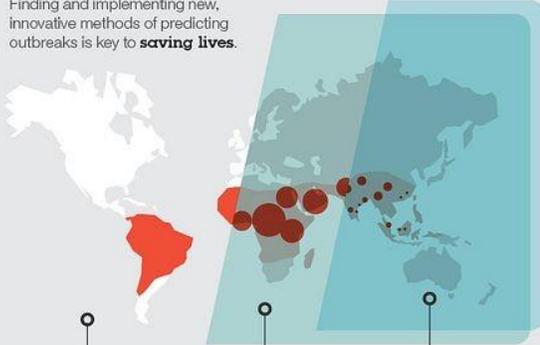
1 million
annual deaths*



Dengue fever has spread to over

100
countries, including the United States

Finding and implementing new, innovative methods of predicting outbreaks is key to **saving lives**.



Data from the World Health Organization showing general sensitivity to outbreaks



Addition of population models (including humans and mosquitos)



Integrating climate data for temperature and precipitation can more accurately pinpoint potential outbreaks



These model results make it possible to **predict where disease incidence** is most likely to increase or decrease.

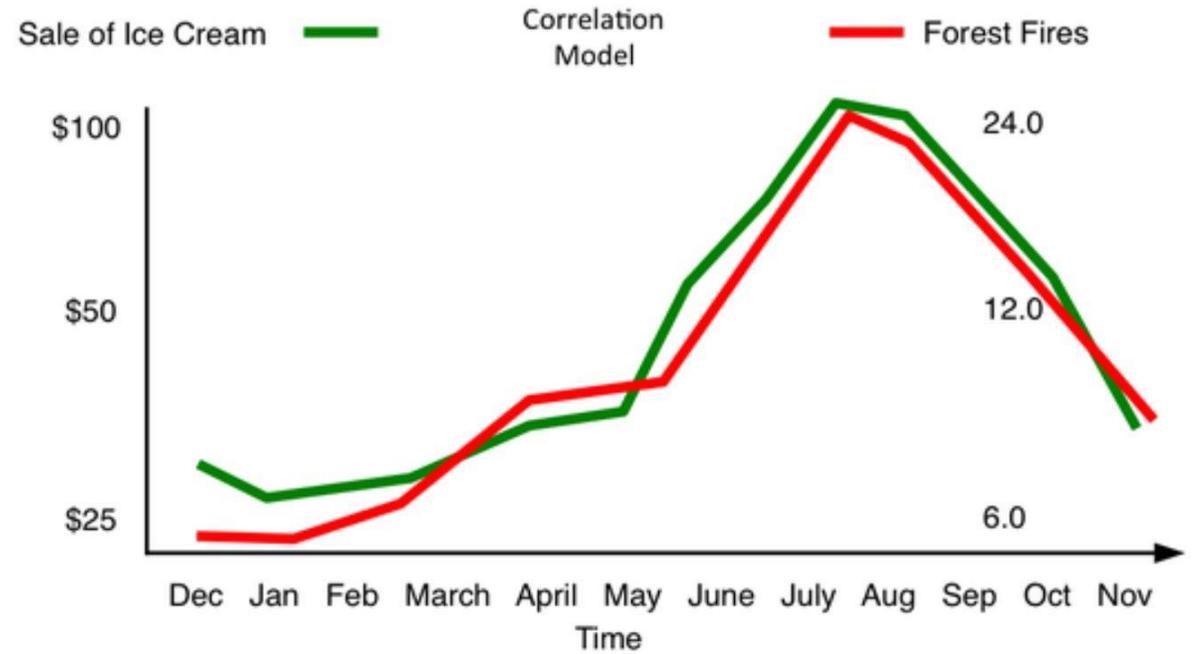
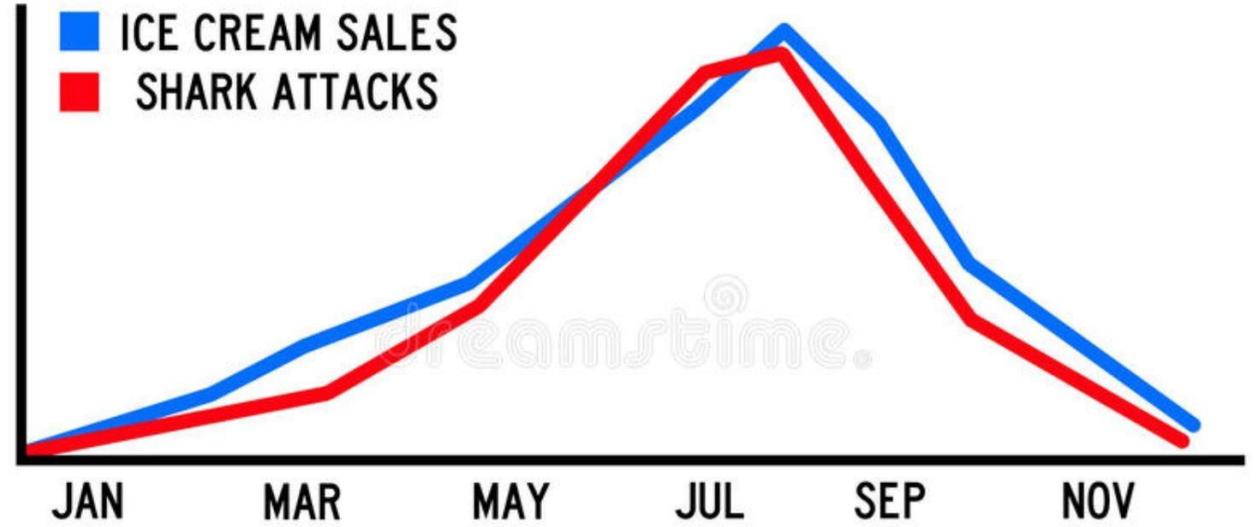
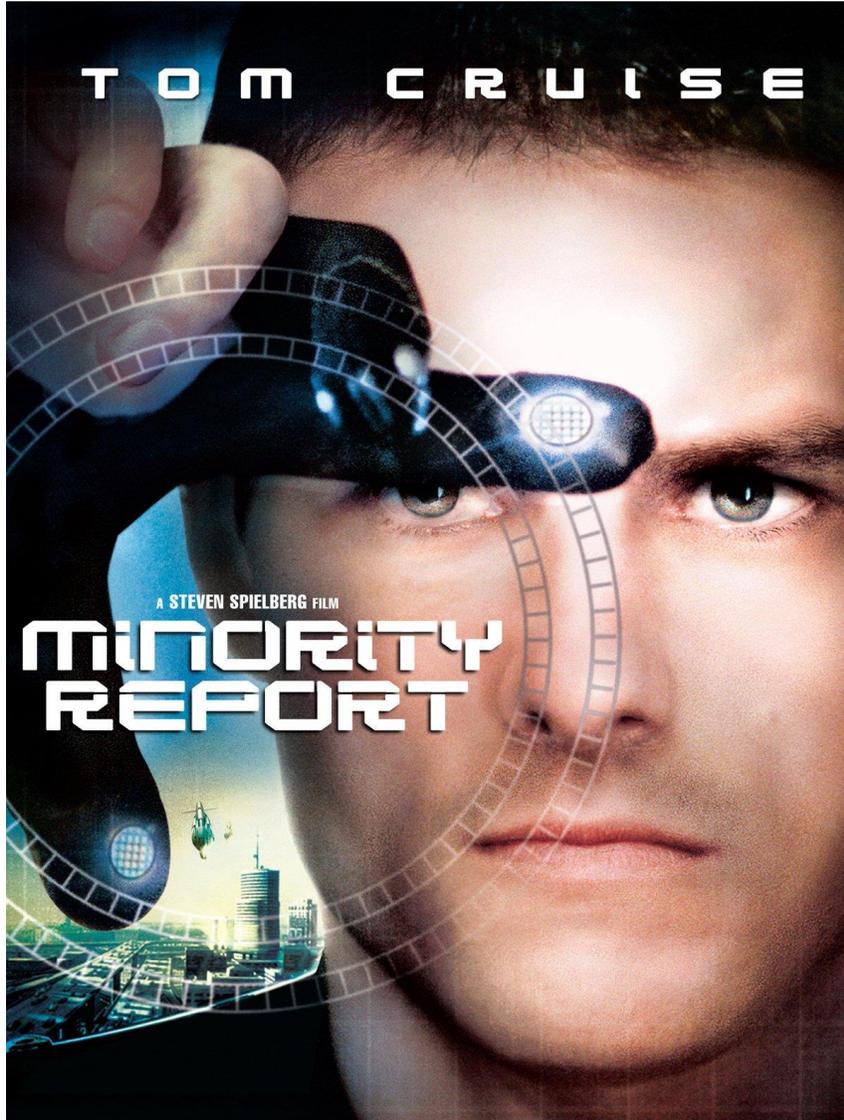


Public health organizations can then be informed, and specific **intervention strategies** can be developed in locations where they can have the greatest impact.



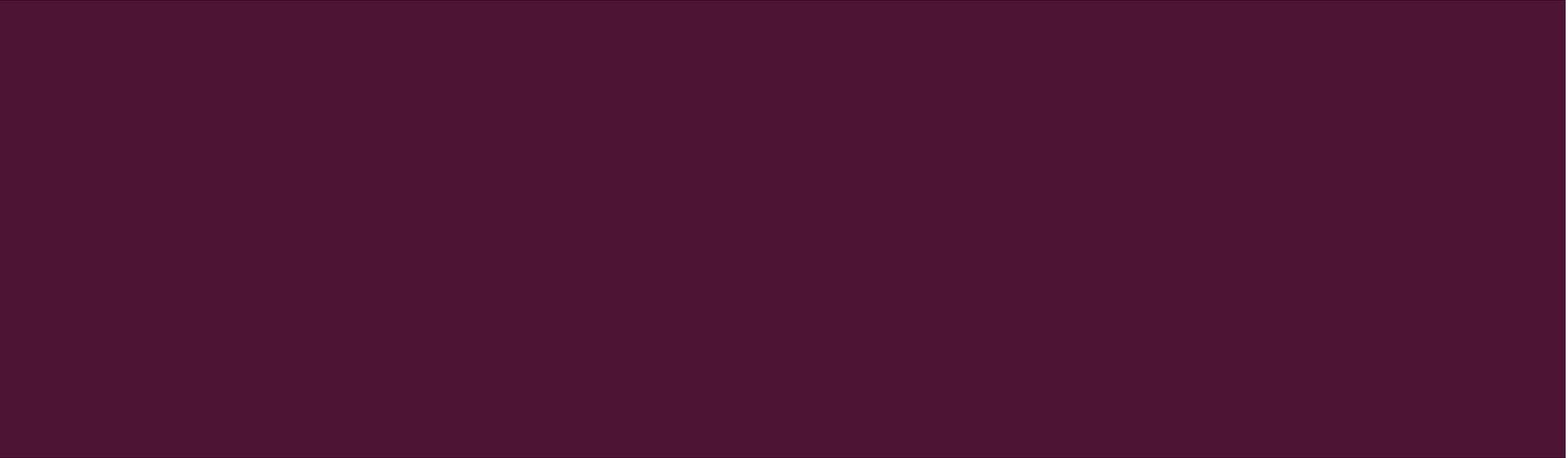
Genoma humano impresso (ACTG):
262.000 páginas
450 kg

PROBLEMAS





ETL – EXTRACT, TRANSFORM, LOAD



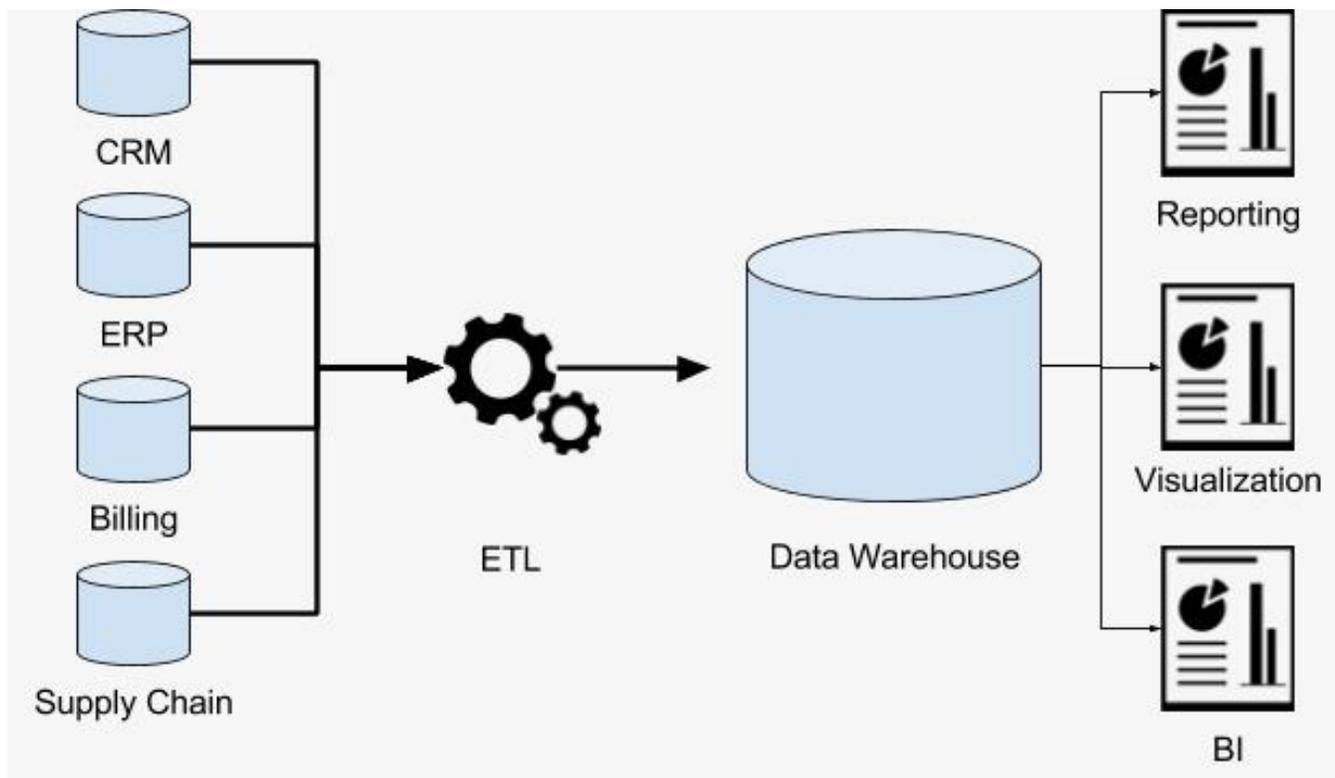
INTRODUÇÃO

- **ETL**, ou *Extract Transform Load*, são práticas ou serviços dentro de um software cuja função é a extração e estruturação de dados de diversos sistemas para uma análise mais transparente.



ETL

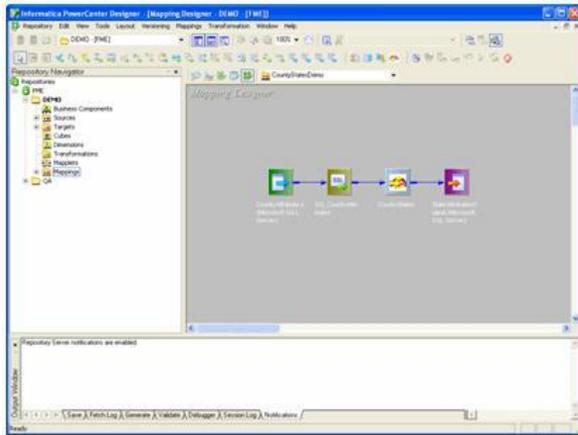
- É uma abordagem alternativa, projetada para jogar o processamento para o banco de dados, de modo a aprimorar a performance, trazendo dados mais completos.



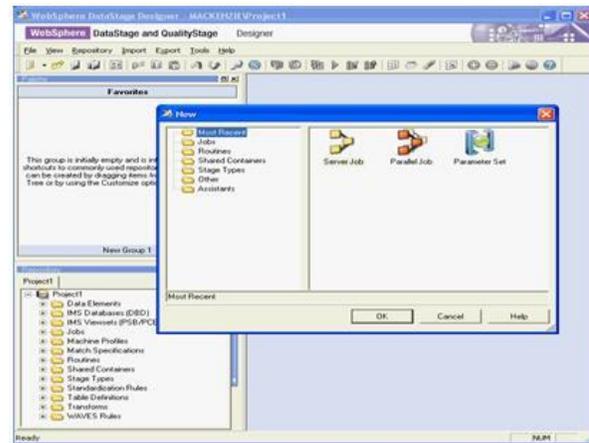
ETL

- O processo de ETL é composto de três partes:
 1. **Extração:** determina a origem, seus metadados, tamanhos e disponibilidades. Geralmente os dados são heterogêneos.
 2. **Transformação:** nesta etapa são aplicadas as regras de negócio, seja selecionando apenas algumas colunas da lista de dados de origem, conversão de caracteres de um determinado campo em outro padrão, derivando novas colunas, ordenando as informações, etc.
 3. **Carga:** a carga é o passo final, onde os dados (depois de extraídos e transformados) são carregados no destino, geralmente um *data warehouse*. *E então prontos para ser analisados.*

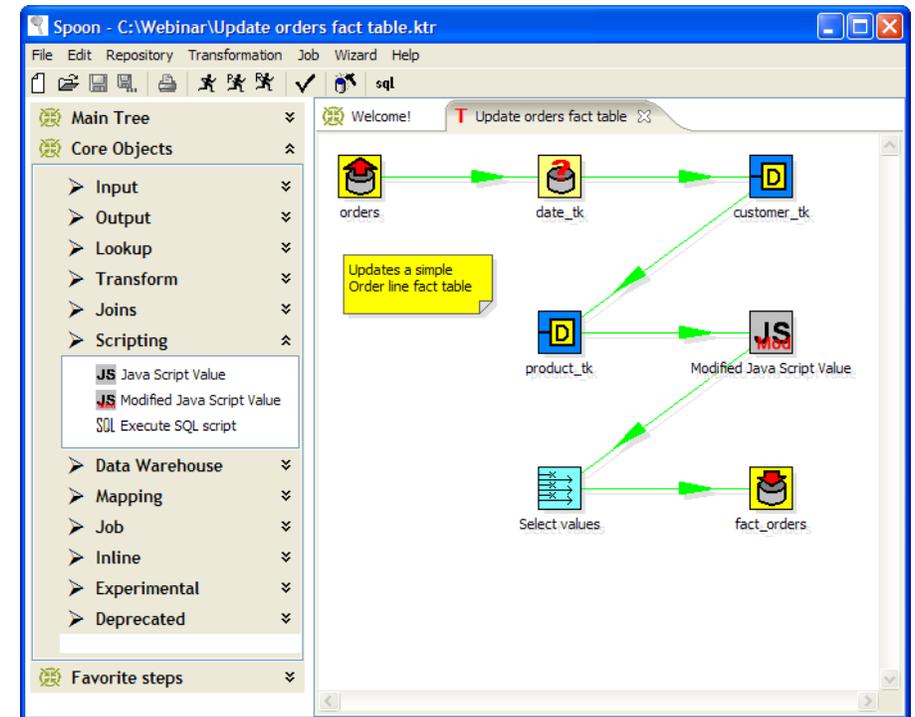
FERRAMENTAS DE ETL



PowerCenter



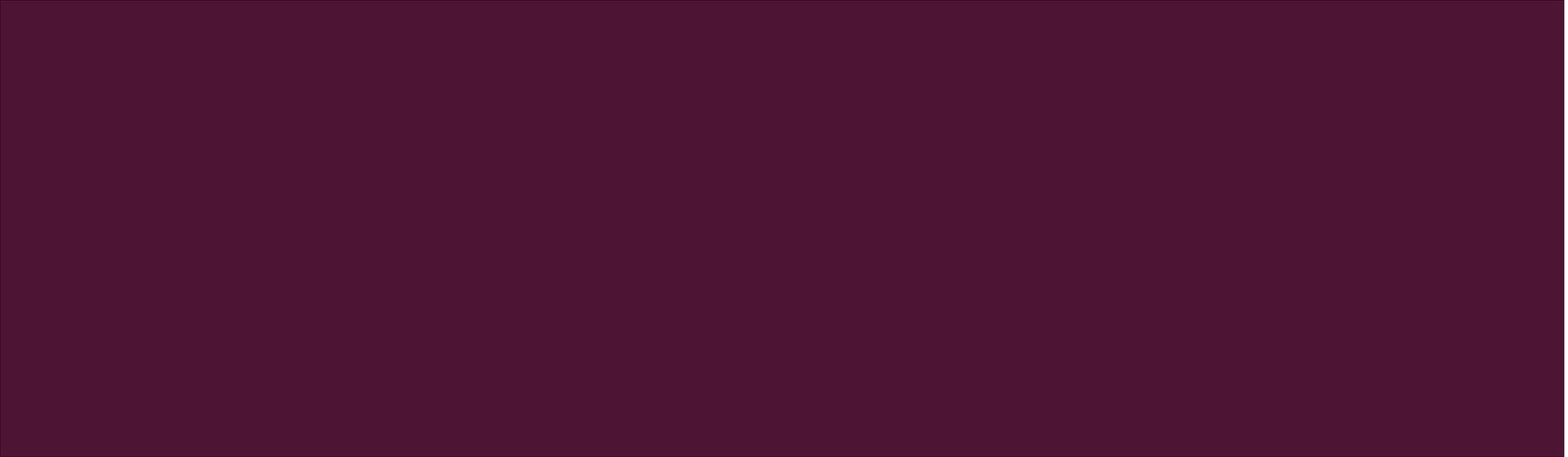
DataStage



Pentaho Data Integration



DATA MINING



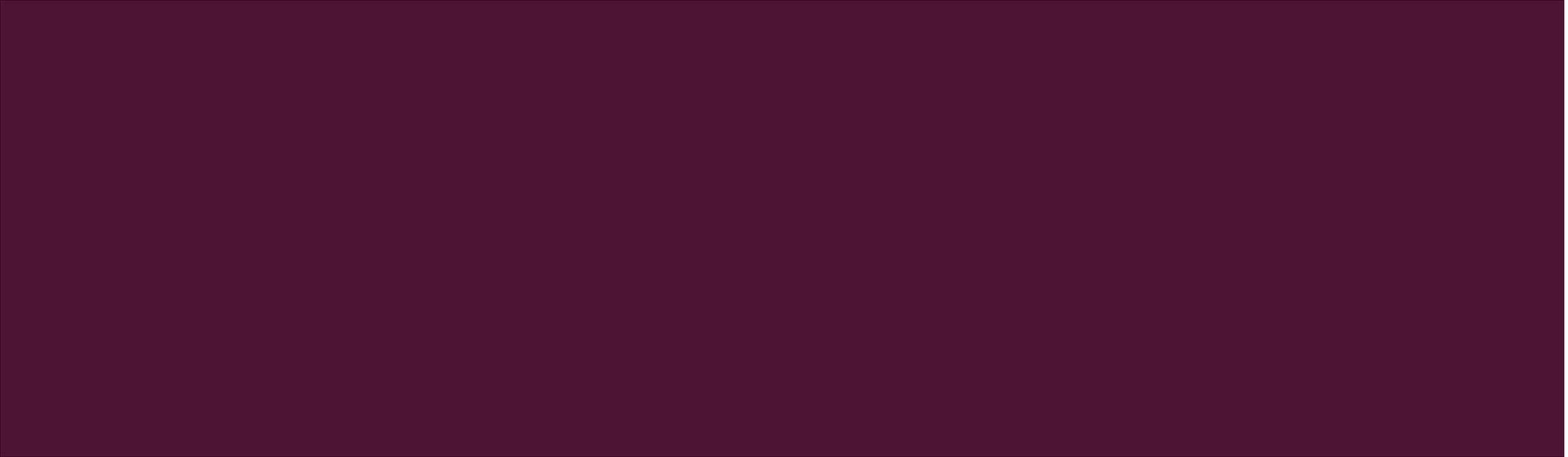
DEFINIÇÕES



- Explorar grandes quantidades de dados;
 - Buscar padrões;
- Detectar relacionamentos entre variáveis;



ETAPAS DO DATA MINING



DEFINIÇÃO DO PROBLEMA



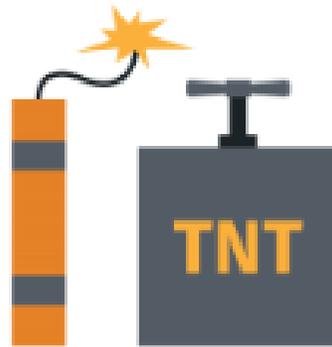
- Definir os Objetivos do processo de Mineração;
 - Metas a serem atingidas;

INTEGRAÇÃO DE DADOS



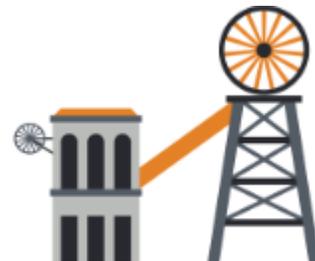
- Identificação das fontes de dados;
- Coleta de dados;

SELEÇÃO DOS DADOS



- Descartar dados que não são necessários;

LIMPEZA DE DADOS



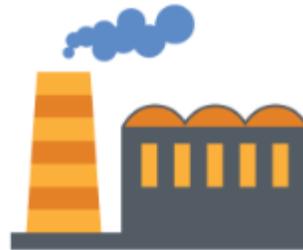
- Limpeza aplicada devido a problemas obtidos na coleta, tais como:
 - Dados inseridos erroneamente;
 - Falta de verificação de restrições;
 - Conflitos de nomes;
 - Contradições;
 - Substituições de dados.

TRANSFORMAÇÃO DE DADOS



- Dados filtrados ainda incapazes de ser utilizados;
- São consolidados adequadamente, formatados e convertidos em formato pronto para uso;

MINERAÇÃO DE DADOS



- Dados foram coletados e tratados;
- Aplicado técnicas de mineração para identificar padrões, correlações ou relacionamentos entre os conteúdos da base.

AVALIAÇÃO DOS PADRÕES



- Padrões identificados na etapa anterior são inspecionados e avaliados;
- Descoberta de quais padrões são importantes para serem utilizados;

APRESENTAÇÃO DO CONHECIMENTO



- A informação do passo anterior é transformada em conhecimento;
- O conhecimento é o produto final apresentado aos stakeholders.



OBRIGADO!

