

Formação Engenheiro de Dados

**Sua Carreira Elevada
a Outro Nível**



Seja muito Bem-Vindo(a)



Data Science Academy

www.datascienceacademy.com.br



Data Science Academy



O que é a Formação Engenheiro de Dados?



Data Science Academy

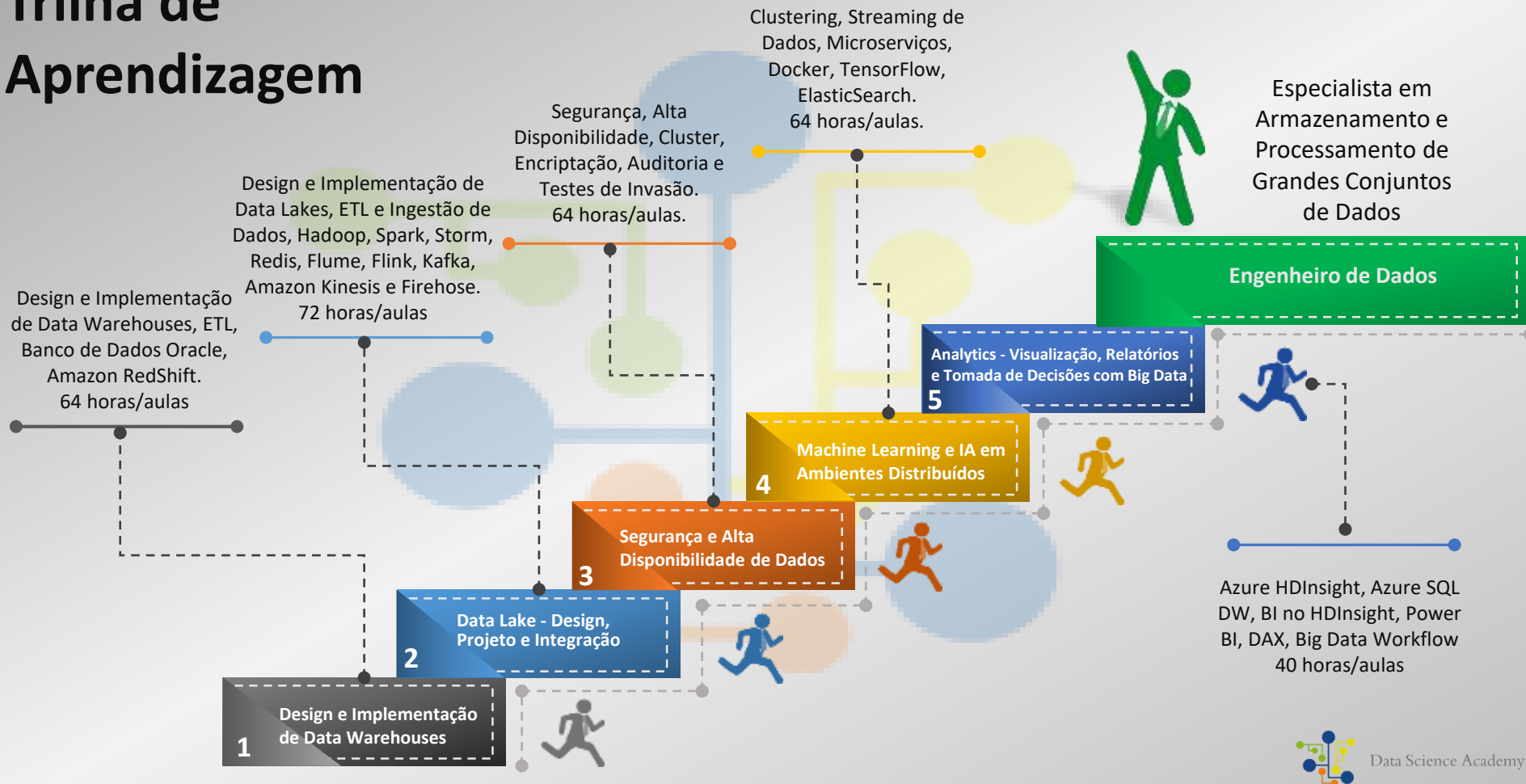
Treinamentos 100% online e 100% Português



Nossa Trilha de Aprendizagem



Trilha de Aprendizagem



Esperamos de Você



Avaliação Final em Cada Curso da Formação
Certificado Engenheiro de Dados

Cenários e Troubleshooting
Sistema Operacional Linux
Oracle, AWS RedShift, Ecosistema Hadoop, Azure



Laboratórios Práticos
Materiais Complementares
Referências e Links Úteis



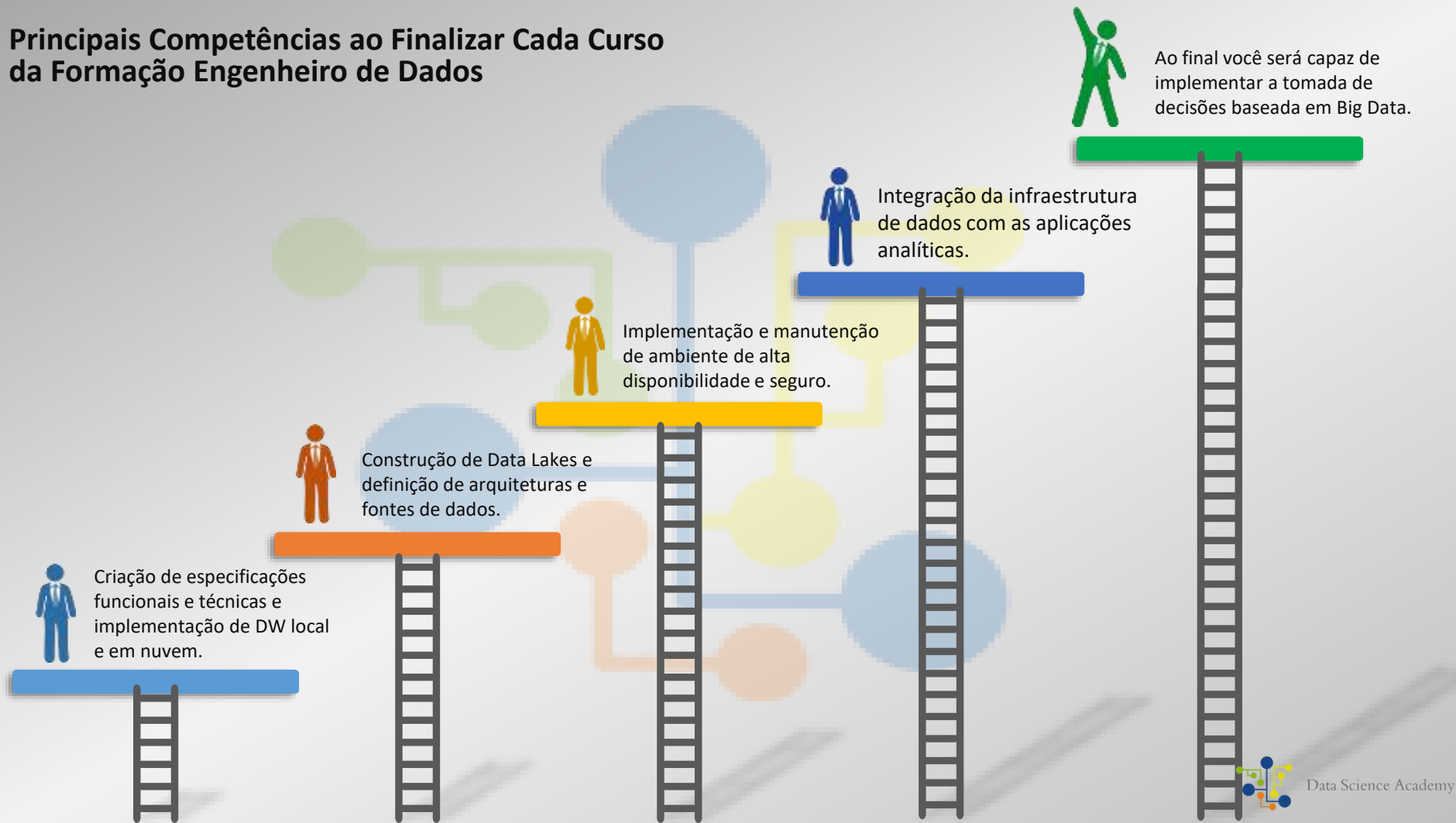
Módulos de Aprendizagem
Quizzes e Exercícios
Fórum de Discussão



Recomendamos dedicação diária de 2 a 4 horas,
de acordo com a sua disponibilidade.



Principais Competências ao Finalizar Cada Curso da Formação Engenheiro de Dados



Como Obter Sucesso Neste Treinamento?

Design e Implementação

- 304 horas de conteúdo de qualidade e alto nível
- Assuntos integrados e complementares
- Curso Bônus – Sistema Operacional Linux
- Construção de um Serverless Data Lake
- Estudo de caso para construção de um MS SQL DW

Uma sequência de cursos profissionais para você aprender na prática como implementar um ambiente de armazenamento e processamento distribuído de Big Data e assim se tornar um profissional com alta empregabilidade.



Seja bem-vindo(a) ao curso Design e Implementação de Data Warehouses

Este é o primeiro curso da
Formação Engenheiro de
Dados





Introdução



Design e Implementação de Data Warehouses



Instrutores



Apresentação dos Instrutores



Data Science Academy



Apresentação da Empresa



Módulos



Apresentação dos Módulos do Curso



Metodologia de Trabalho



Apresentação da Metodologia de Trabalho



Data Science Academy



Data Science Academy

Nossa Academia

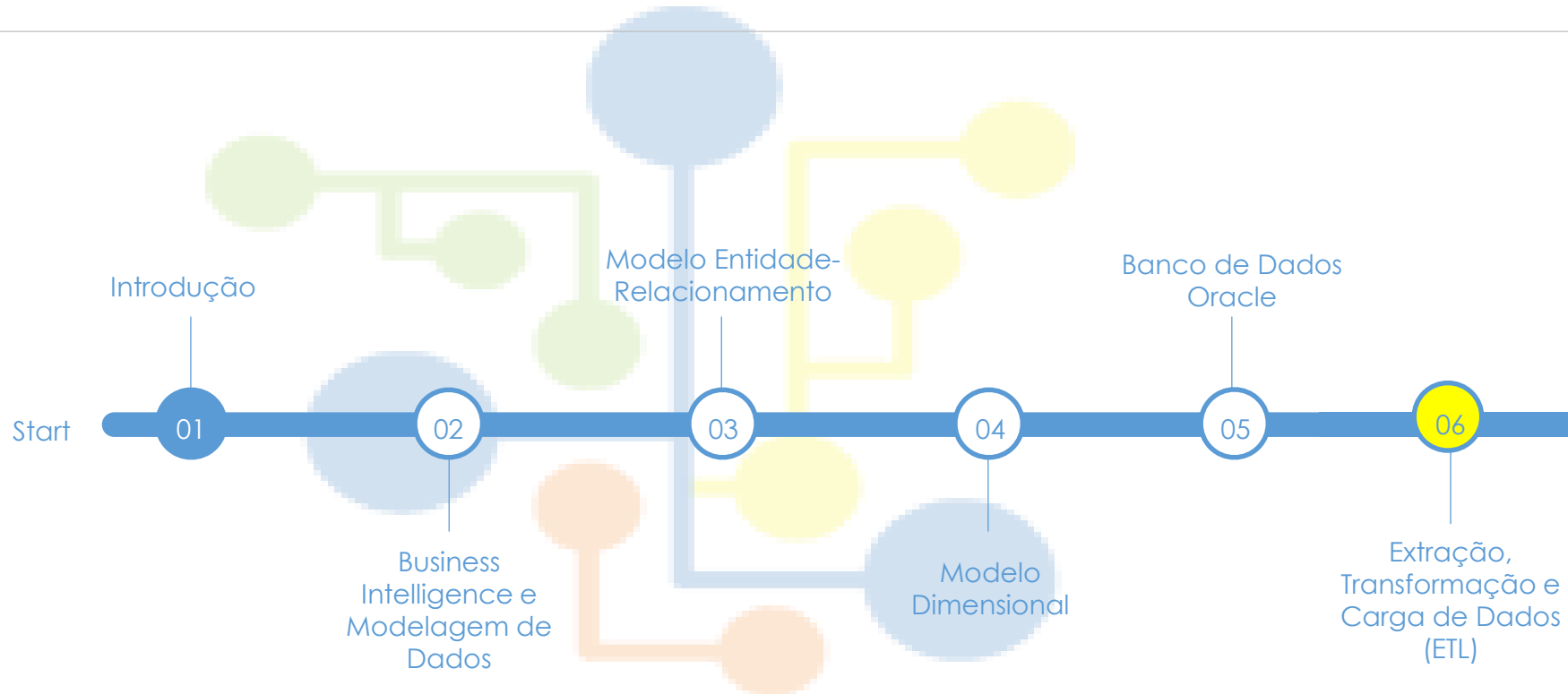
A Data Science Academy (DSA) é um portal de ensino online especializado em Big Data, Machine Learning, Inteligência Artificial, Desenvolvimento de Chatbots e tecnologias relacionadas. Nosso objetivo é fornecer aos alunos conteúdo de alto nível por meio do uso de computador, tablet ou smartphone, em qualquer lugar, a qualquer hora, 100% online e 100% em português.

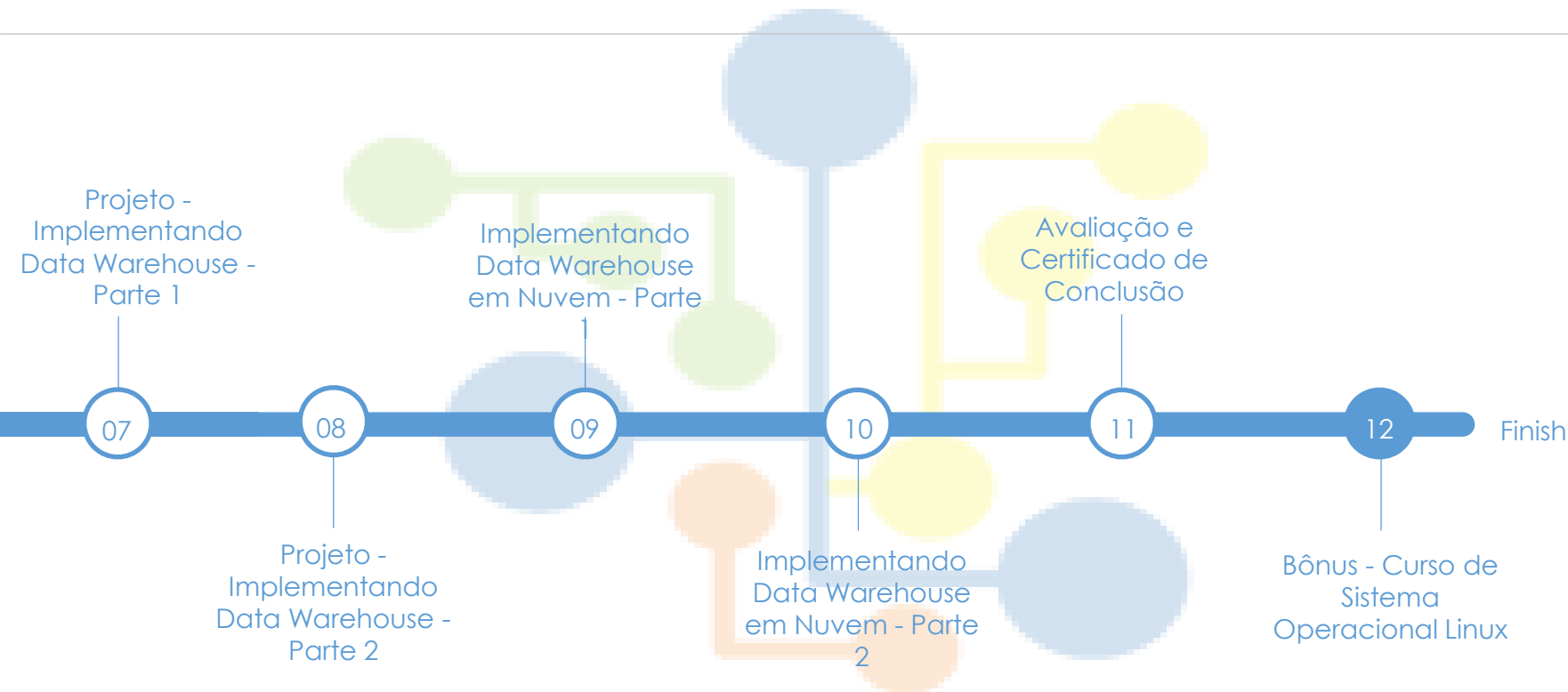


Data Science Academy - Localização

No Brasil e no Mundo

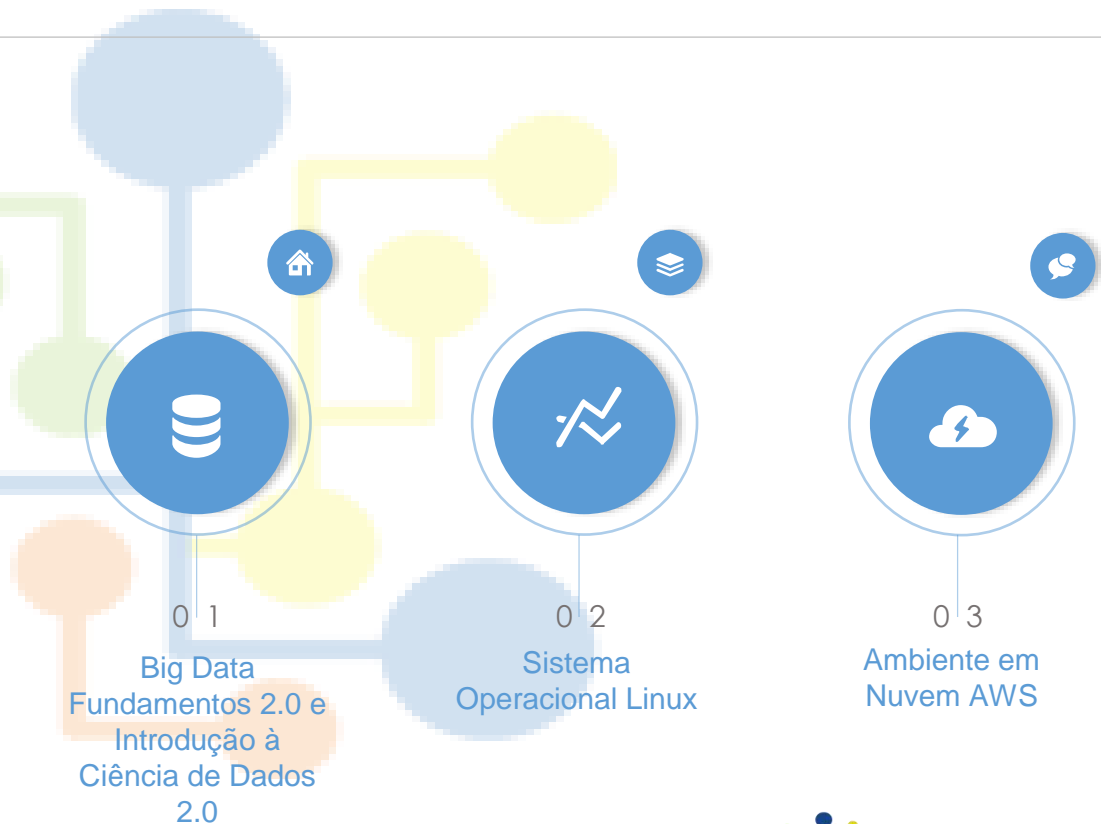




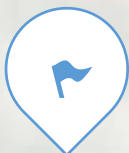


Curso técnico, voltado para profissionais de tecnologia.

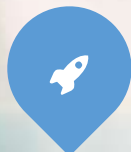
Esperamos que você tenha um bom conhecimento em Sistemas Operacionais (especialmente sistema Linux) e este curso oferece como bônus um curso completo de Introdução ao Sistema Operacional Linux, para ajudar aqueles que não estejam confortáveis com este sistema operacional. Conhecimento de como funciona a Infraestrutura de TI e noções de Business Intelligence.



Teoria



Dinamismo



Leitura



Muito
Conteúdo



Alguns
Conceitos
Avançados



Objetividade





Aulas em Vídeo

Exposição teórica sobre o conteúdo.



Laboratórios

Cenários e Troubleshooting



Pesquisa Adicional

Pesquisa sobre temas relacionados.
Bibliografia, referências e links úteis ao final de cada capítulo.



Quizzes e Exercícios

Quizzes e exercícios para testar seu conhecimento.



Divirta-se

Comunique-se, aprenda e divirta-se em nossa Comunidade.

Bibliografia

Leia a bibliografia adicional, acesse os links úteis e realize os quizzes ao final dos capítulos.

1 a 2 horas

de dedicação por semana.

Leitura do Material

E-books e material complementar!

Interação

Utilize nossas Apps e interaja na rede com outros alunos, no fórum exclusivo e na timeline da Comunidade em nosso site.





3 Tentativas



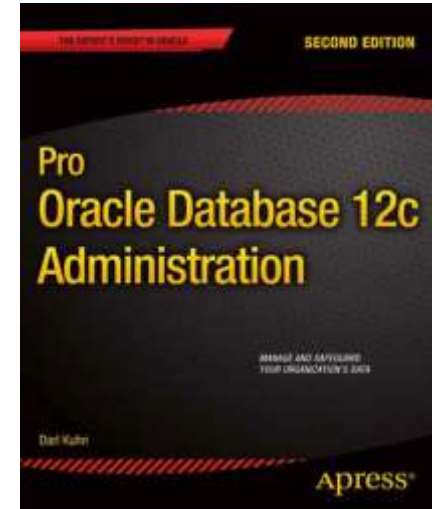
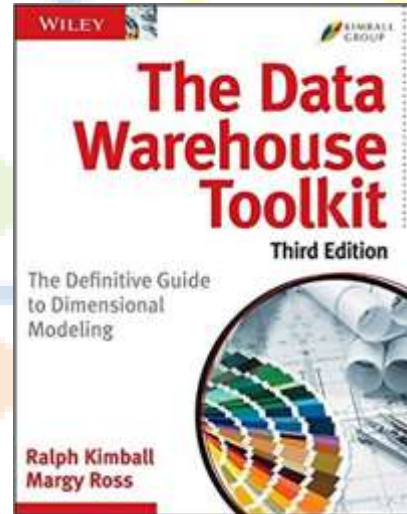
50 Questões

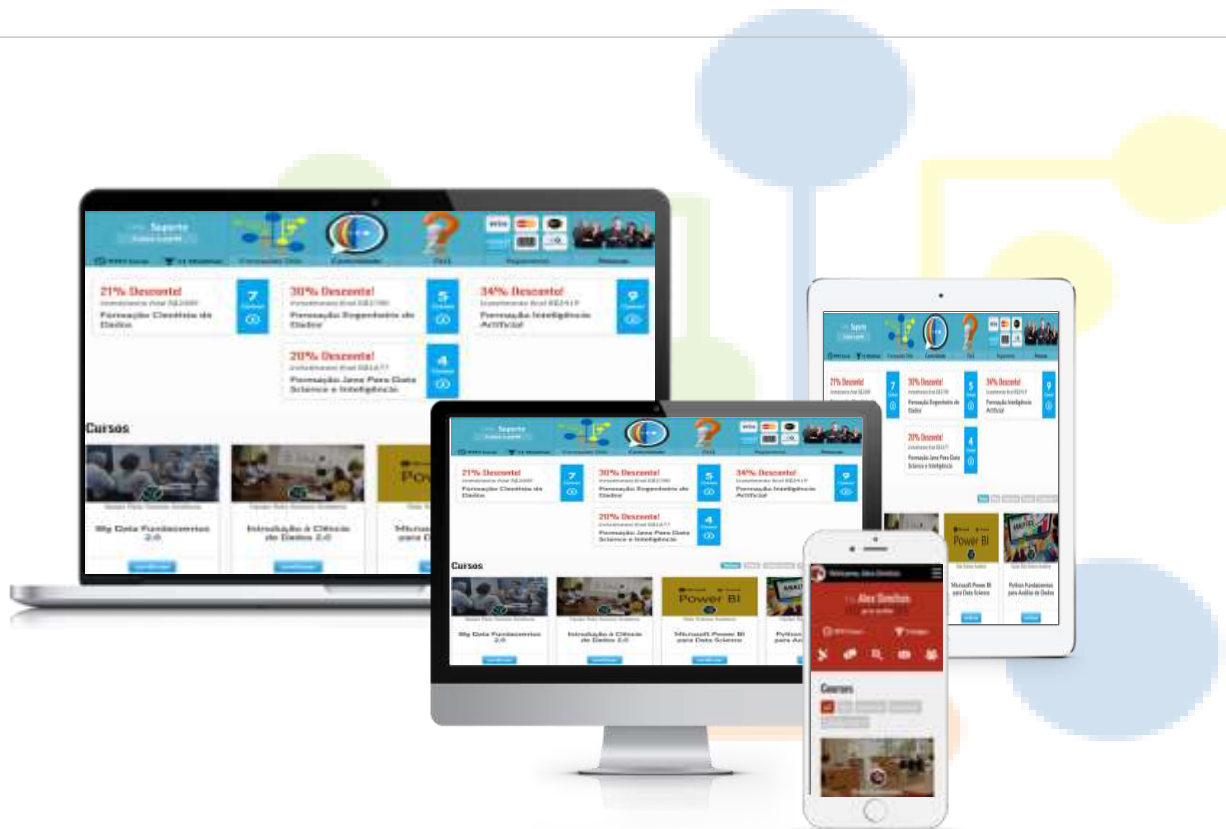


70 % Aproveitamento



Bibliografia Sugerida







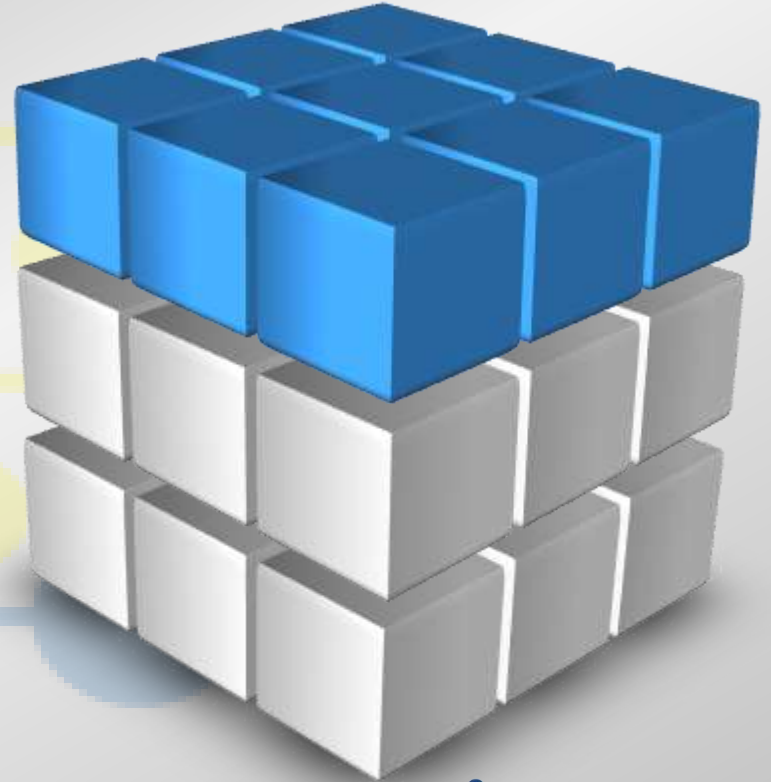


O QUE É ENGENHARIA DE DADOS?



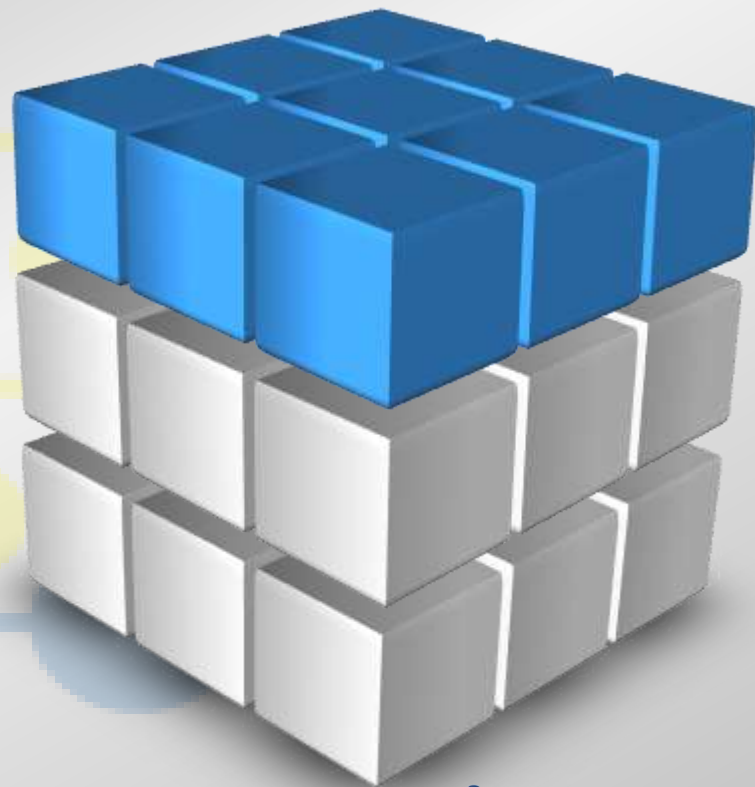
O QUE É ENGENHARIA DE DADOS?

Se a Engenharia é a prática de utilizar ciência e tecnologia para criar e construir sistemas que resolvem problemas, podemos definir a Engenharia de Dados como sendo a área da Engenharia dedicada a processar e tratar dados para aplicações que utilizarão Big Data.



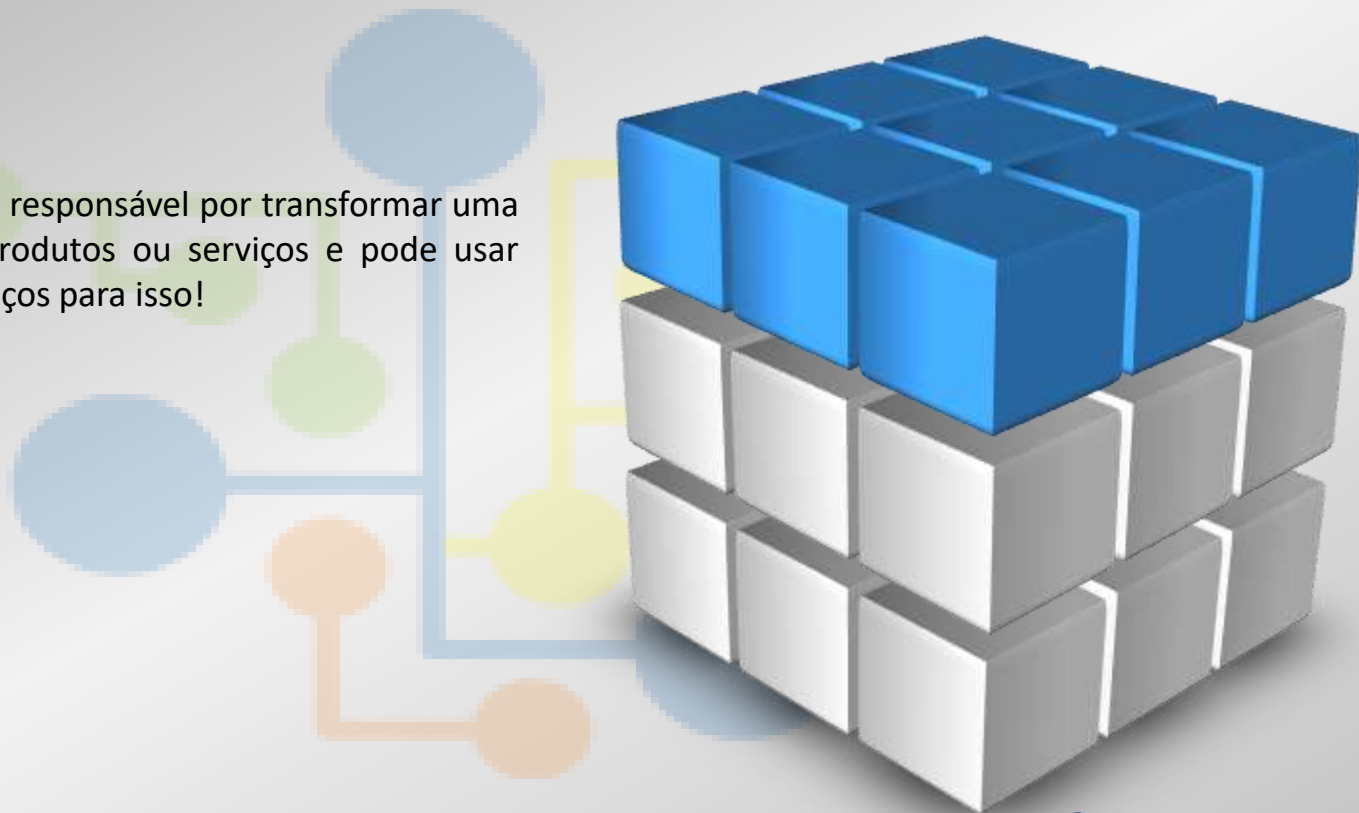
O QUE É ENGENHARIA DE DADOS?

Engenheiros de Dados utilizam conhecimento em ciência da computação para criar sistemas e resolver problemas de processamento de dados em tempo real e manipular quantidades imensas de dados.



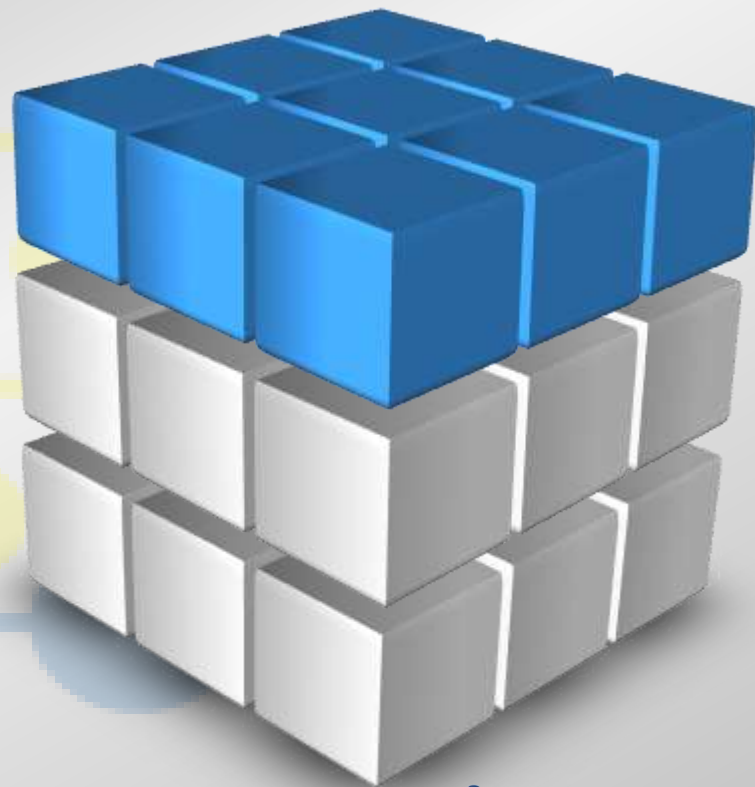
O QUE É ENGENHARIA DE DADOS?

O Engenheiro de Dados é responsável por transformar uma aplicação analítica em produtos ou serviços e pode usar arquitetura de Microserviços para isso!



O QUE É ENGENHARIA DE DADOS?

O propósito do Engenheiro de Dados, é fornecer soluções e plataformas de Big Data modulares e escaláveis para que os Cientistas de Dados possam fazer seu trabalho de análise.





ENGENHEIRO DE DADOS X CIENTISTA DE DADOS

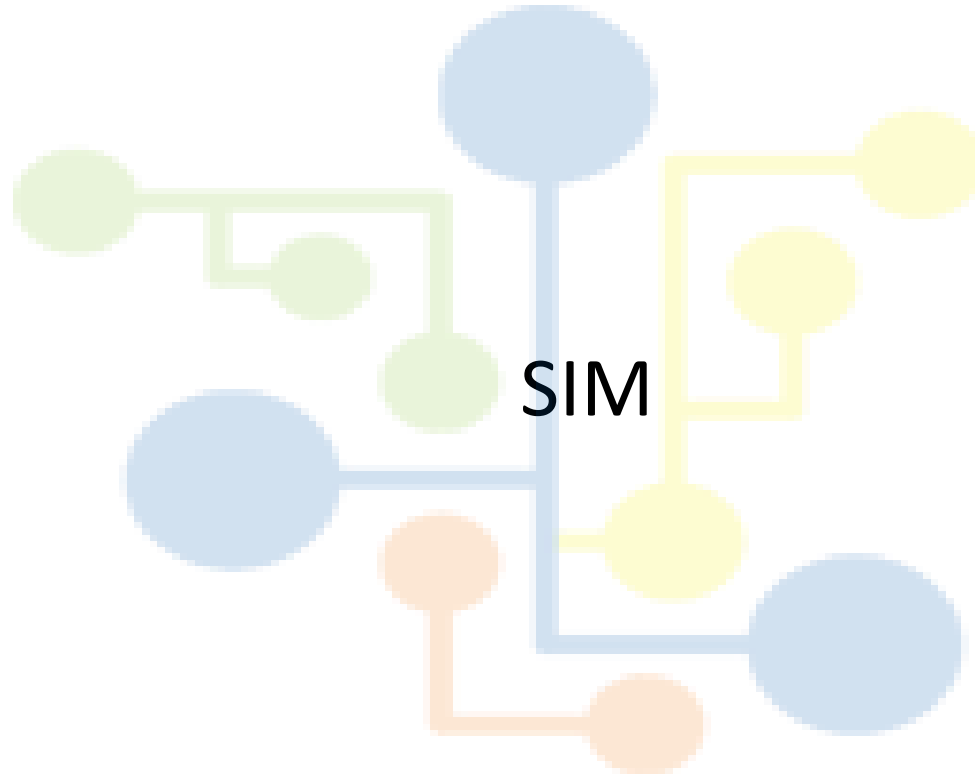


ENGENHEIRO DE DADOS X CIENTISTA DE DADOS

É possível um único profissional
trabalhar como Cientista de Dados e
Engenheiro de Dados?



ENGENHEIRO DE DADOS X CIENTISTA DE DADOS



ENGENHEIRO DE DADOS X CIENTISTA DE DADOS

Se você acredita que não tem capacidade para trabalhar com diversas tecnologias, você está certo!

Se você acredita que tem capacidade para trabalhar com diversas tecnologias, você também está certo!

O único limite é o que você impõe a si mesmo!



ENGENHEIRO DE DADOS X CIENTISTA DE DADOS

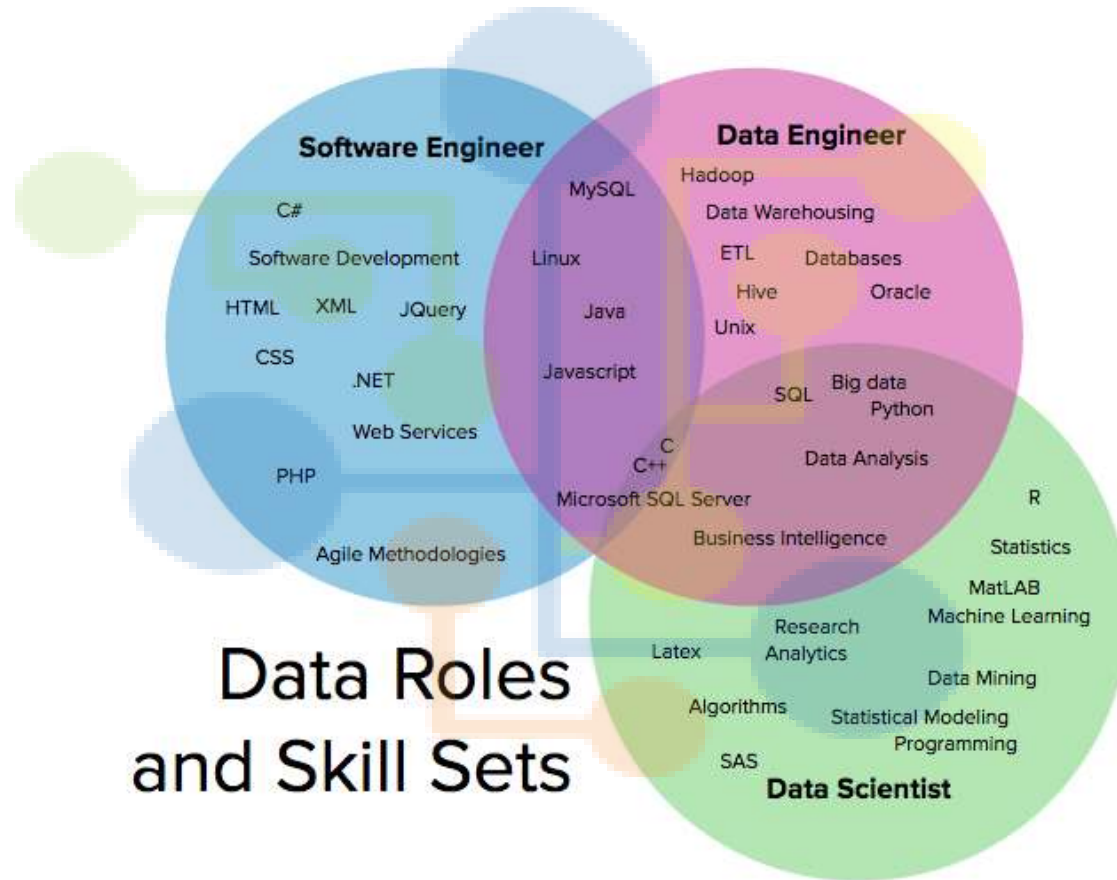
É necessário um curso superior para trabalhar como Engenheiro de Dados?

ENGENHEIRO DE DADOS X CIENTISTA DE DADOS

Não existe um curso superior que capacite alguém em Engenharia de Dados, mas uma graduação na área de Ciências Exatas pode ajudar na preparação deste perfil profissional!



ENGENHEIRO DE DADOS X CIENTISTA DE DADOS



ENGENHEIRO DE DADOS X CIENTISTA DE DADOS

THE DATA SCIENCE HIERARCHY OF NEEDS

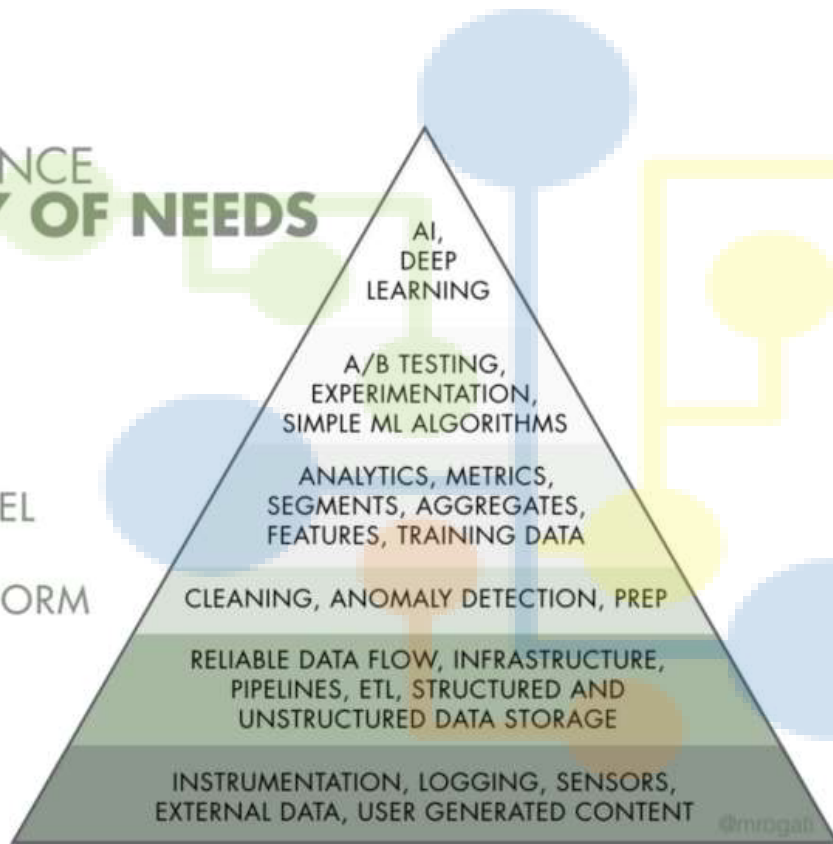
LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



Cientista de Dados

Formação Cientista de Dados,
Formação IA e Formação Java.

Engenheiro de Dados

Formação Engenheiro de Dados.



Data Science Academy



O QUE É DATA WAREHOUSE?

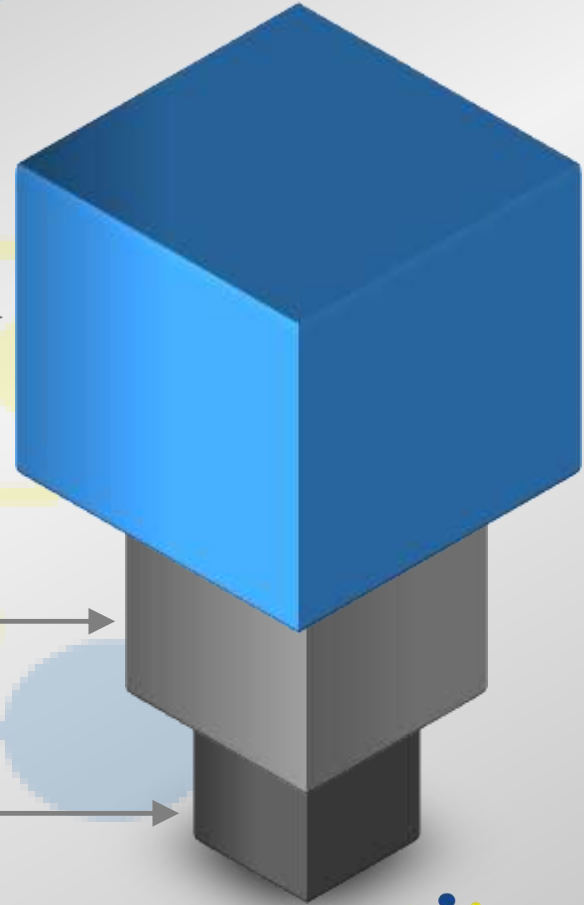


O QUE É DATA WAREHOUSE?

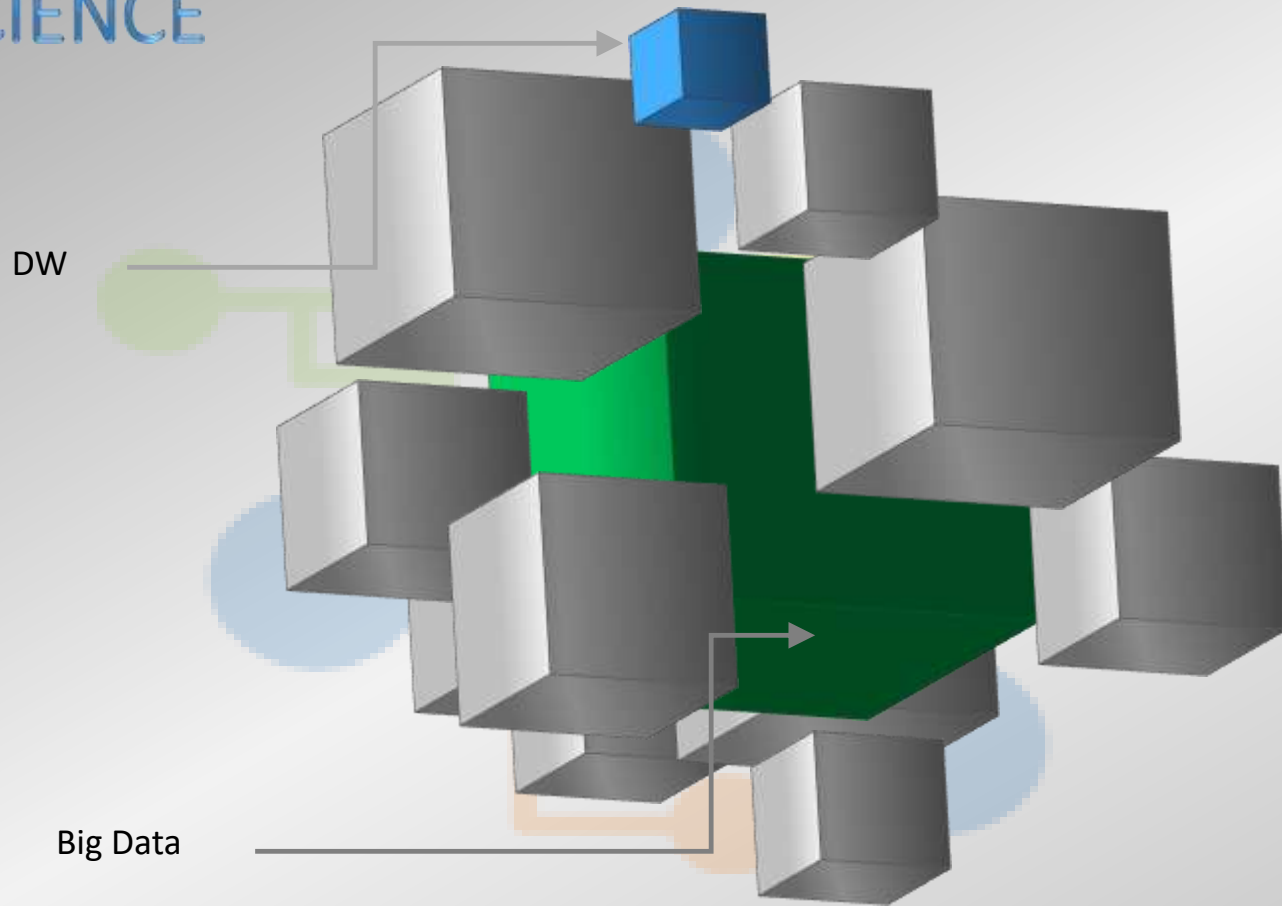
DW e BI
(Data Warehouse e
Business Intelligence)

DSS
(Decision Support
System)

EIS
(Executive Information
System)



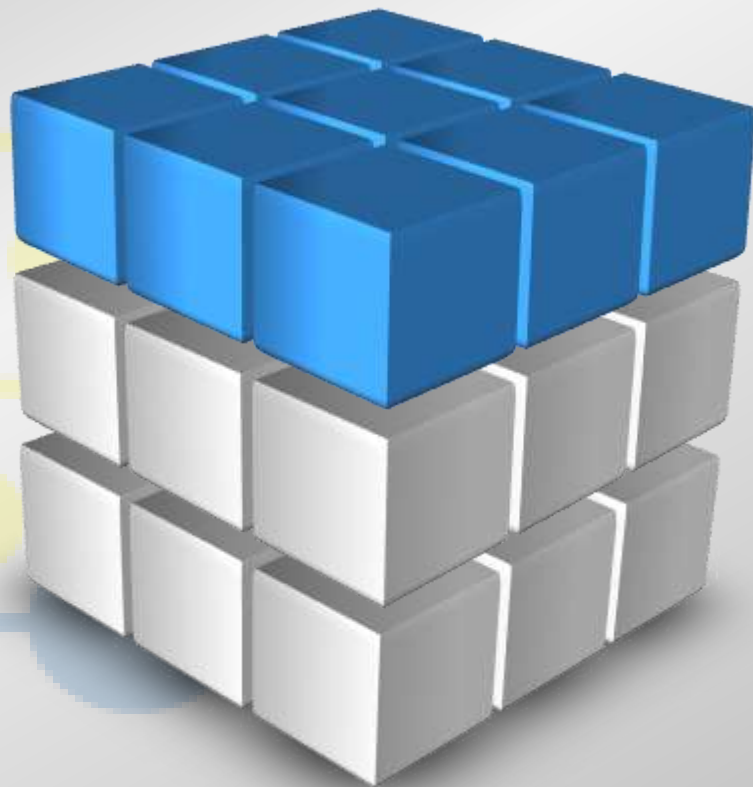
DATA SCIENCE



O QUE É DATA WAREHOUSE?

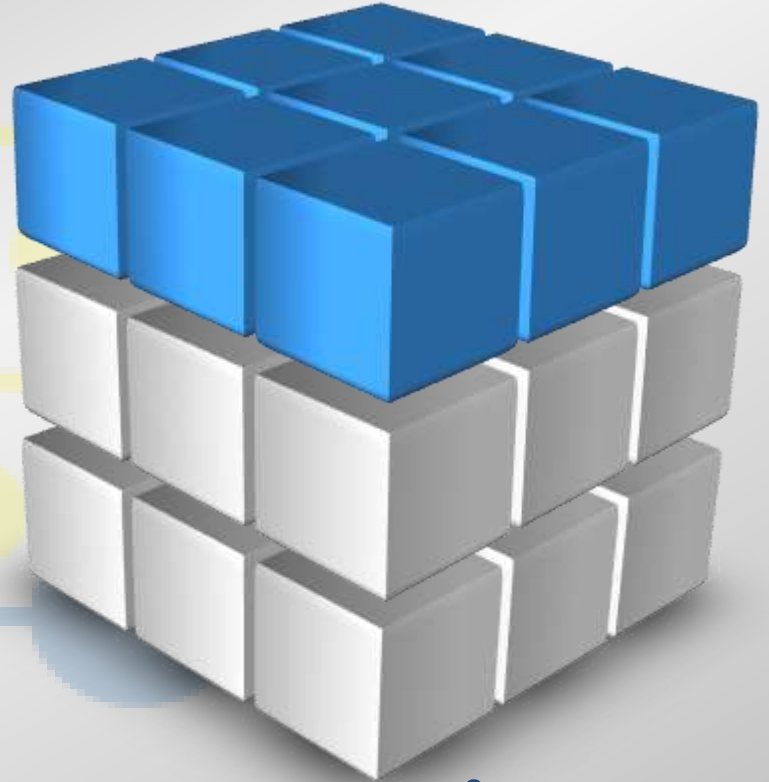
Um Data Warehouse é um banco de dados, usado para armazenar informações relativas às atividades de uma organização de forma consolidada.

Possibilita a análise de grandes volumes de dados, que são coletados a partir de sistemas transacionais (OLTP – Online Transaction Processing).



O QUE É DATA WAREHOUSE?

O Data Warehouse é organizado para dar suporte à tomada de decisões estratégicas da empresa.





DW E BIG DATA

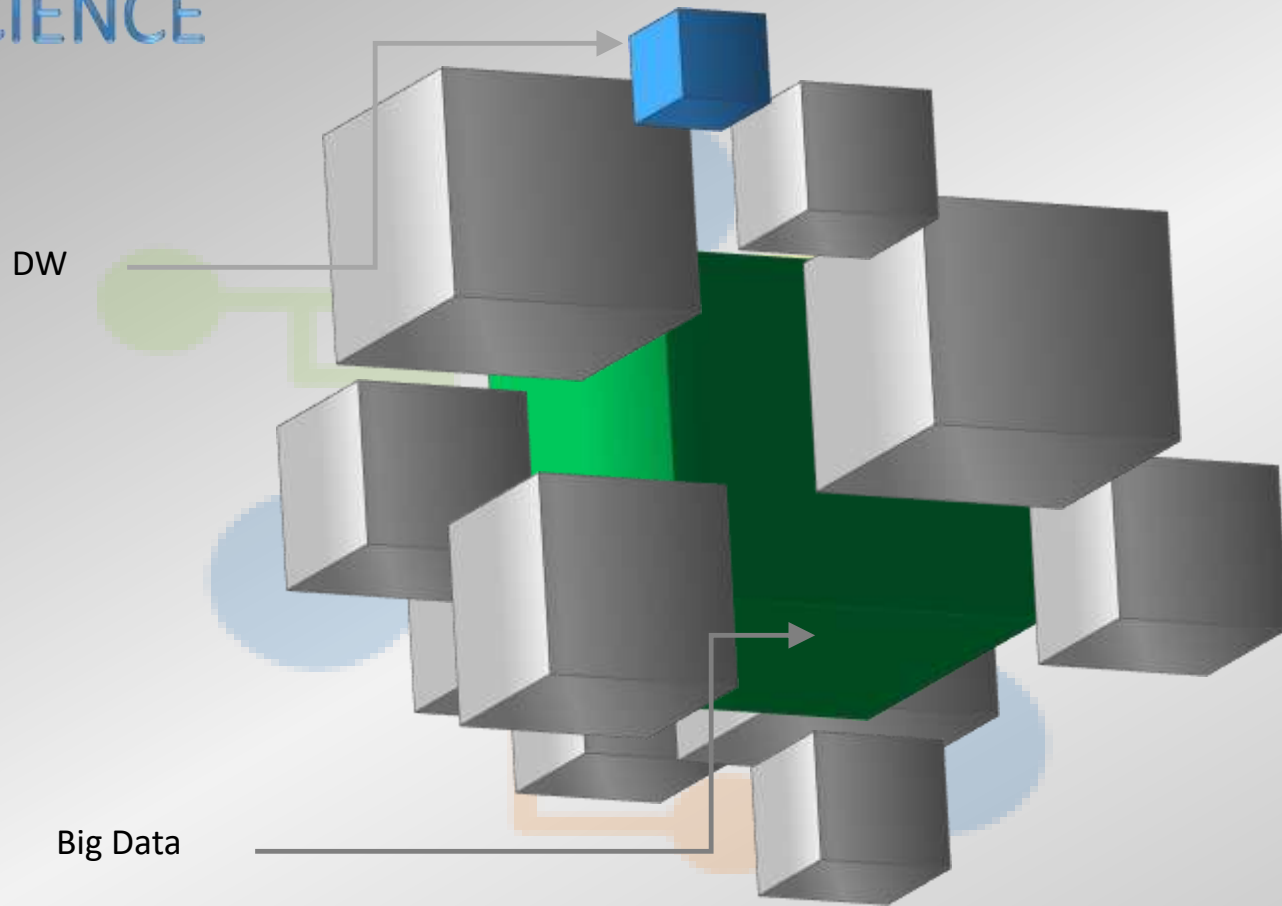


An abstract graphic in the background consisting of several colored circles (blue, green, yellow, orange) connected by thin lines, creating a network-like structure.

Data Warehouses Ainda São Relevantes na Era do Big Data?



DATA SCIENCE



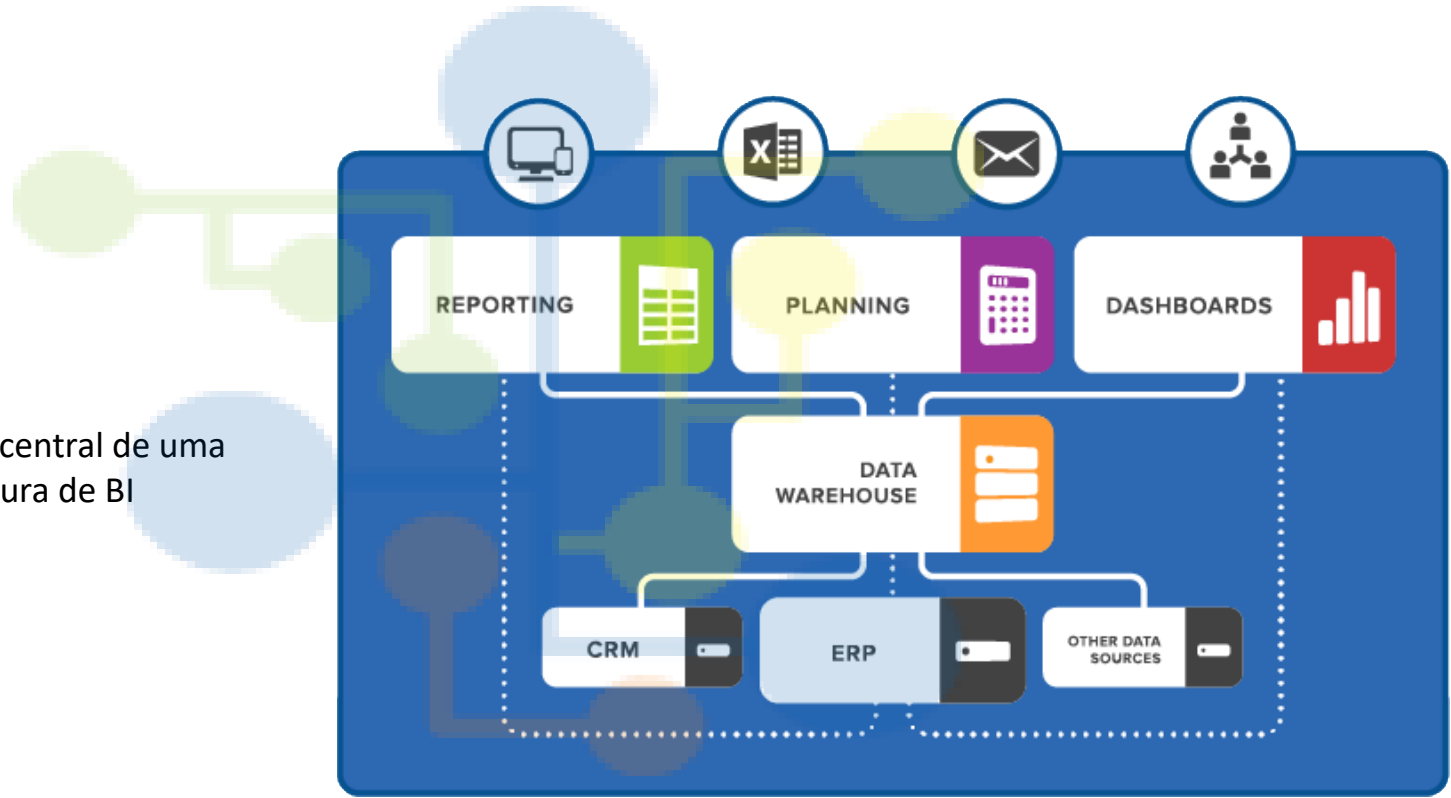


DATA WAREHOUSES, BUSINESS INTELLIGENCE, DATA MARTS E OLTP

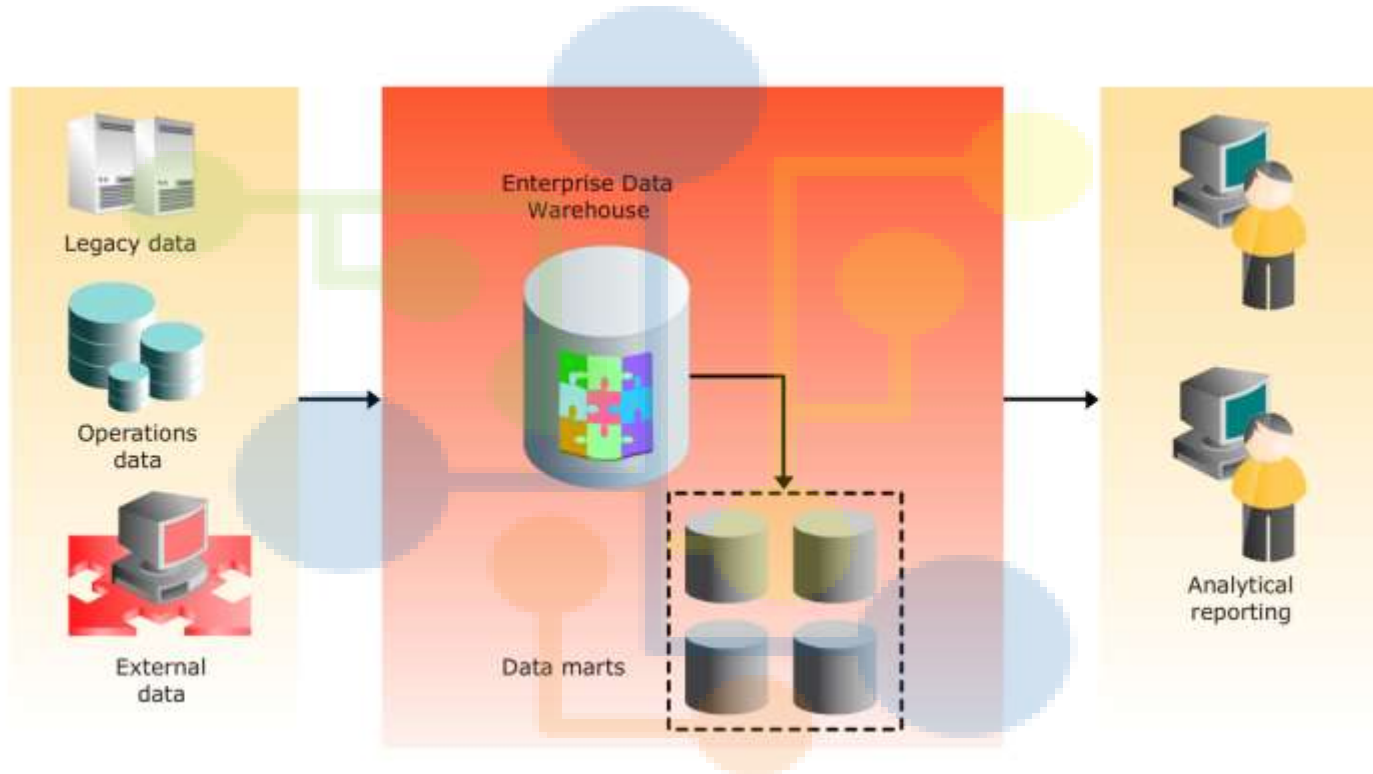


DATA WAREHOUSES, BUSINESS INTELLIGENCE, DATA MARTS E OLTP

O DW é o ponto central de uma infraestrutura de BI



DATA WAREHOUSES, BUSINESS INTELLIGENCE, DATA MARTS E OLTP



OLTP

DW e Data Mart

BI



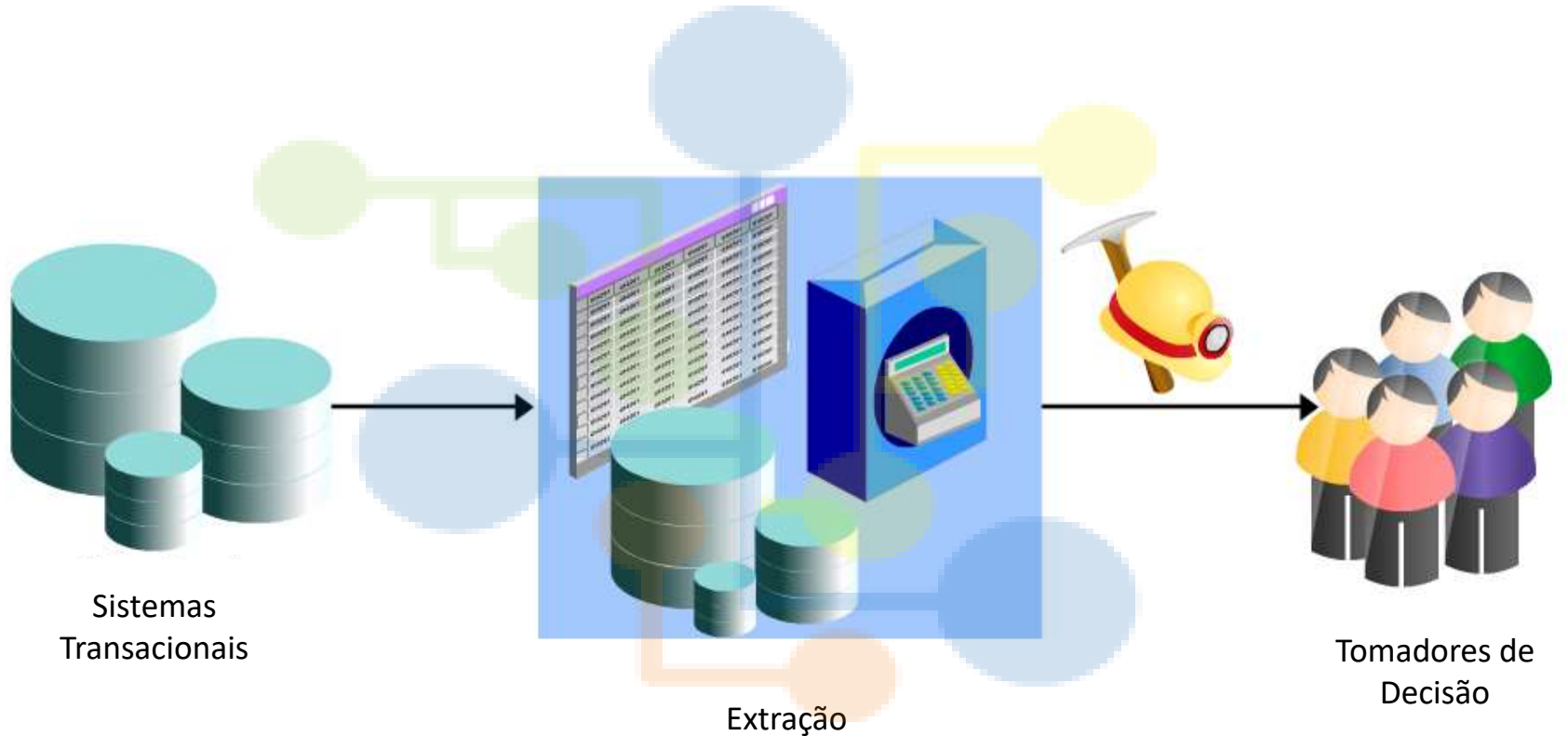
Data Science Academy



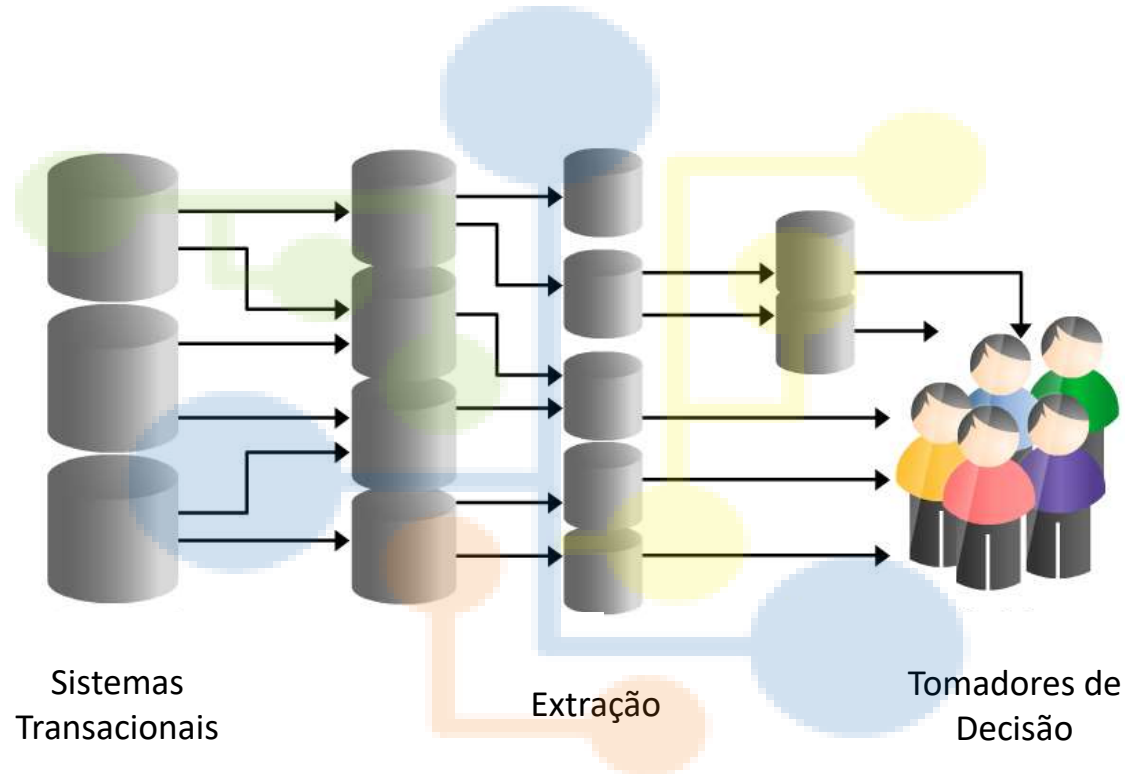
QUANDO USAR UM DATA WAREHOUSE?



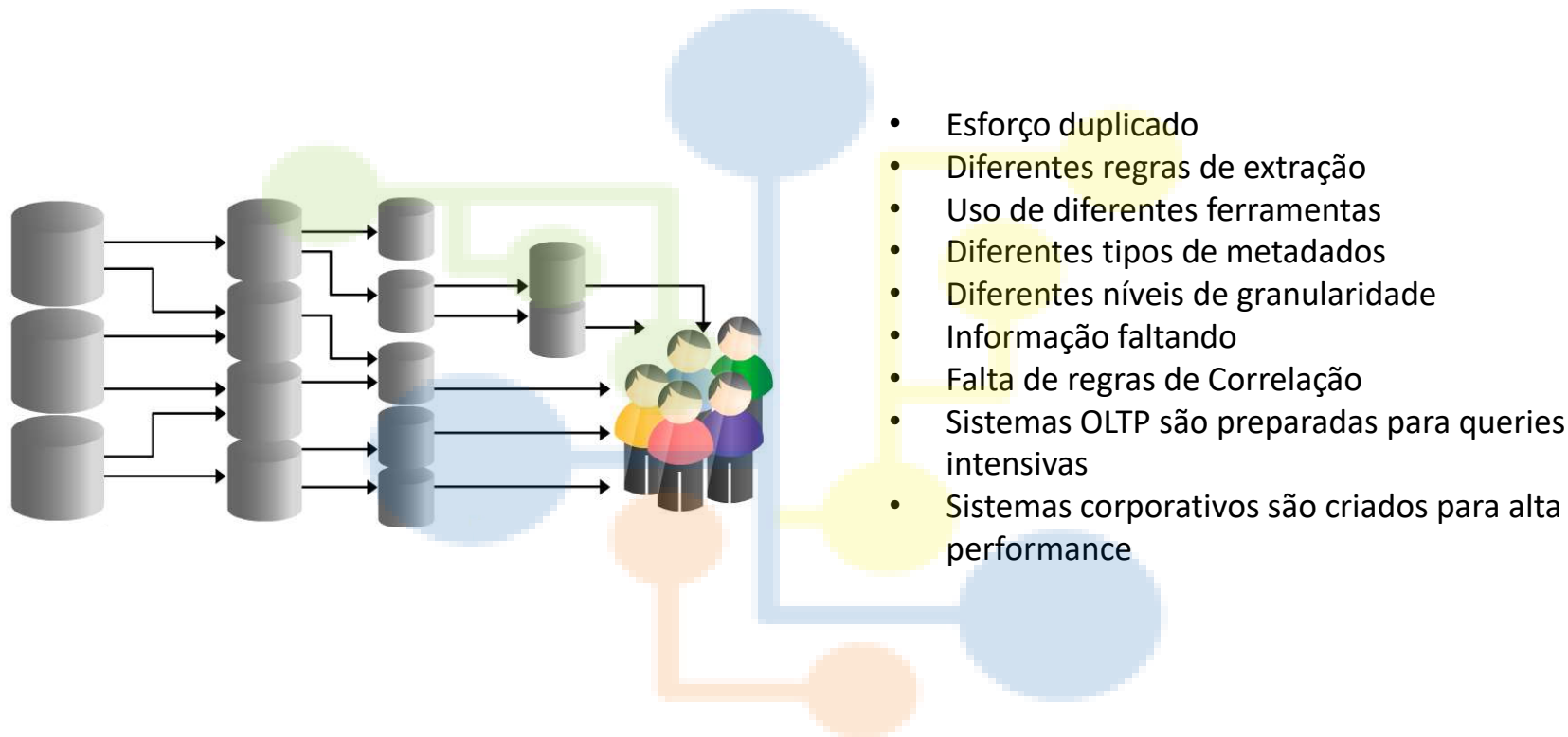
QUANDO USAR UM DATA WAREHOUSE?



QUANDO USAR UM DATA WAREHOUSE?



QUANDO USAR UM DATA WAREHOUSE?



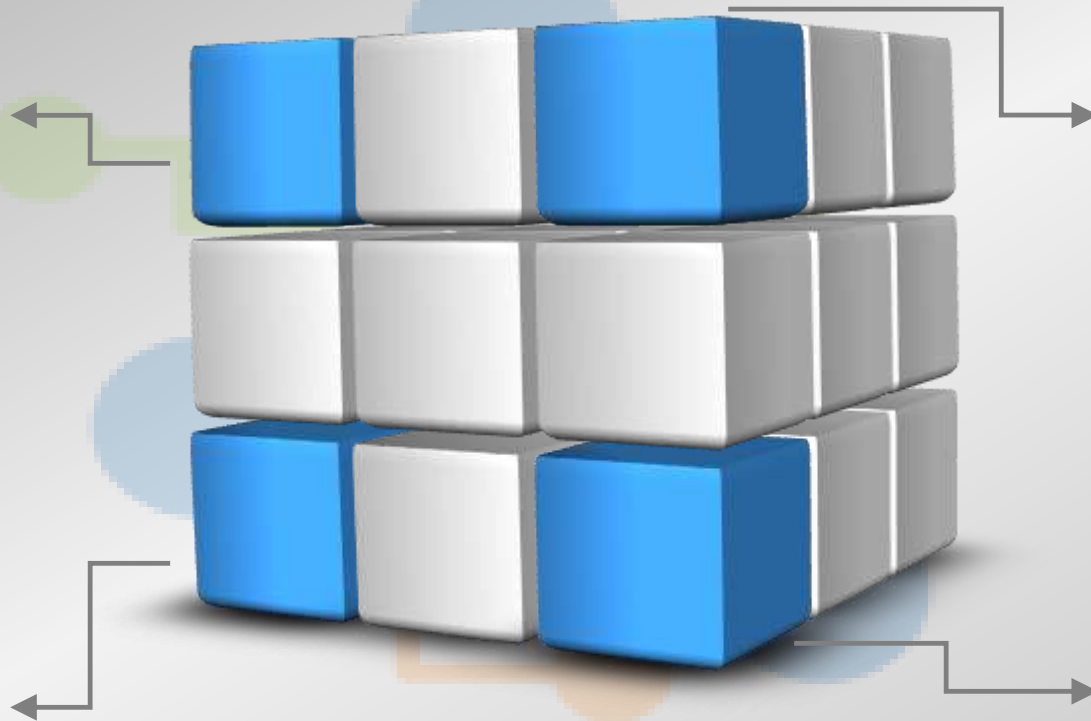
QUANDO USAR UM DATA WAREHOUSE?

Estruturas de Dados
são complexas.

Informação
consolidada é
requerida.

Dados estão
dispersos e sua
representação pode
não ser a ideal.

Apenas informações
relevantes são
necessárias.



QUANDO USAR UM DATA WAREHOUSE?

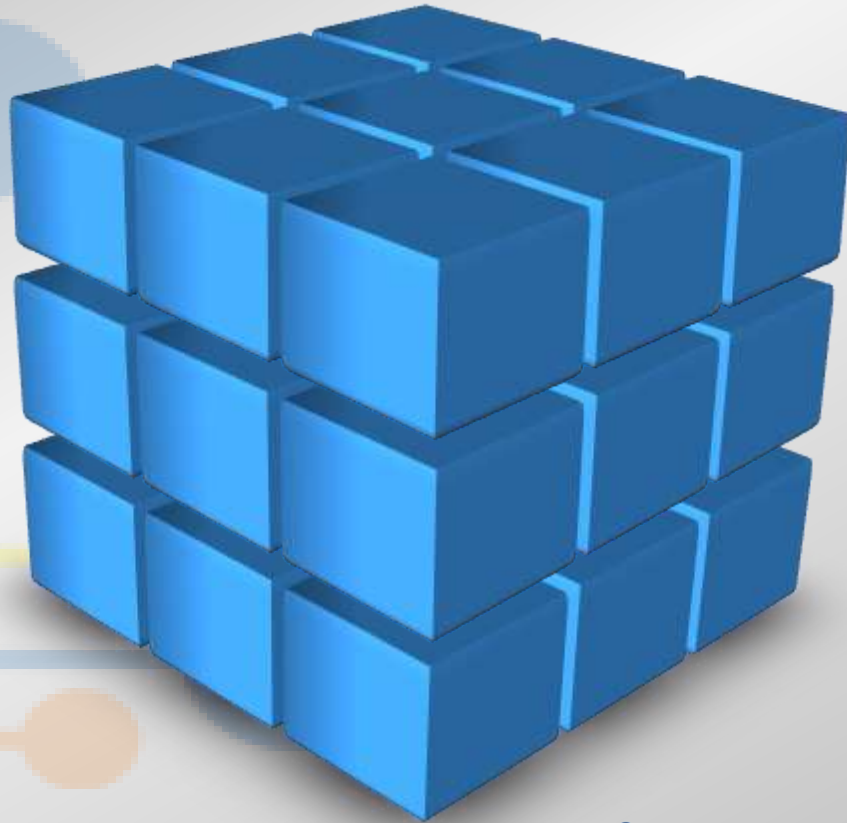
Quando for necessário consolidar os dados mais relevantes para a geração de relatórios de BI e suporte à tomada de decisão.



QUANDO USAR UM DATA WAREHOUSE?

O objetivo principal do DW é suportar o negócio da empresa, reduzindo custos, aumentando lucros e permitindo tomada de decisão com base em dados.

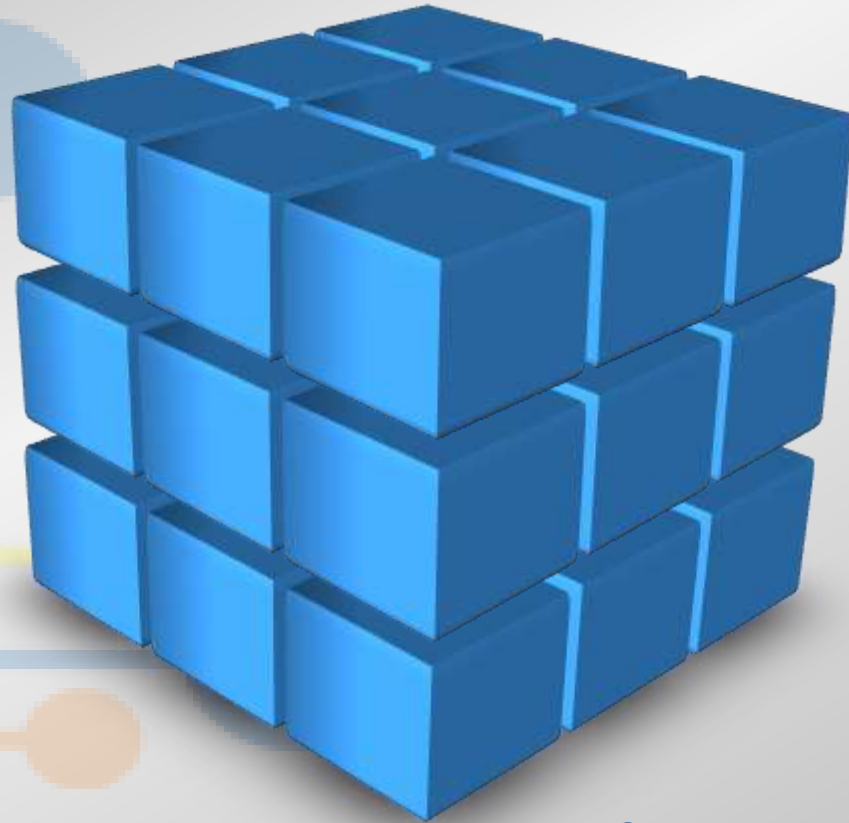
Essencialmente dados estruturados.



QUANDO USAR UM DATA WAREHOUSE?

Data Warehouse

- Hardware
- Sistema Operacional
- Banco de Dados
- Ferramentas de ETL e Consulta
- Técnicas de Indexação
- Ferramentas de Análise
- Ferramentas de Relatórios, Gráficos e Dashboards





OLTP X RELATÓRIOS ANALÍTICOS



OLTP X RELATÓRIOS ANALÍTICOS

OLTP	Relatórios Analíticos
Informação para suportar as decisões diárias da operação de uma empresa	Informação histórica consolidada
Dados são armazenados em nível de transação	Dados são integrados de diferentes fontes e consolidados
Design do banco de dados: normalizado	Design do banco de dados: desnormalizado, Star Schema





VANTAGENS E LIMITAÇÕES DO DW



DATA WAREHOUSE

Data Warehouse é um conceito abstrato que indica que você está consolidando dados de diversas fontes e aplicações de uma organização em um único "local" facilitando suas consultas e eventualmente a manipulação.



DATA WAREHOUSE

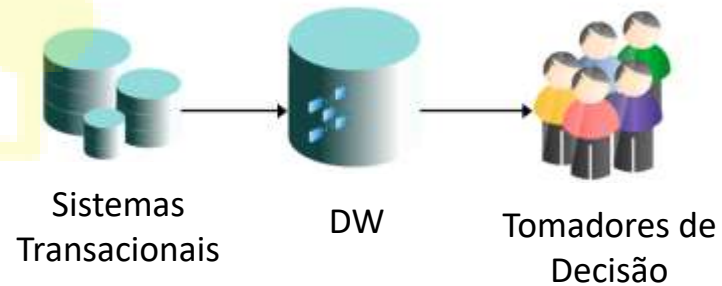
Data Warehouse é um conceito abstrato que indica que você está consolidando dados de diversas fontes e aplicações de uma organização em um único "local" facilitando suas consultas e eventualmente a manipulação.

Normalmente isto é feito através de um banco de dados que depois é alimentado com dados de fontes externas a ele, tornando o DW independente.



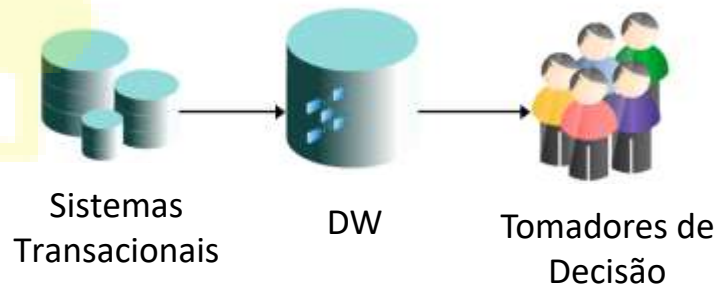
VANTAGENS DO DW

- Informação controlada e confiável.
- Qualidade dos dados.
- Fonte única de informação.
- Evita problemas de conflitos de dados e restrições de Drill-down.
- Dados possuem representação e significado.
- Não há necessidade de várias ferramentas de extração de dados.
- Ele é um grande indexador de dados.
- Costuma ser fácil acessar dados históricos.
- O controle de acesso e permissões aos dados é facilitado.
- Em tese reduz a necessidade de intervenção do pessoal de TI nas esferas Gerenciais.
- Não afeta os sistemas operacionais.



LIMITAÇÕES DO DW

- Não é fácil coletar e transformar os dados de fontes diversas em algo útil para o DW. Costuma-se dizer que o trabalhoso do DW é a transformação que pode reduzir a relação de custo e benefício. É extremamente complicado compatibilizar todos os dados.
- Apesar de controlar o acesso no DW pode haver problemas de segurança e privacidade pelo fato do dado sair da sua fonte que pode ter políticas diferentes.
- A implementação e a manutenção para manter em conformidade com todos os sistemas existentes que estão em constante mutação e novos sistemas não é simples e é caro.





CARACTERÍSTICAS DO DW



CARACTERÍSTICAS DO DW

Orientado ao Negócio

Não Volátil

Variante no Tempo



Data Science Academy

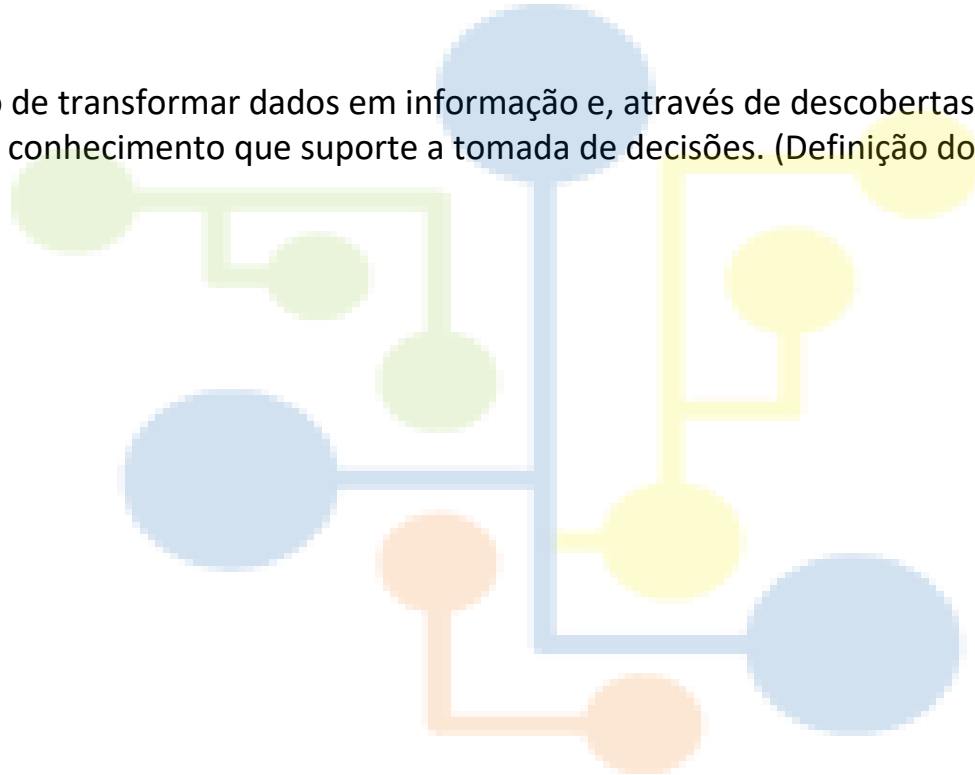


BUSINESS INTELLIGENCE X DATA SCIENCE



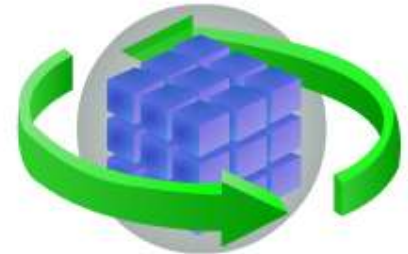
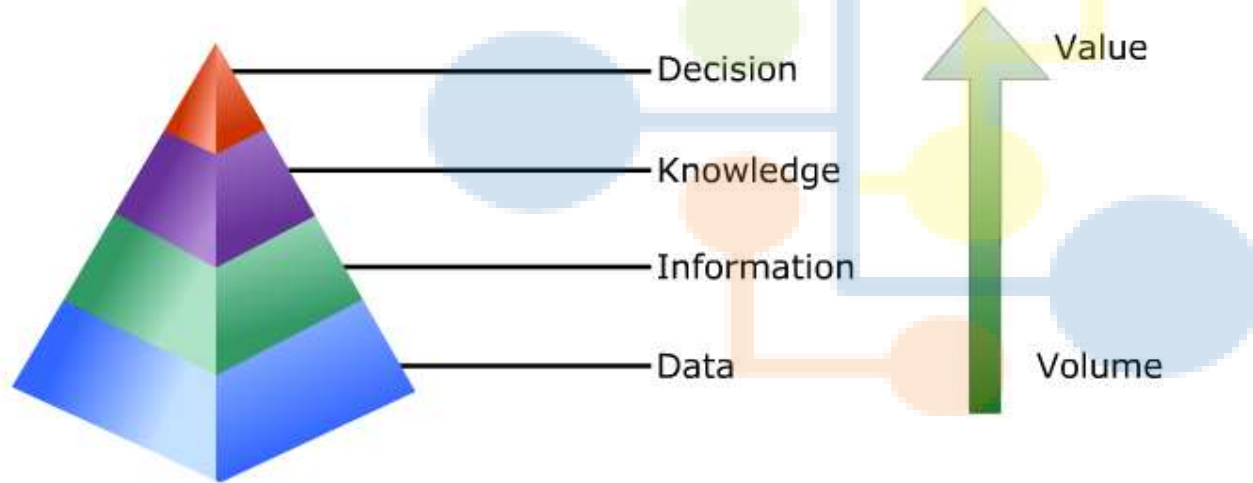
BUSINESS INTELLIGENCE X DATA SCIENCE

É o processo de transformar dados em informação e, através de descobertas, transformar informação em conhecimento que suporte a tomada de decisões. (Definição do Gartner Group)



BUSINESS INTELLIGENCE X DATA SCIENCE

Business Intelligence é o processo de transformar dados em informação e, através de descobertas, transformar informação em conhecimento que suporte a tomada de decisões. (Definição do Gartner Group)



BUSINESS INTELLIGENCE X DATA SCIENCE

BI ajuda a responder as seguintes perguntas:

- Como estão as vendas até este período do ano e como elas se comparam ao mesmo período do ano passado?
- Como está o turnover dos funcionários em comparação ao ano passado?
- Como o custo com fraudes tem afetado o faturamento durante os últimos anos?
- Quais são os produtos mais lucrativos?



BUSINESS INTELLIGENCE X DATA SCIENCE





OLAP E O MODELO DIMENSIONAL



OLAP E O MODELO DIMENSIONAL

O mais comum é usar ferramentas de OLAP (Online Analytical Processing) para trabalhar com DWs, em oposição ao OLTP (Online Transaction Processing) que é usado em bancos de dados ditos operacionais da organização.

Em geral, no DW os dados são usados de forma mais analítica, para suportar a tomada de decisão.



OLAP E O MODELO DIMENSIONAL

Modelagem OLTP x Modelagem OLAP



OLAP E O MODELO DIMENSIONAL

O OLAP fornece para organizações um método de acessar, visualizar, e analisar os dados corporativos com alta flexibilidade e performance.

Portanto, o OLAP é frequentemente utilizado para integrar e disponibilizar informações gerenciais contidas em bases de dados operacionais como sistemas ERP, CRM e sistemas contábeis. Estas características tornaram-no uma tecnologia essencial em diversos tipos de aplicações de suporte à decisão.



OLAP E O MODELO DIMENSIONAL

A arquitetura OLAP possui ferramentas que são classificadas em cinco tipos: ROLAP, MOLAP, HOLAP, DOLAP e WOLAP (além de XOLAP).



OLAP E O MODELO DIMENSIONAL

A aplicação do OLAP é bastante diversificada e seu uso encontra-se em diversas áreas de uma empresa. Alguns tipos de aplicação onde a tecnologia é empregada são:

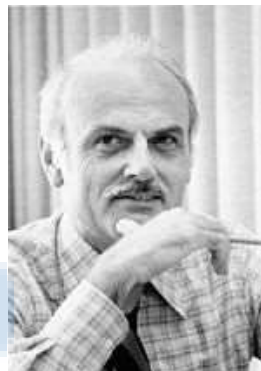
Finanças	Análise de L&P, Relatórios L&P, Orçamento, Análise de Balanço, Fluxo de Caixa, Contas a Receber,
Vendas	Análise de vendas (por região, produto, vendedor, etc.), Previsões, Lucratividade de Cliente/Contrato, Análise de Canais de Distribuição, ...
Marketing	Análise de Preço/Volume, Lucratividade de Produto, Análise de Mercados, ...
Recursos Humanos	Análise de Benefícios, Projeção de Salários, Análise de "Headcount", ...
Manufatura	Gerência de Estoque, Cadeia de Fornecimento, Planejamento de Demanda, Análise de custos de matéria-prima, ...



OLAP E O MODELO DIMENSIONAL

Descrito por E. F. Codd, em 1992, o termo OLAP possui 12 (doze) regras utilizadas para avaliar ferramentas ou produtos OLAP. Essas doze regras são:

1. Conceito de visão multidimensional;
2. Transparência;
3. Acessibilidade;
4. Performance consistente de relatório;
5. Arquitetura cliente/servidor;
6. Dimensionamento genérico;
7. Tratamento dinâmico de matrizes esparsas;
8. Suporte a multiusuários;
9. Operações de cruzamento dimensional irrestritas;
10. Manipulação de dados intuitiva;
11. Relatórios flexíveis;
12. Níveis de dimensões e agregações ilimitados.



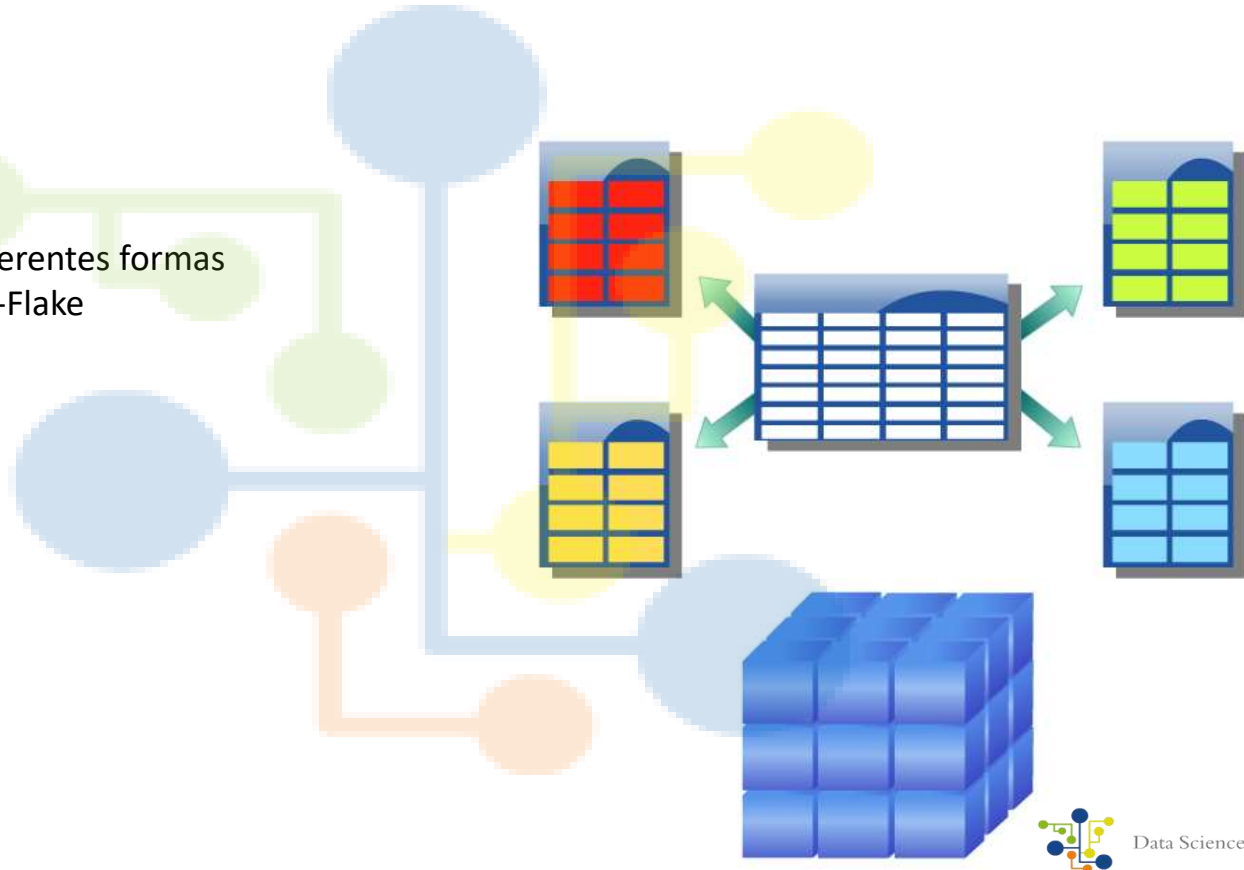
E. F. Codd



OLAP E O MODELO DIMENSIONAL

Modelo Dimensional

- Pode ser criado de diferentes formas
- Star Schema ou Snow-Flake
- Fatos
- Dimensões
 - Hierarquias
 - Níveis
 - Atributos



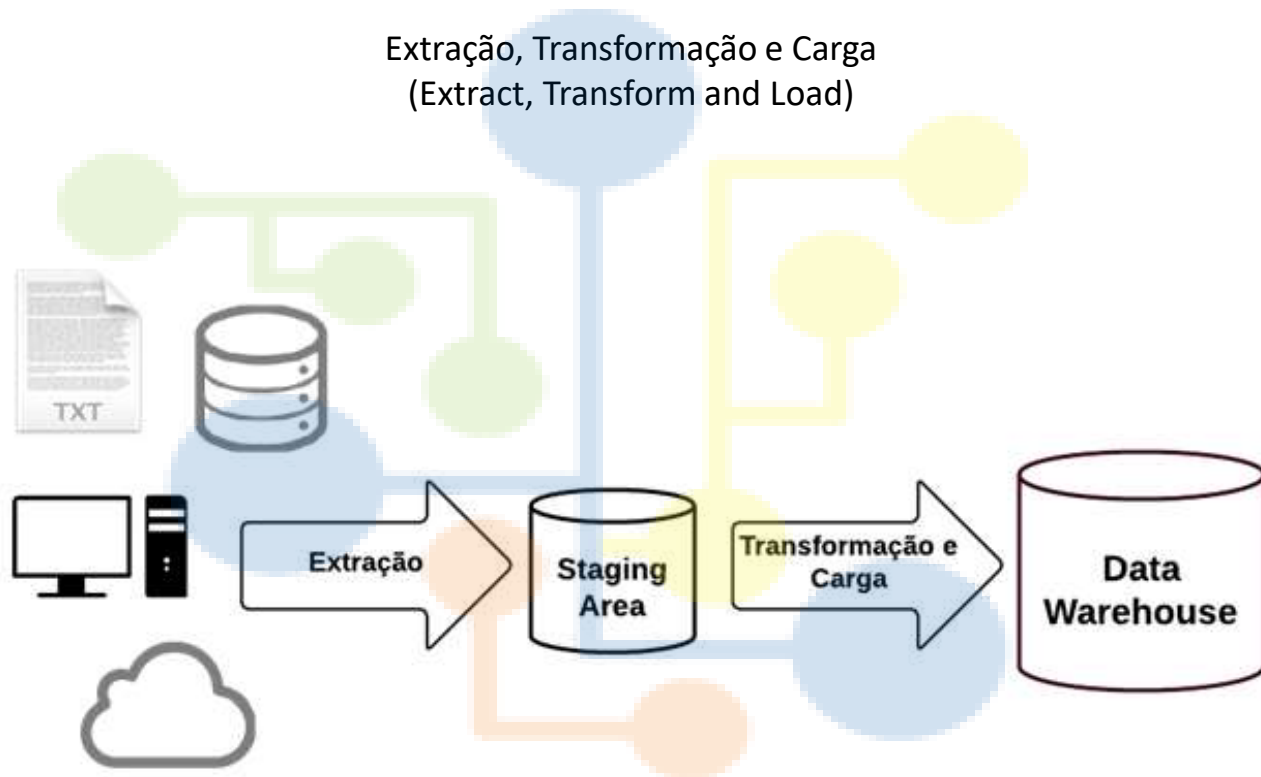


ETL



ETL

Extração, Transformação e Carga
(Extract, Transform and Load)



ETL

Extração, Transformação e Carga (Extract, Transform and Load)

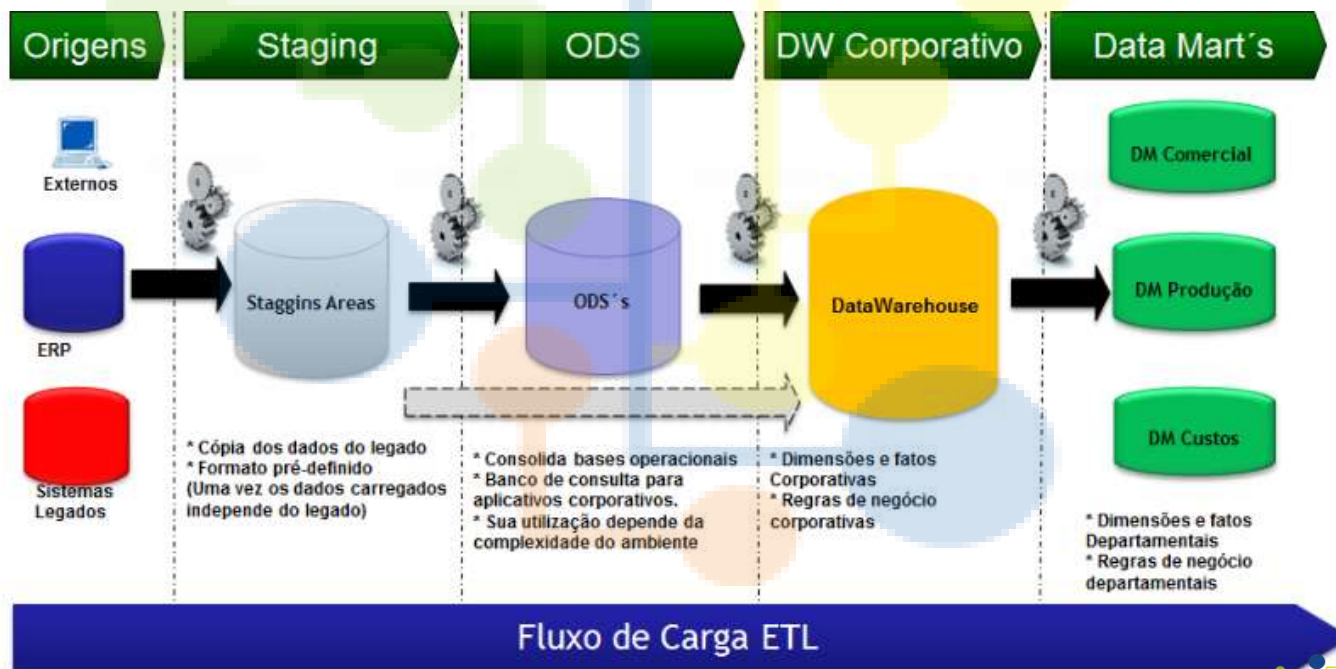
EXTRAÇÃO: fase em que os dados são extraídos dos OLTPs e conduzidos para a staging area (área de transição ou área temporária), onde são convertidos para um único formato.

TRANSFORMAÇÃO: É nesta etapa que realizamos os devidos ajustes, podendo assim melhorar a qualidade dos dados e consolidar dados de duas ou mais fontes.

CARGA: Consiste em fisicamente estruturar e carregar os dados para dentro da camada de apresentação seguindo o modelo dimensional.



Extração, Transformação e Carga (Extract, Transform and Load)





Muito Obrigado!

É um prazer ter você aqui.
Tenha uma excelente jornada de aprendizagem.



Oportunidade



Disponibilidade



Conhecimento