



Osvaldo Aranha

Head of Data & AI
@Microsoft

MBA em Business Intelligence pela FIAP , pós-graduação em Tecnologia pela USP-Poli, e extensão em gestão de TI pela FGV-EAESP.

Formado em Processamento de Dados pela FATEC e possui diversas certificações de tecnologia.

Especialista em Analytics e BI, é Head da área de Azure Data & AI na Microsoft

Ex-Diretor da área de Analytics na Deloitte Touche Tohmatsu, mentor de startups e professor de pós-graduação.

SOBRE O MÓDULO

01

INTRODUÇÃO

02

PERFIL DE UM
ANALISTA DE DADOS

03

TIPOS DE ARQUIVOS

04

TIPOS DE GRÁFICOS

05

TIPOS DE ANÁLISES

06

OVERVIEW DE
PROJETOS +
FECHAMENTO

INTRODUÇÃO

INTRODUÇÃO



Conceitos fundamentais e práticos sobre a análise de dados, os tipos de dados tipicamente encontrados nas empresas,



Metodologias dos processos de coleta, tratamento e análise de dados estruturados/não estruturados.



Serão apresentadas as abordagens, melhores práticas e as ferramentas mais utilizadas atualmente, com preparação para futuro caso prático.

4 BIG DATA MACRO TRENDS YOU SHOULD DEFINITELY WATCH

LATEST NEWS MARKET TRENDS UNCATEGORIZED

by Princy Lalawat / July

TECNOLOGIA

Sales big d

JULY 7, 2015 BY TIMO ELLIOTT

Could Data Science Have Saved Greece?

crisis, many
avoided if

DESCUBRA A

Era do cidadão Cientista de Dados chegou e o que isso mudará para você?

Ferramentas de software estão capacitando analistas de negócios a gerar modelos analíticos e insights sem a ajuda de engenheiros de software

Home > Noti

Clint Boulton, CIO (EUA)
30/10/2019 às 8h24

Clou recup

Segundo
investir e

Da Redação

10/11/2020 às 18h31

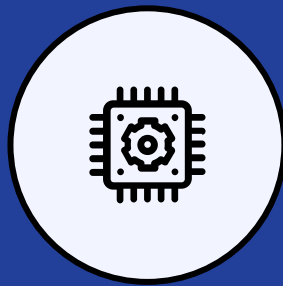


ial criteria for
eek deficit
he Greek
atistics to the EU
t country's
% limit.

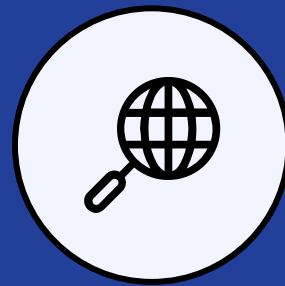
resas devem

Digital

EM ATÉ 3 ANOS



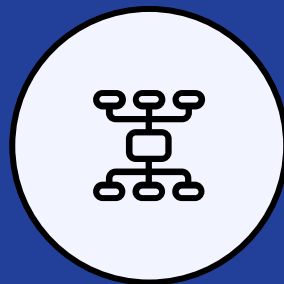
Serão criados cerca de 2 megabytes de novas informações por segundo, para cada ser humano do planeta.



Haverá mais de 50 bilhões de objetos conectados no mundo, todos desenvolvidos para coletar, analisar e compartilhar dados.



Mais de 55 zettabytes (ou 55 trilhões de GB) de dados estarão disponíveis no mundo. **Desse total, 35% serão úteis para análise (IDC)**



Pelo menos dois terços de todos os dados serão enviados por meio da nuvem.



Existirão mais de 6 bilhões de usuários de smartphones no mundo.

CENÁRIO



Cybersecurity

A cibersegurança se tornou relevante nas organizações. Pense no seguinte: se uma empresa tem 10 colaboradores e todos eles estão trabalhando remotamente, cada um se torna um alvo em potencial para ataques.



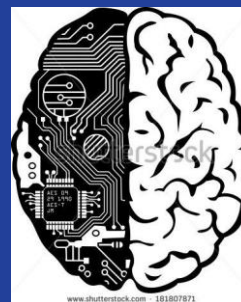
IoT

Para o próximo ano, a expectativa é que 35 bilhões de dispositivos IoT estejam funcionando em todo o mundo.



5G

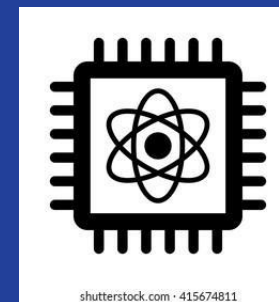
O uso de aplicações em nuvem para essa colaboração digital exigiu uma atenção redobrada na conectividade. De uma rede segura a maior largura de banda, o 5G será capaz de manter as operações on-line, quase sem interrupções.



Democratização da IA / GPT-3

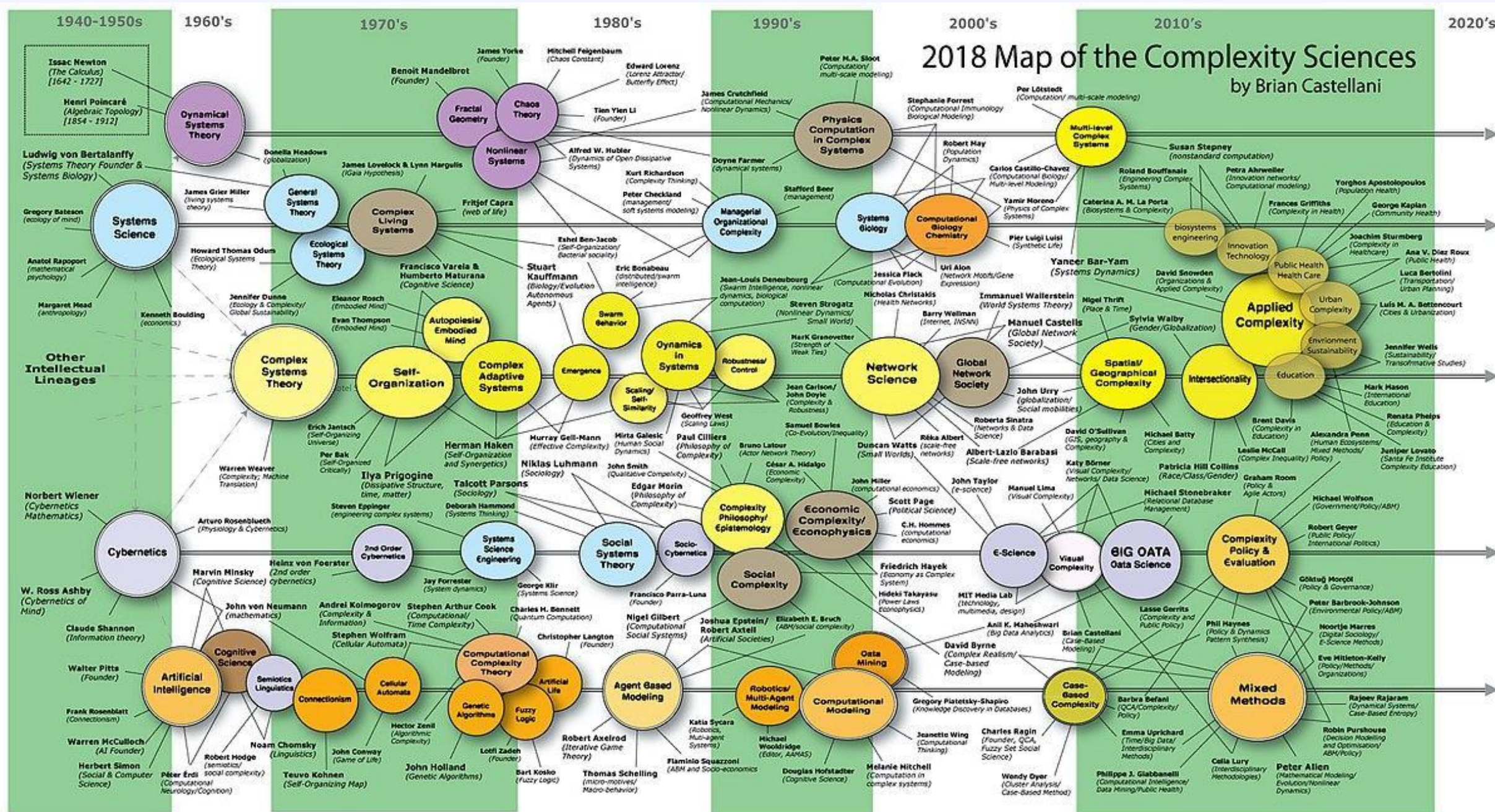
Diversas organizações começaram a dar atenção aos seus dados e essas tecnologias, até então emergentes, ajudaram a organizá-los para que uma tomada de decisão assertiva fosse feita.

Inteligência Artificial ficará cada vez mais democrática.



Computação Quântica

São poucos os casos de uso da tecnologia, mas, logo, todos os setores poderão aproveitar a computação quântica para consultar, monitorar, analisar e agir com facilidade e em escala sobre os dados que eles possuem.





Exploration

Refinement



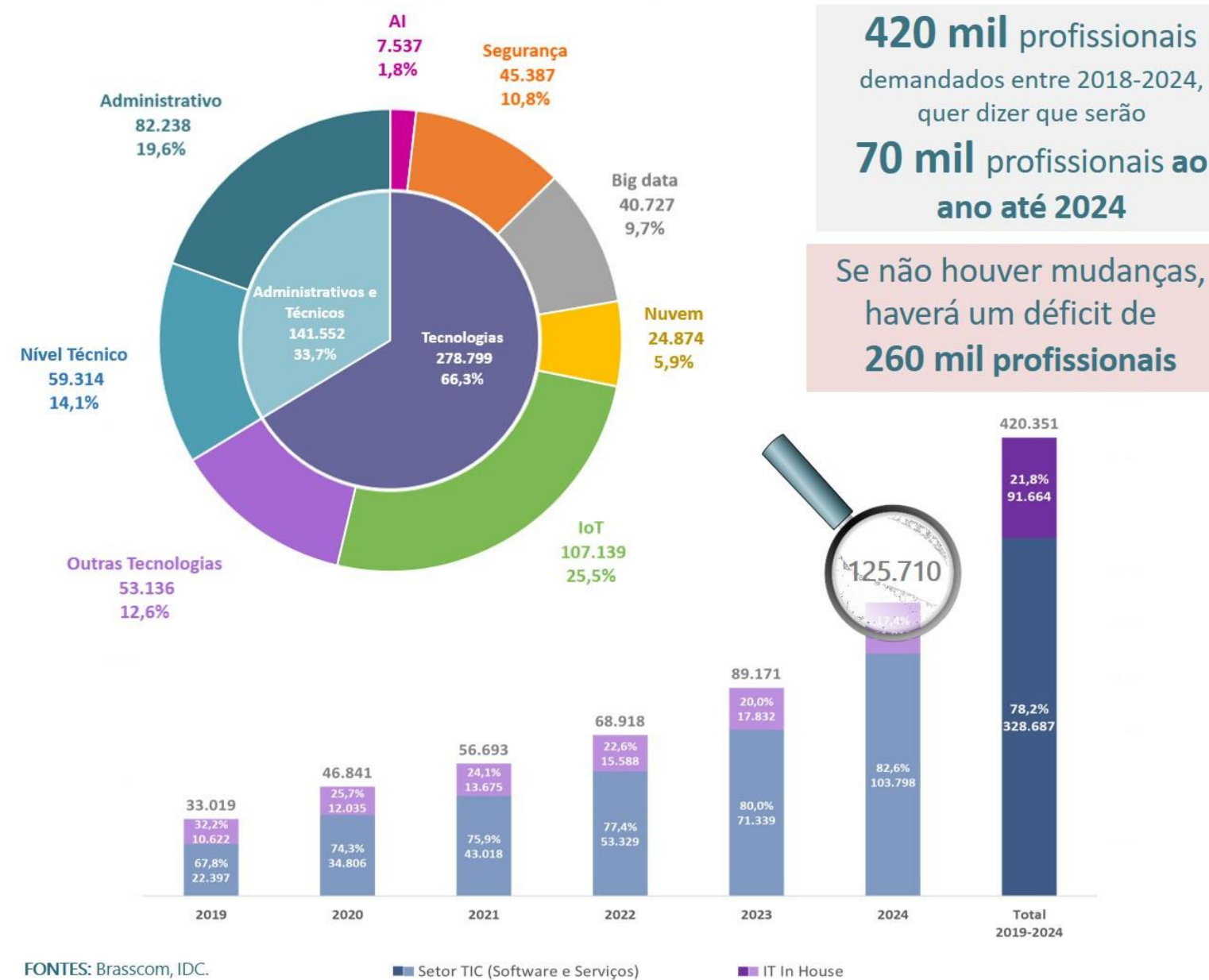
Extraction

Application

PERFIL DE UM ANALISTA DE DADOS

PERFIL DE UM ANALISTA DE DADOS

Demanda de Empregos por Tecnologias e TI In House (2019-2024)



TECNOLOGIAS DE TRANSFORMAÇÃO DIGITAL



Nuvem
R\$ 77,8 bi

26% a.a.



Robótica
R\$ 23,0 bi

17% a.a.



IoT
R\$ 155,2 bi

19% a.a.



Social
R\$ 9,3 bi

14% a.a.



Big Data & Analytics
R\$ 61,1 bi

10% a.a.



Realidade virtual/aumentada
R\$ 6,0 bi

52% a.a.



Segurança da Informação
R\$ 8,9 bi

15% a.a.



Impressão 3D
R\$ 0,2 bi

22% a.a.



Inteligência Artificial
R\$ 2,5 bi

29% a.a.



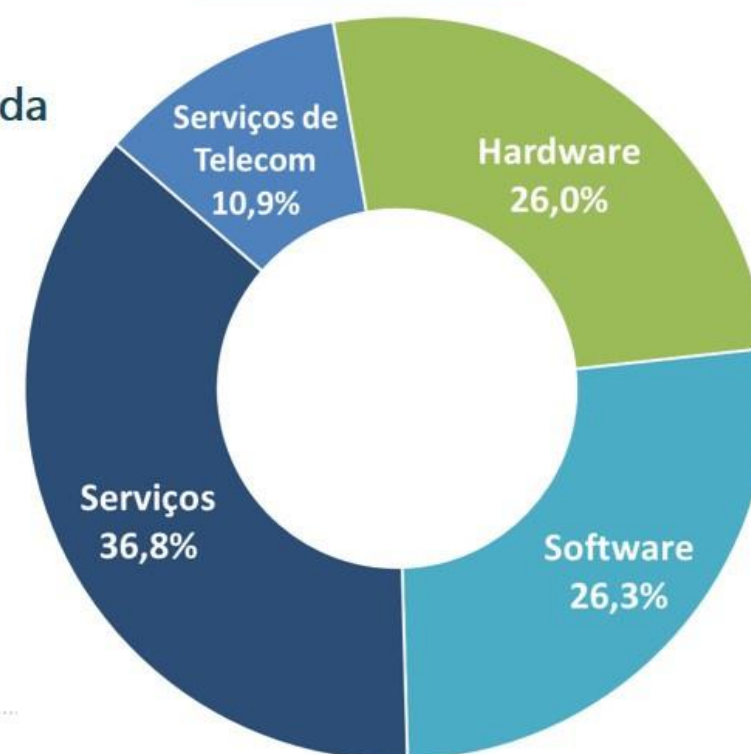
Blockchain
R\$ 1,4 bi

64% a.a.

R\$ 345,5 bi

Transformação Digital

19,3% a.a.



FONTES: Brasscom, IDC.

PERFIL DO PROFISSIONAL DE DADOS



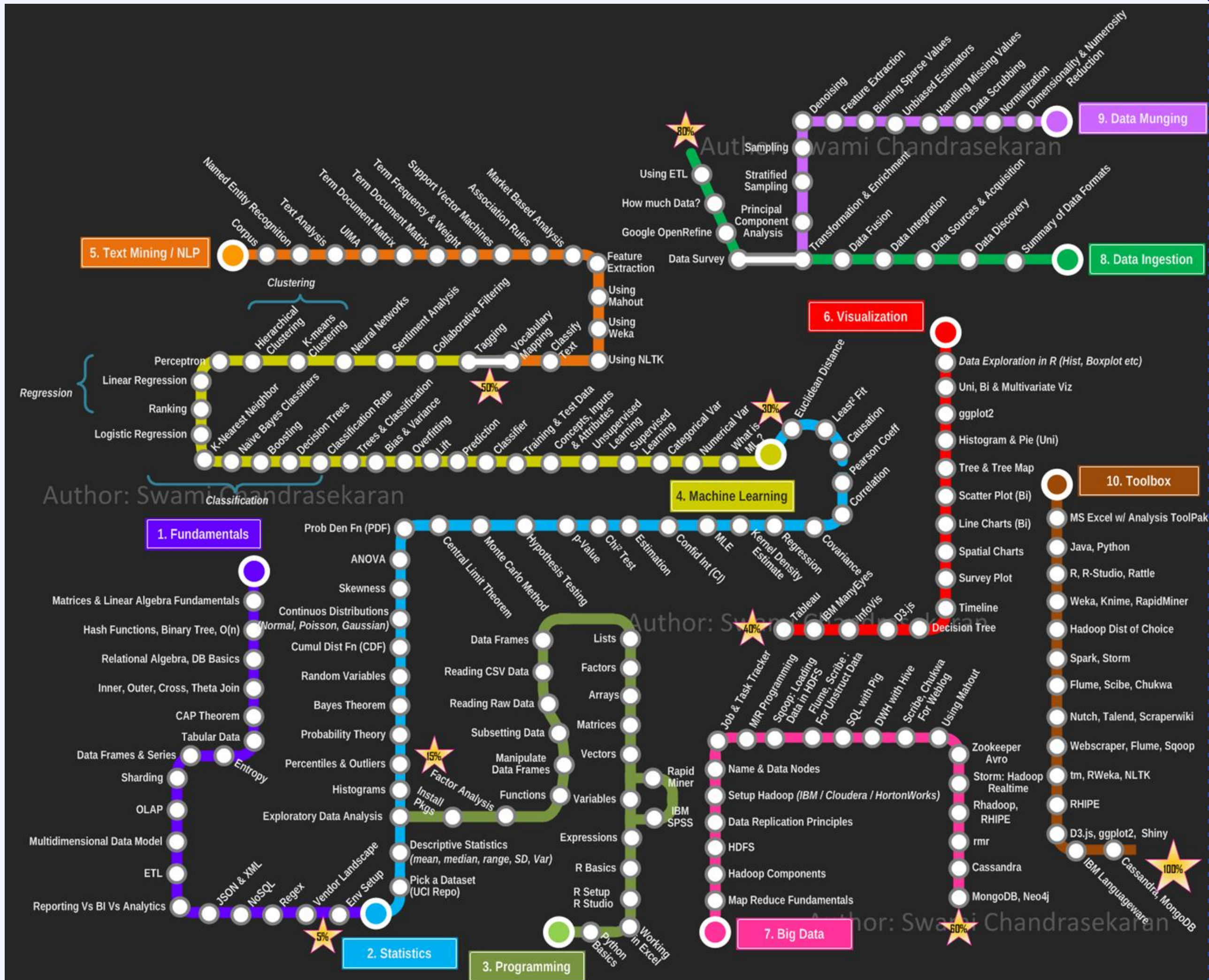
O QUE ESPERAR DO PROFISSIONAL

- Generalista / Multi-Task
- Empreendedorismo
- Life Long Learner
- Problem Solver
- Curiosidade/ Criatividade
- Conhecimento do negócio
- Boa comunicação

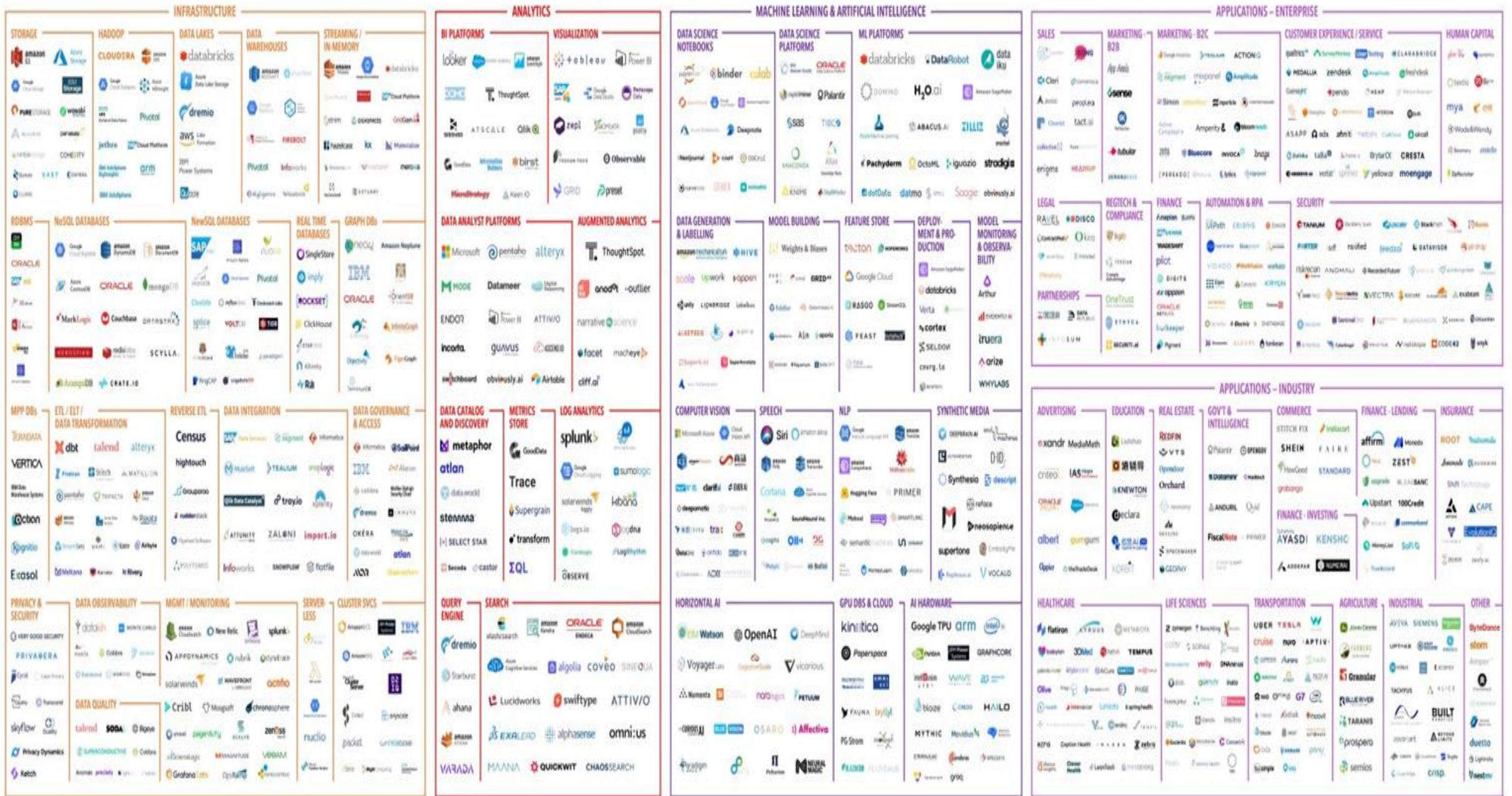


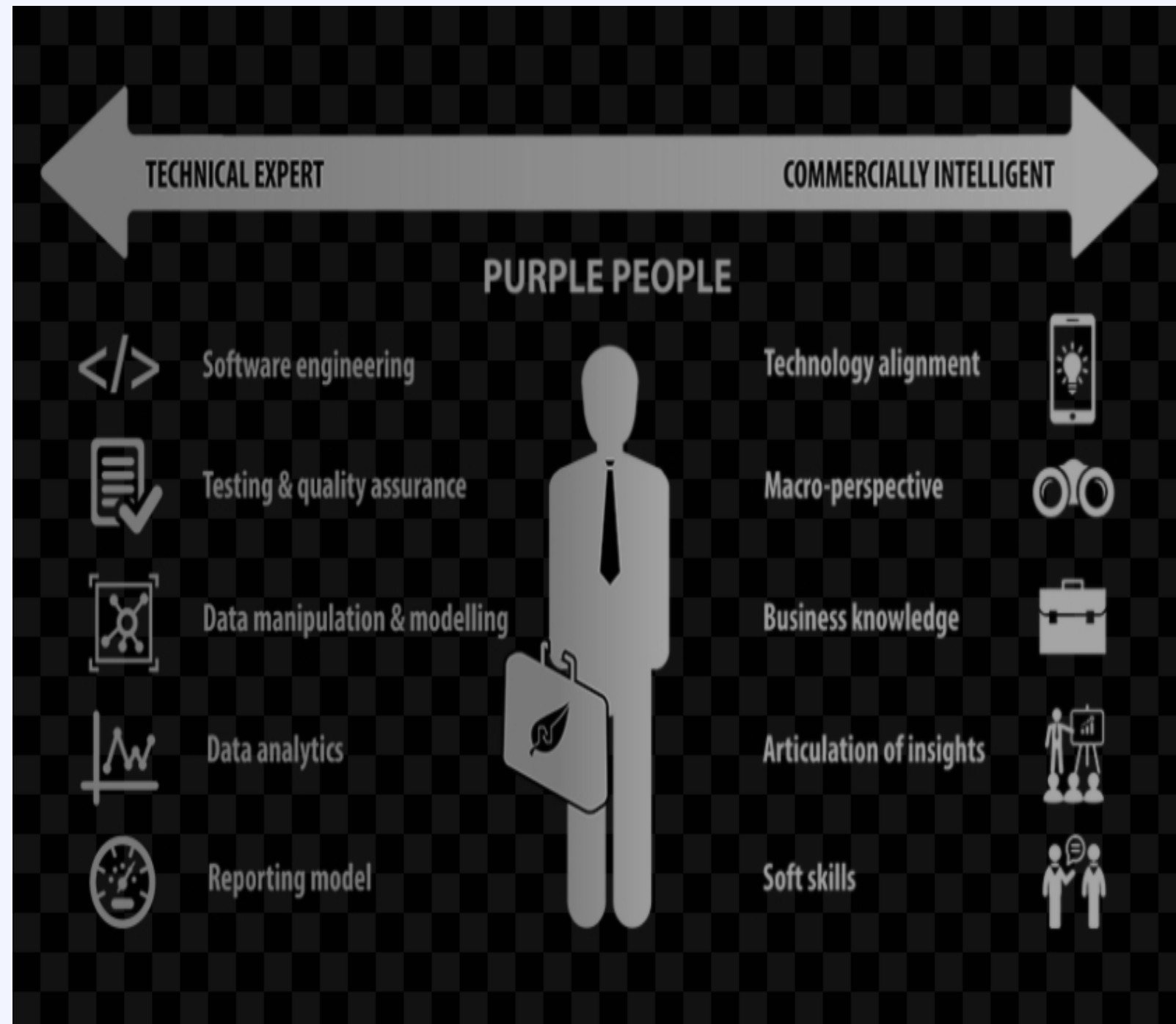
O QUE ESPERAR DA EMPRESA

- Oportunidades (diferente de plano de carreira)
- Cultura bem estabelecida
- Responsabilidade ESG
- Remuneração Justa
- Justiça/Equidade



MACHINE LEARNING, ARTIFICIAL INTELLIGENCE, AND DATA (MAD) LANDSCAPE 2021





MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

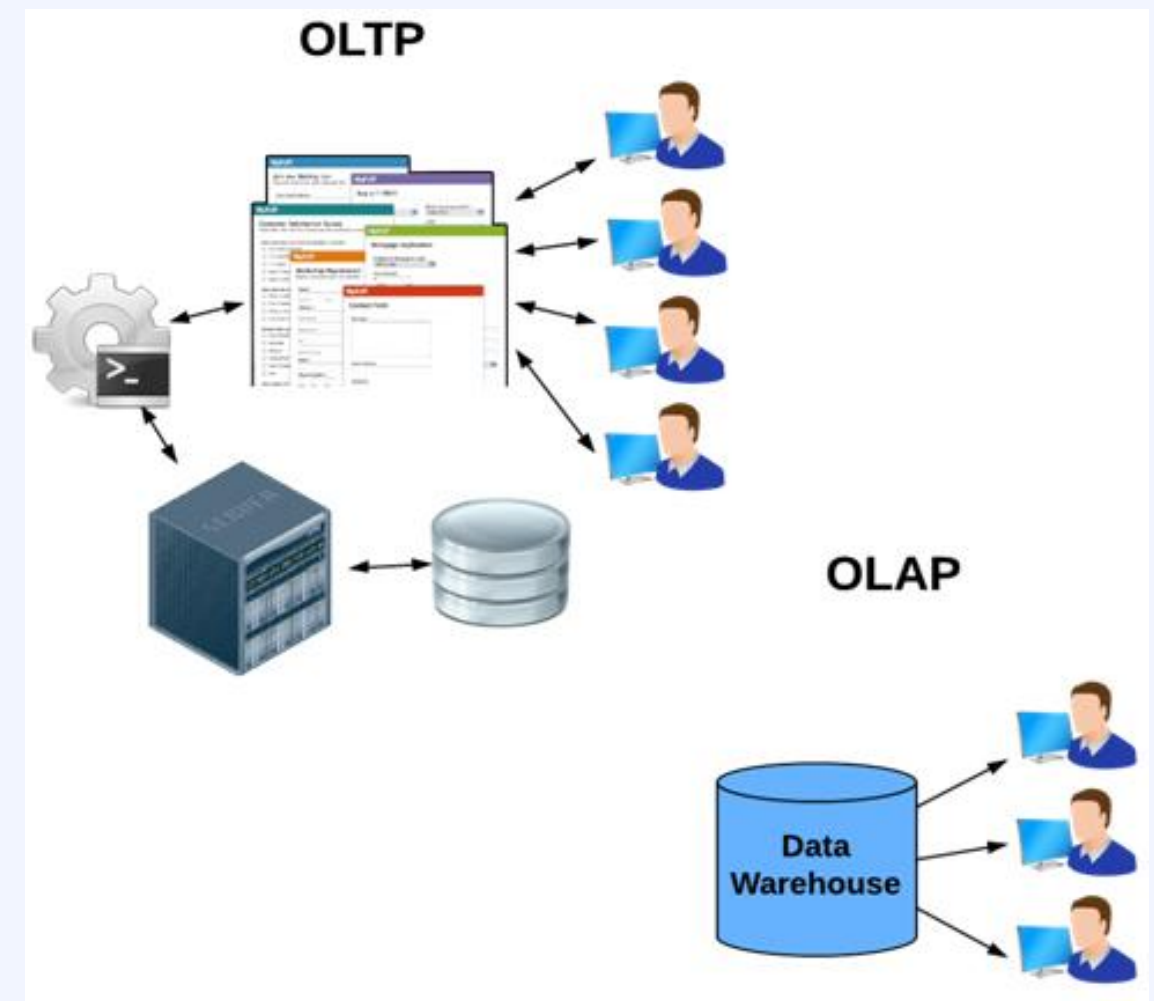
- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

TIPOS DE ARQUIVOS E ARQUITETURAS

CONCEITOS FUNDAMENTAIS



	OLAP	OLTP
Foco	Foco no nível estratégico da organização. Visa a análise empresarial e tomada de decisão.	Foco no nível operacional da organização. Visa a execução operacional do negócio.
Performance	Otimização para a leitura e geração de análises e relatórios gerenciais.	Alta velocidade na manipulação de dados operacionais, porém ineficiente para geração de análises gerenciais.
Estrutura dos dados	Os dados estão estruturados na modelagem dimensional. Os dados normalmente possuem alto nível de sumarização.	Os dados são normalmente estruturados em um modelo relacional normalizado, otimizado para a utilização transacional. Os dados possuem alto nível de detalhes.
Armazenamento	O armazenamento é feito em estruturas de <i>Data Warehouse</i> com otimização no desempenho em grandes volumes de dados.	O armazenamento é feito em sistemas convencionais de banco de dados através dos sistemas de informações da organização.
Abrangência	É utilizado pelos gestores e analistas para a tomada de decisão.	É utilizado por técnicos e analistas e engloba vários usuários da organização.
Frequência de atualização	A atualização das informações é feita no processo de carga dos dados. Frequência baixa, podendo ser diária, semanal, mensal ou anual (ou critério específico).	A atualização dos dados é feita no momento da transação. Frequência muito alta de atualizações.
Volatilidade	Dados históricos e não voláteis. Os dados não sofrem alterações, salvo necessidades específicas (por motivos de erros ou inconsistências de informações).	Dados voláteis, passíveis de modificação e exclusão.
Tipos de permissões nos dados	É permitido apenas a inserção e leitura. Sendo que para o usuário está apenas disponível a leitura.	Podem ser feito leitura, inserção, modificação e exclusão dos dados.

NoSQL



Gaming



Social



IoT



Web



Mobile



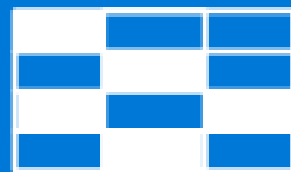
Enterprise



Key/value store



Document
database



Column family store

SQL



Web



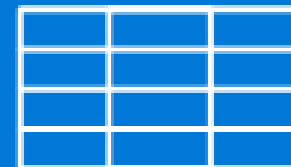
Mobile



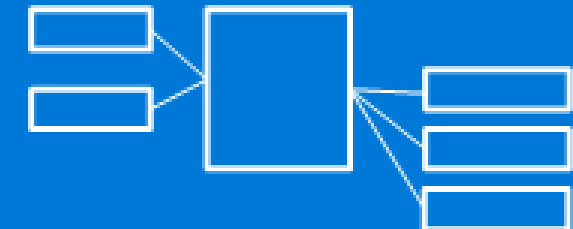
Enterprise



Data mart

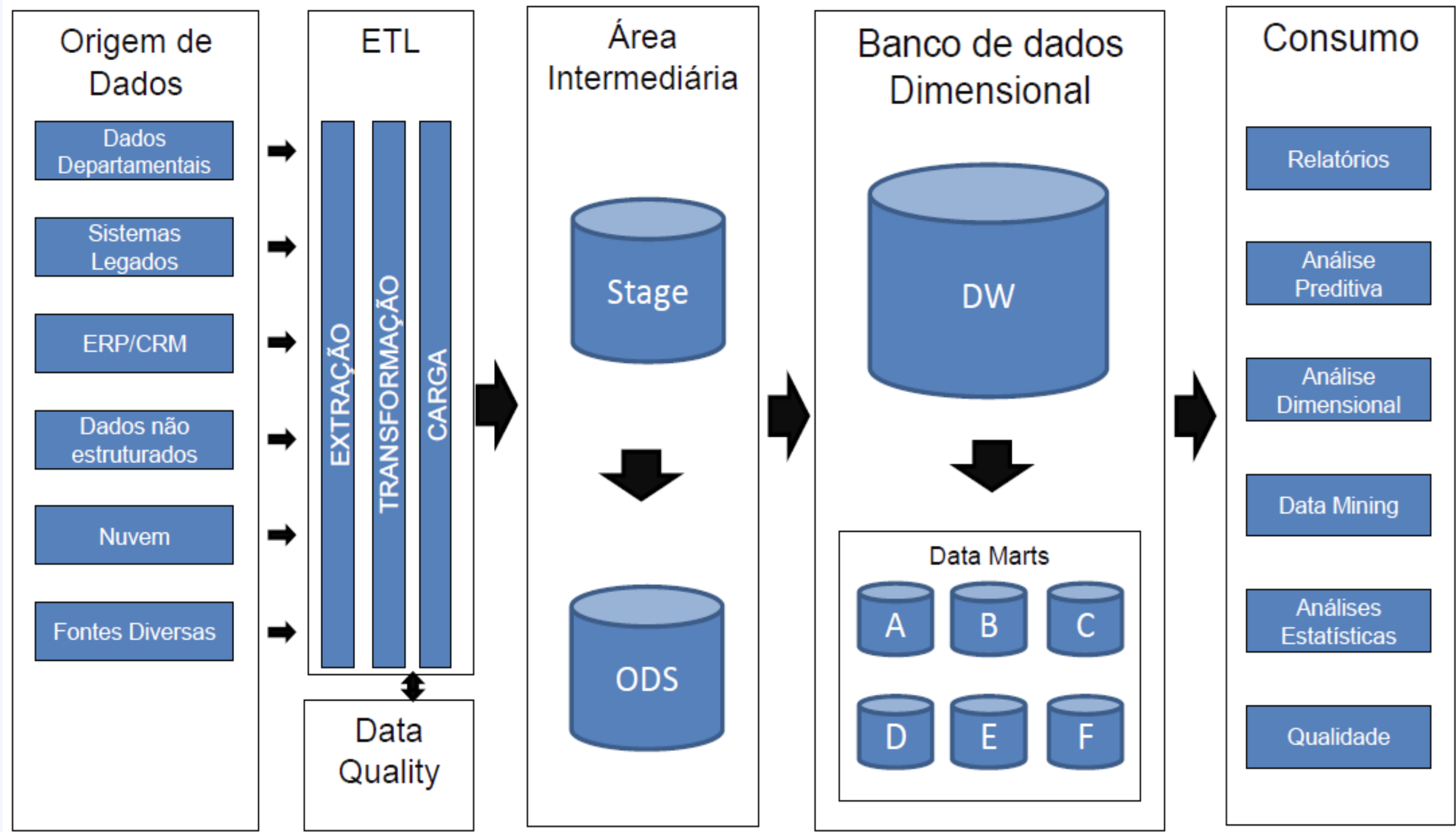


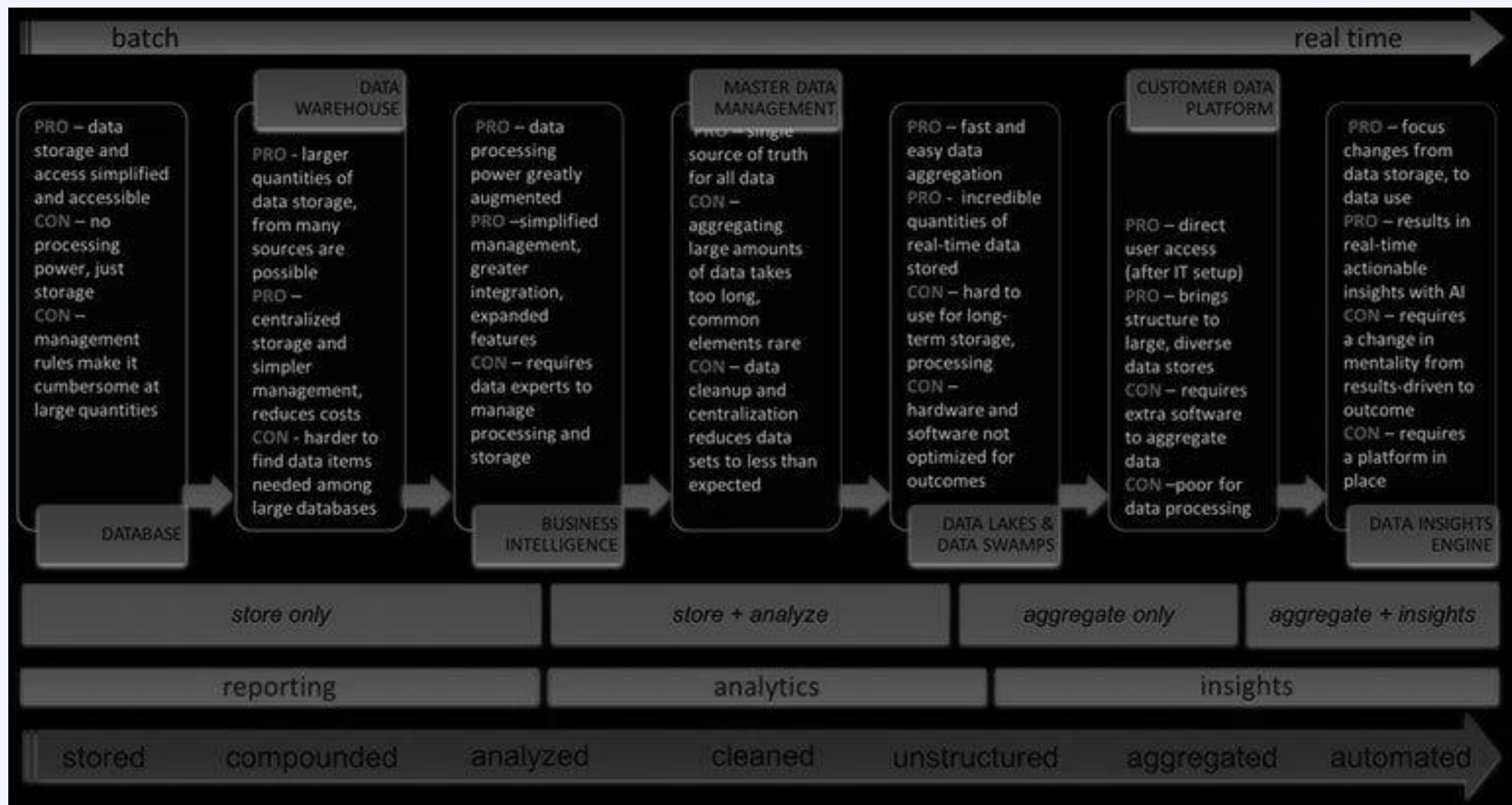
Relational table storage

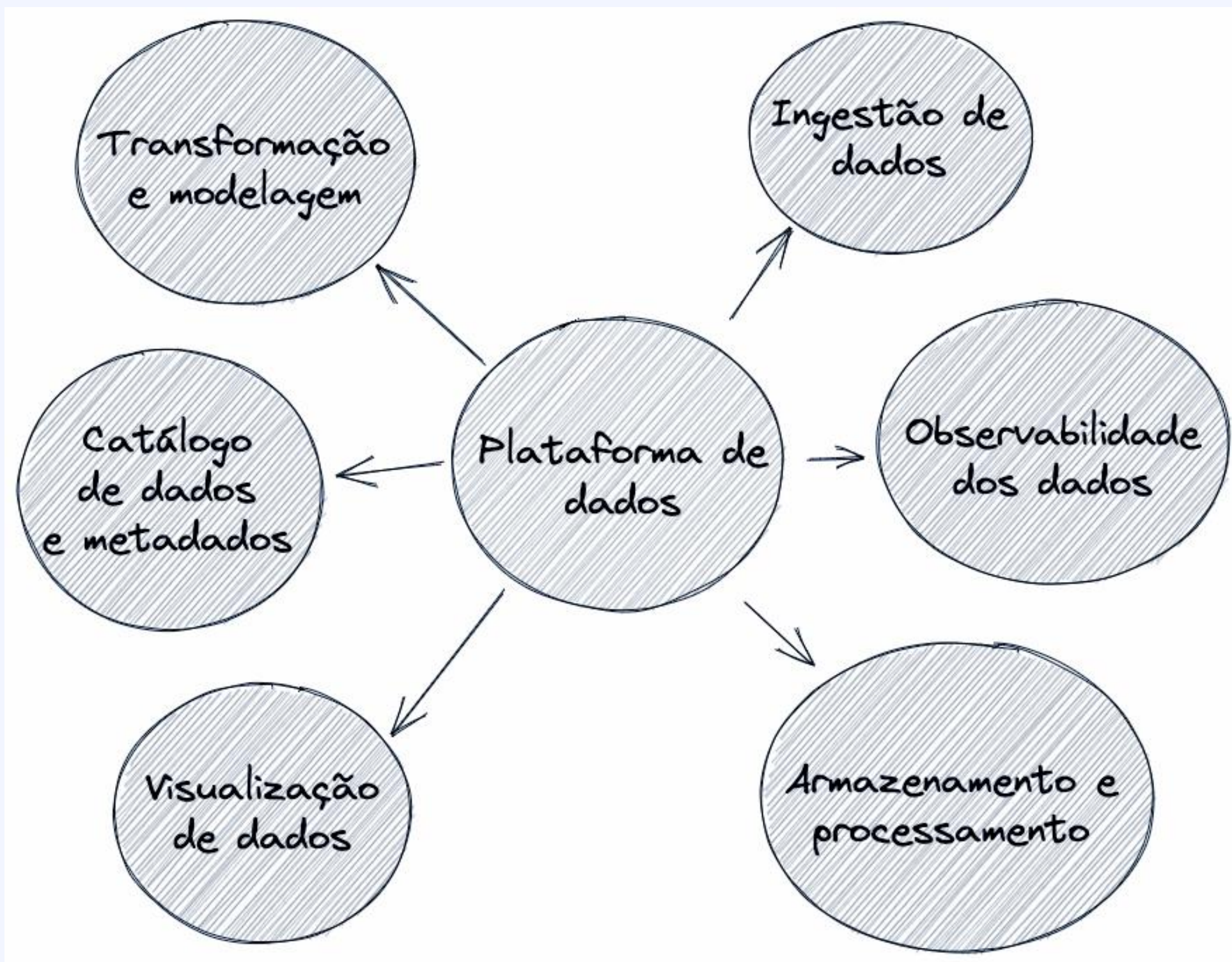


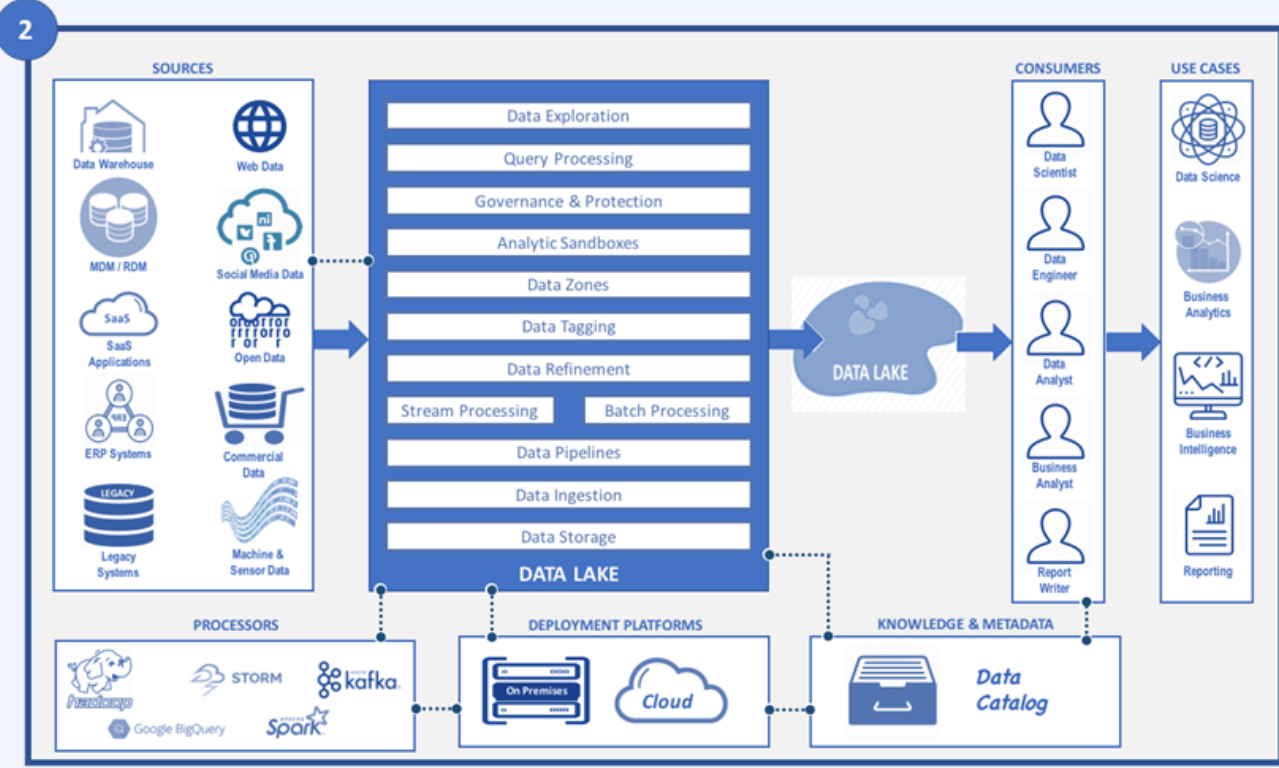
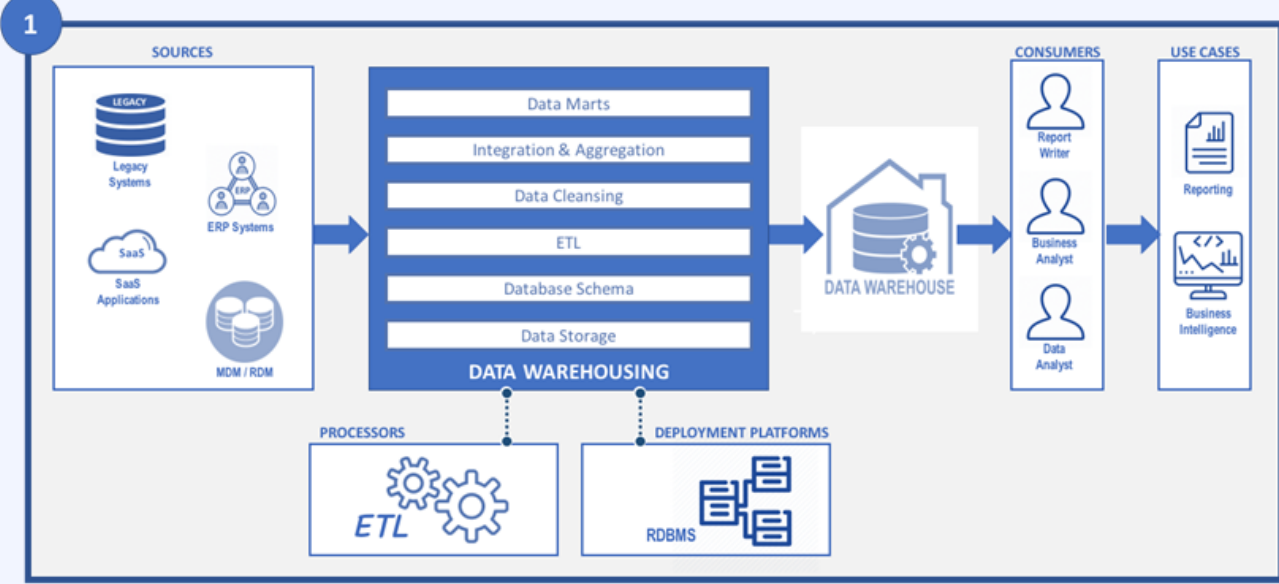
Relationships use joins

Critério de Classificação	Questões de Levantamento
Disponibilidade	Quão disponível e acessível está o dado? Há obstáculos técnicos ao acesso? Há problemas de autorização de acesso?
Entendimento	Quão simples é o entendimento dos dados? Ele é bem documentado? Alguém na organização tem conhecimento em profundidade? Quem trabalha frequentemente com esses dados?
Estabilidade	Com que frequência a estrutura dos dados muda? Qual o histórico de mudança dos dados? Qual o tempo de vida das origens potenciais?
Exatidão	Quão confiável é o dado? As pessoas que trabalham com o dado confiam nele?
Prontidão	Quando e com qual frequência os dados são atualizados? Quão atual é o dado? Quanto de histórico está disponível? Quanto ele está disponível para extração?
Completude	O escopo dos dados corresponde ao escopo do Data Warehouse? Há dados faltando?
Granularidade	A fonte tem o dado mais granular disponível para o dado a ser analisado?

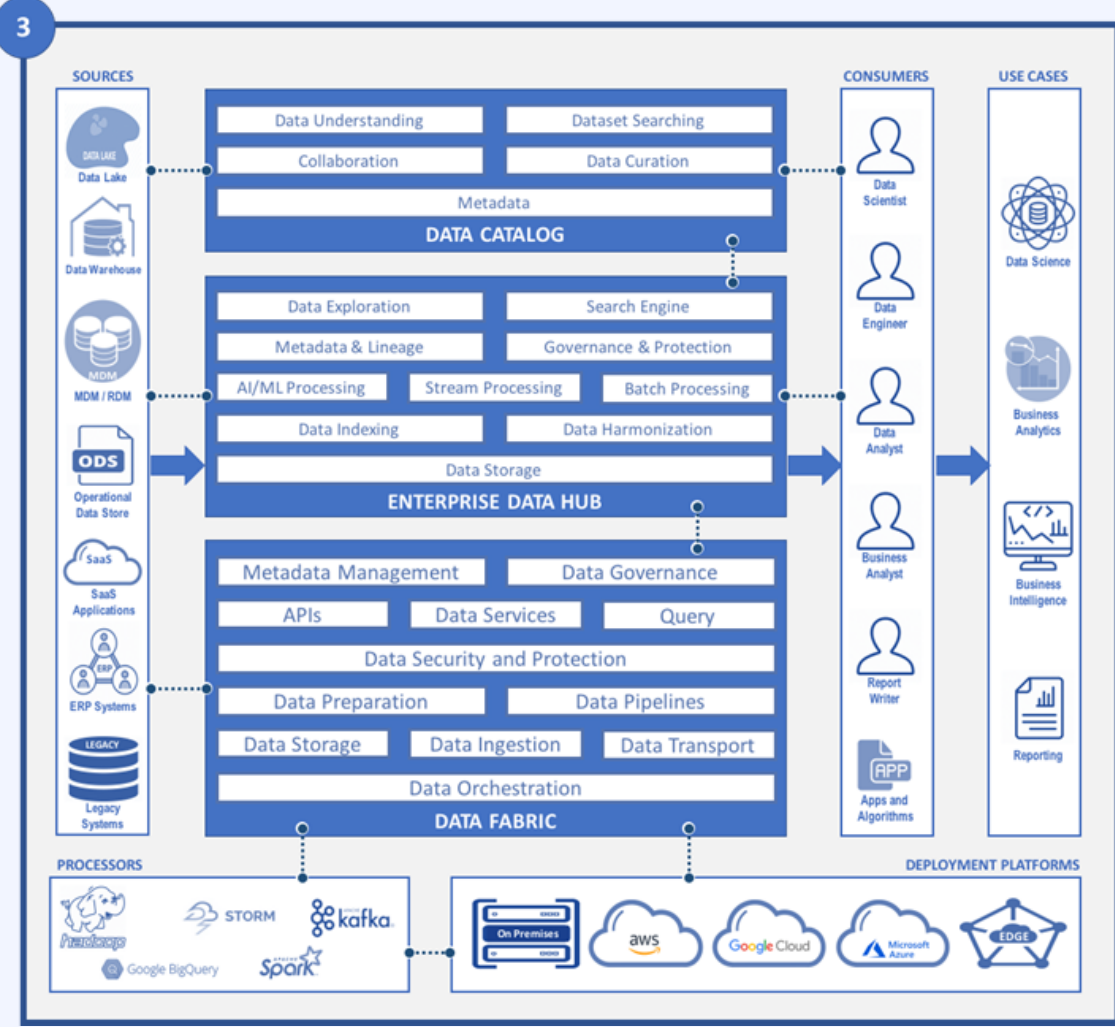
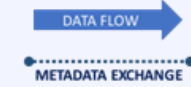








- 1st Generation Data Management – Data Warehousing**
Focus is on data integration and single version of truth using a one-size-fits-all data model.
- 2nd Generation Data Management – Data Lake**
Data warehouses become sources for the data lake.
- 3rd Generation Data Management – Enterprise Data Hub + Data Fabric**
Data lake & warehouses become sources for the enterprise data hub. Data catalog becomes a centerpiece of data management.



Desafios de Arquitetura



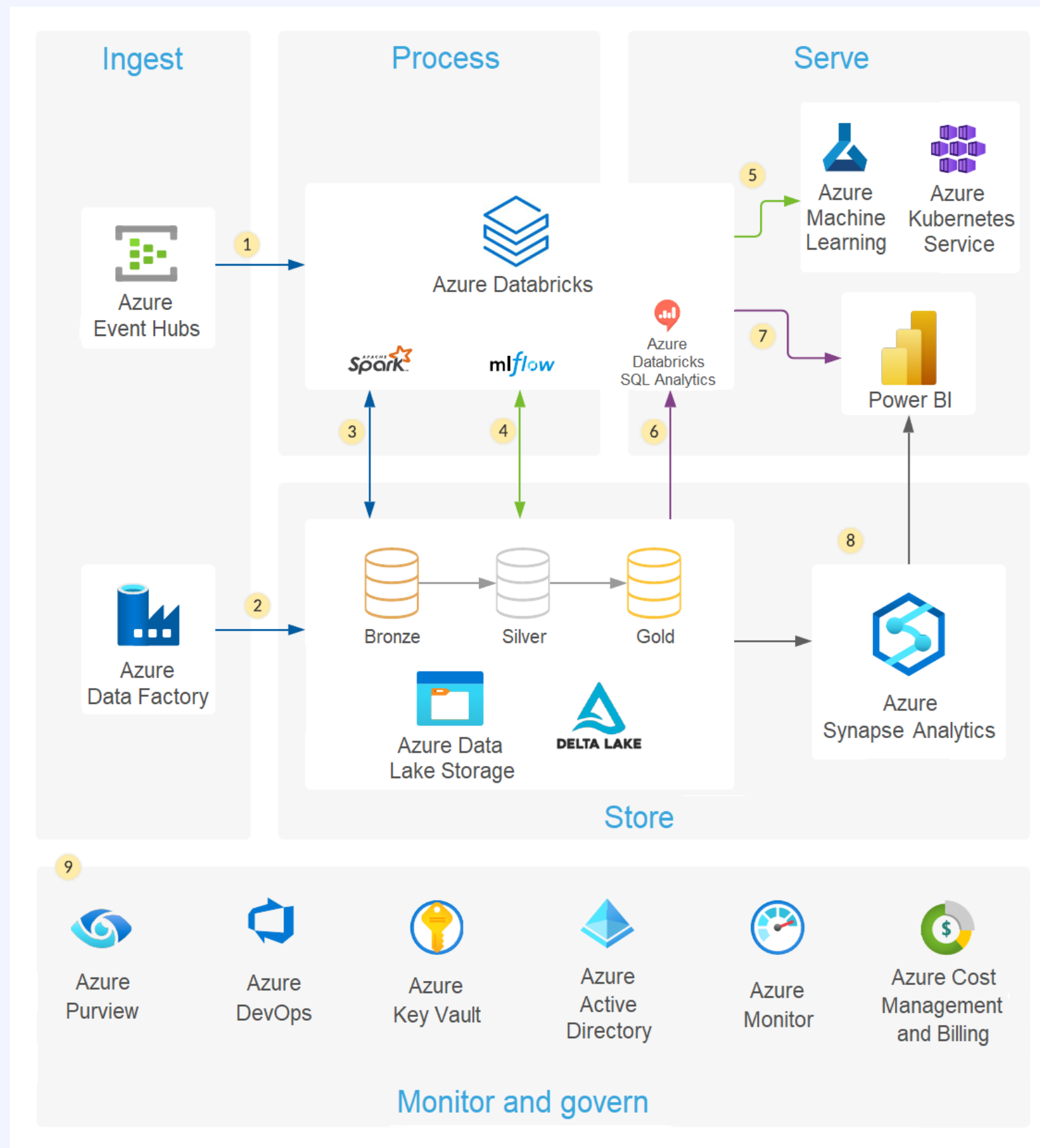
Unificar dados, análises e cargas de trabalho

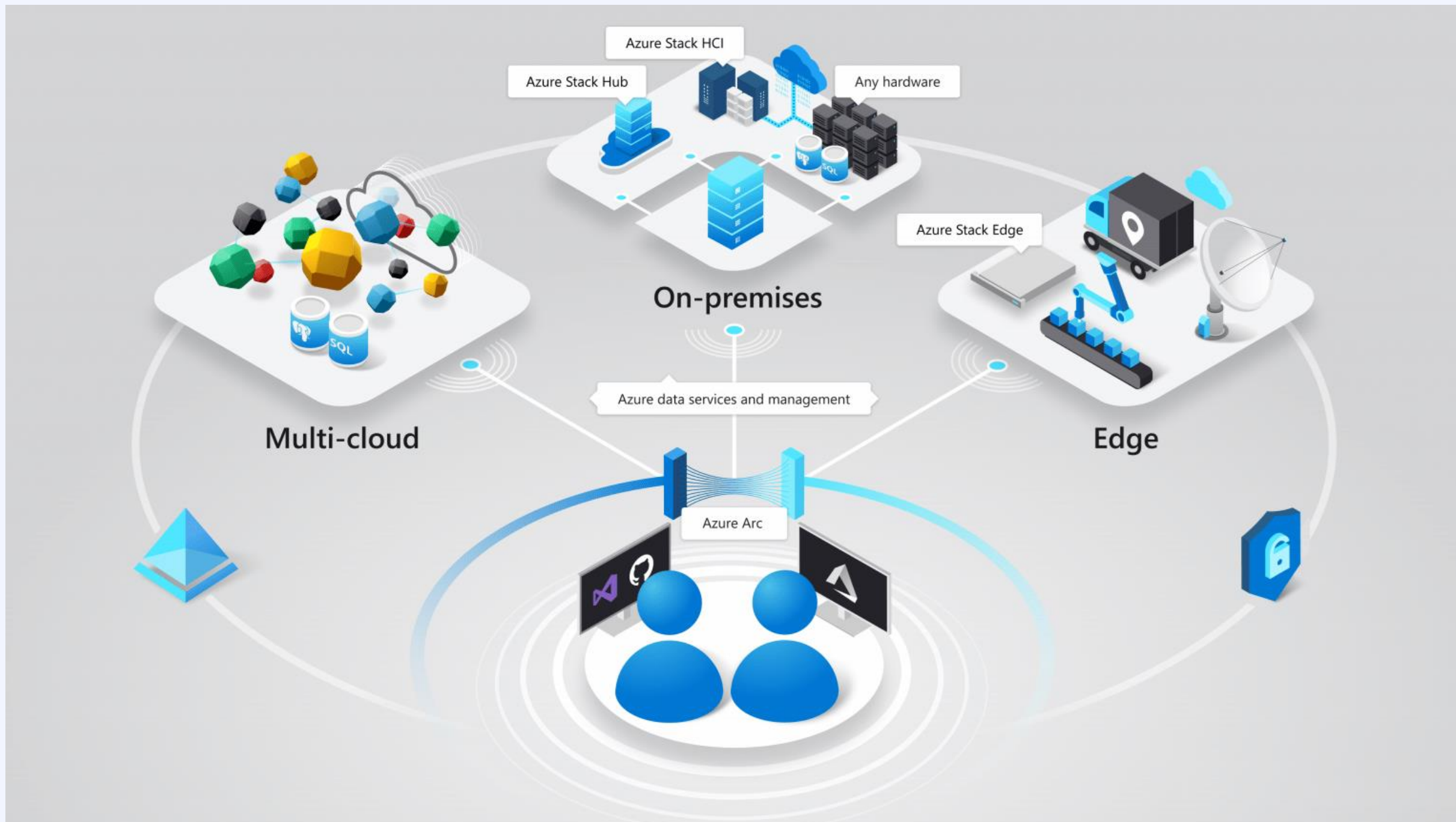


Executar de forma eficiente e confiável em qualquer escala



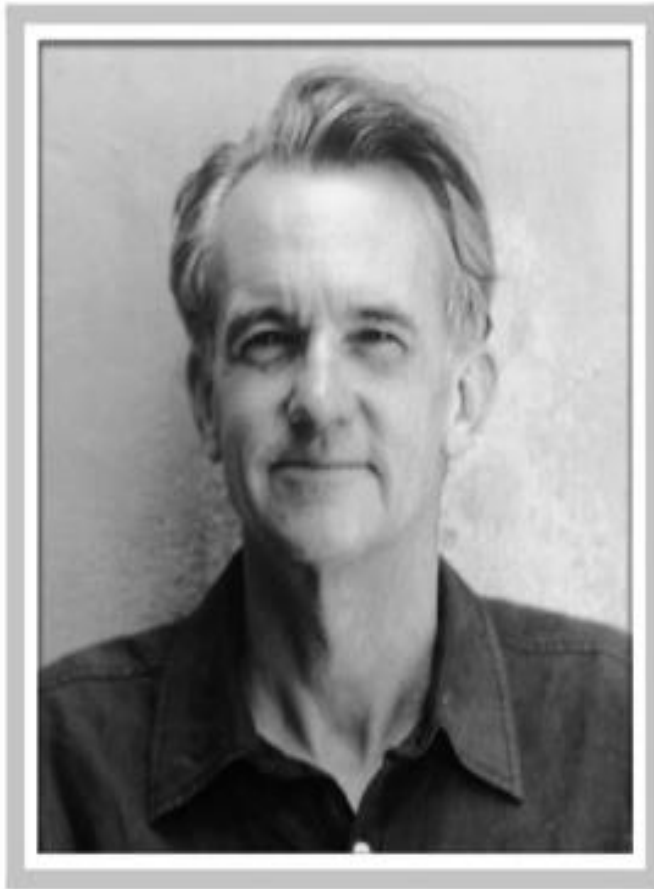
Fornecer insights por meio de painéis de análise, relatórios operacionais ou análises avançadas





TIPOS DE GRÁFICOS

CONCEITOS FUNDAMENTAIS



Edward Tufte

Data Visualization Pioneer
Professor, Yale University

PowerPoint is
Evil!

WIRED
Sep 2003

Stephen Few:

“A visualização de dados é a exibição gráfica de informação abstrata para duas finalidades: a tomada de decisão e comunicação”.

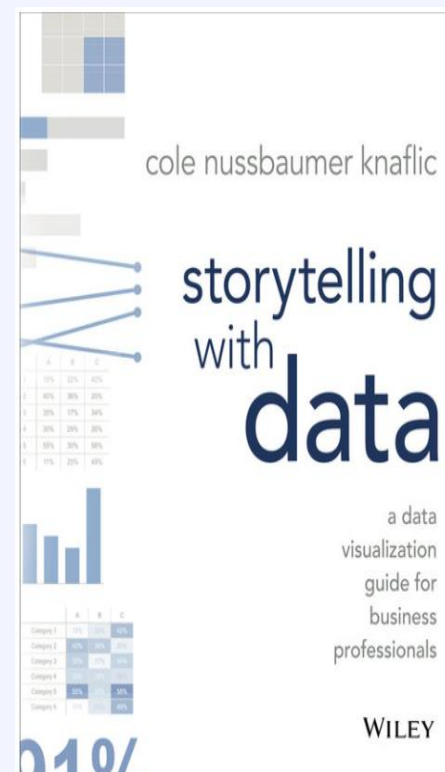
90% da informação transmitida ao cérebro é visual. No cérebro as informações visuais são processados 60.000 vezes mais rápidas do que texto.

Os conceitos apresentados nessa aula são baseados no livro **Storytelling com Dados**, da autora **Cole Nussbaumer Knaflic**, que se destina a qualquer pessoa que precisa se comunicar utilizando dados.

Uma visualização de dados eficaz pode significar a diferença entre o sucesso e o fracasso na hora de comunicar as constatações de um estudo/análise para o cliente, seu chefe, ou para seu público.

No geral, as pessoas que trabalham com funções analíticas possuem conhecimento quantitativo atuando nas etapas de coleta, tratamento, análise e modelagem, mas não necessariamente sabem (ou se interessam em saber) de que forma será a comunicação da análise. Sendo que, essa é justamente a etapa do processo que o público irá ver.

E como o mundo está cada vez mais voltado para os dados, mesmo quem não possui o conhecimento técnico precisa também aprender como se comunicar com dados.

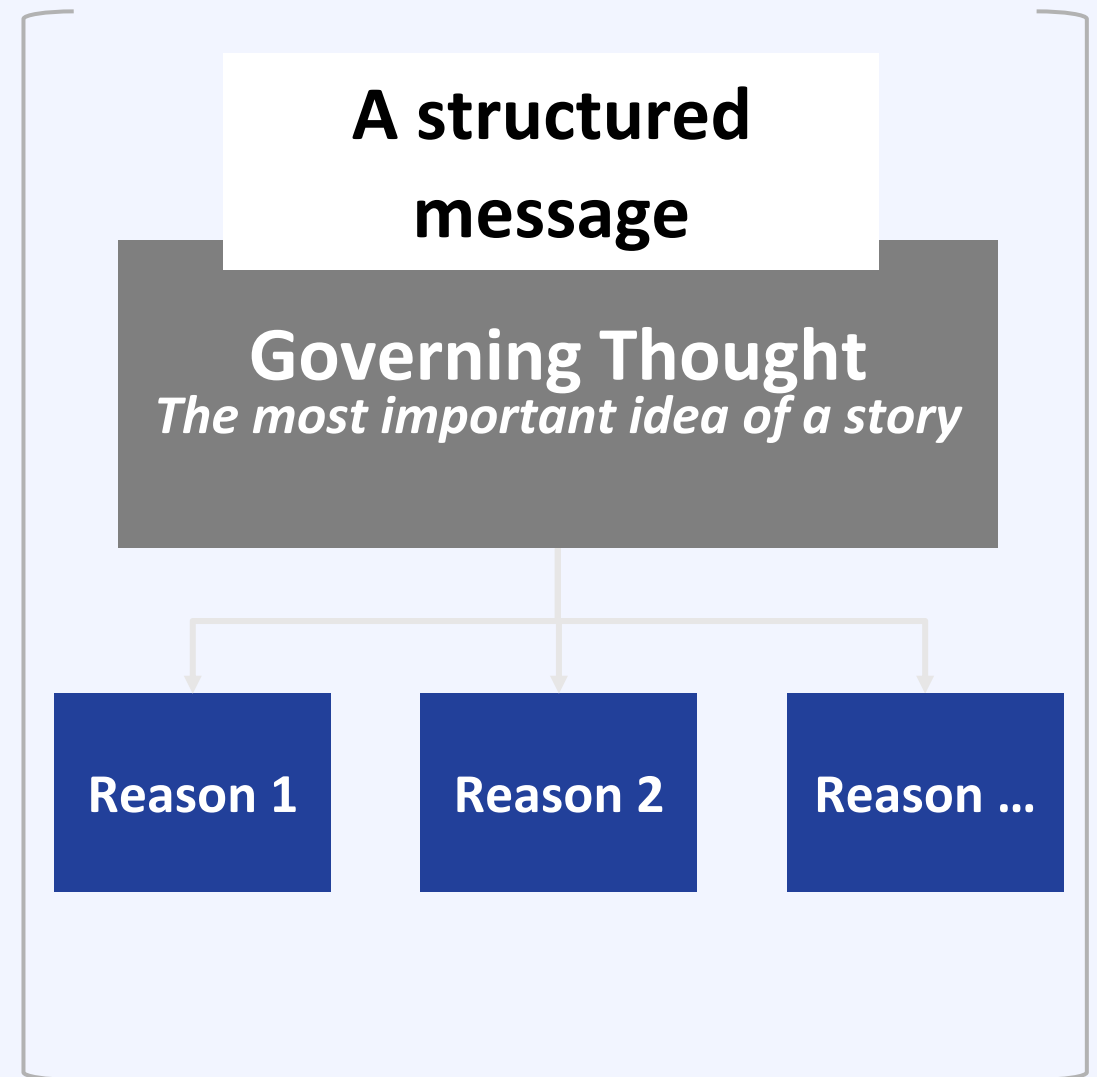
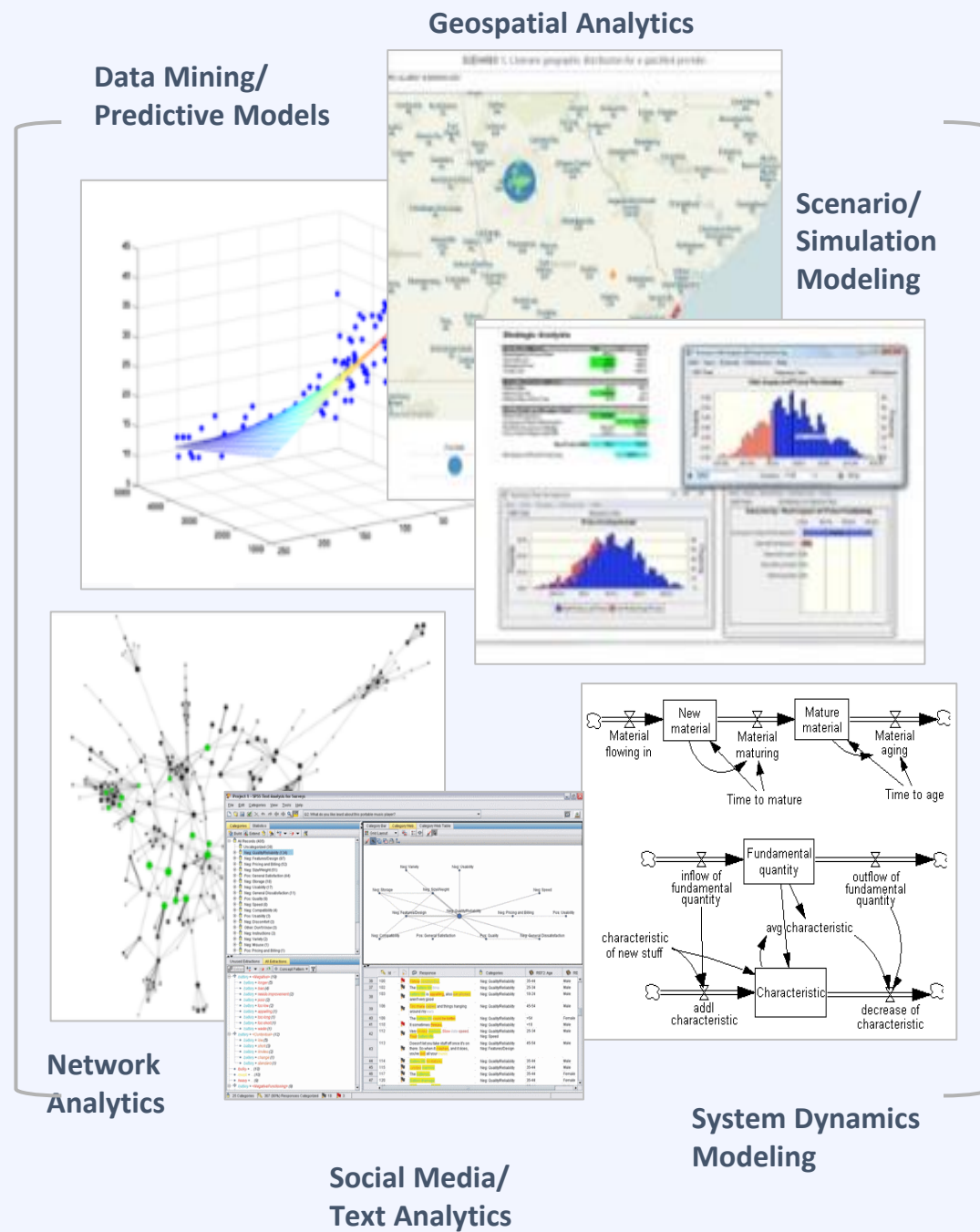


Data Science, Data Visualizations, SQL

The art and science of working with data

Storytelling with Data

The art and science of telling a story



Como contar histórias e suportar decisões a partir dos dados?

Análise Exploratória X Análise Explanatória

Concentre-se apenas na informação que seu público precisa saber.

Quem, O quê e Como?

Quem é o público-alvo?

O que queremos comunicar para o nosso público?

Quais dados tenho disponíveis e **como** utiliza-los para apresentar minha ideia?

História de 3 minutos

Se você tivesse apenas 3 minutos fazer sua apresentação, o que diria a seu público?

A Grande Ideia

Uma única frase que atenda os 3 componentes abaixo:

1. Deve articular seu ponto de vista único;
2. Deve transmitir o que está em jogo;
3. Deve ser uma frase completa.

Storyboard

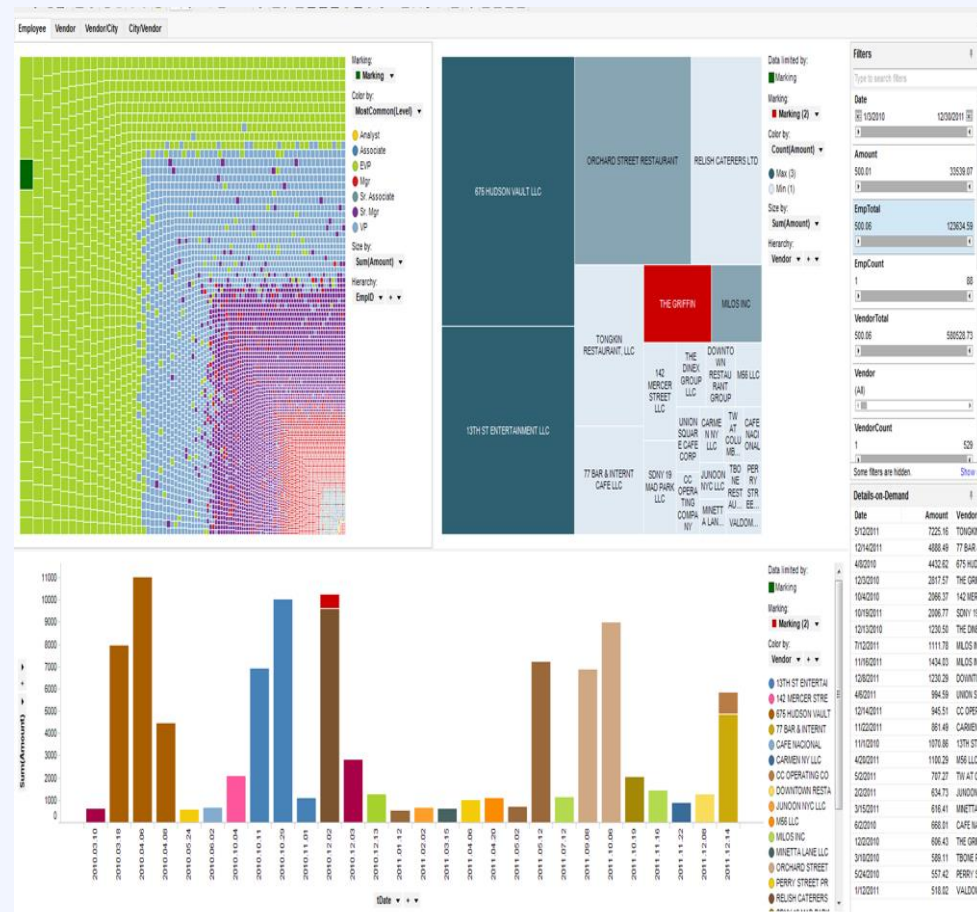
Não comece a gerar slides de modo aleatório.

O primeiro passo é estruturar o conteúdo visual a ser criado.



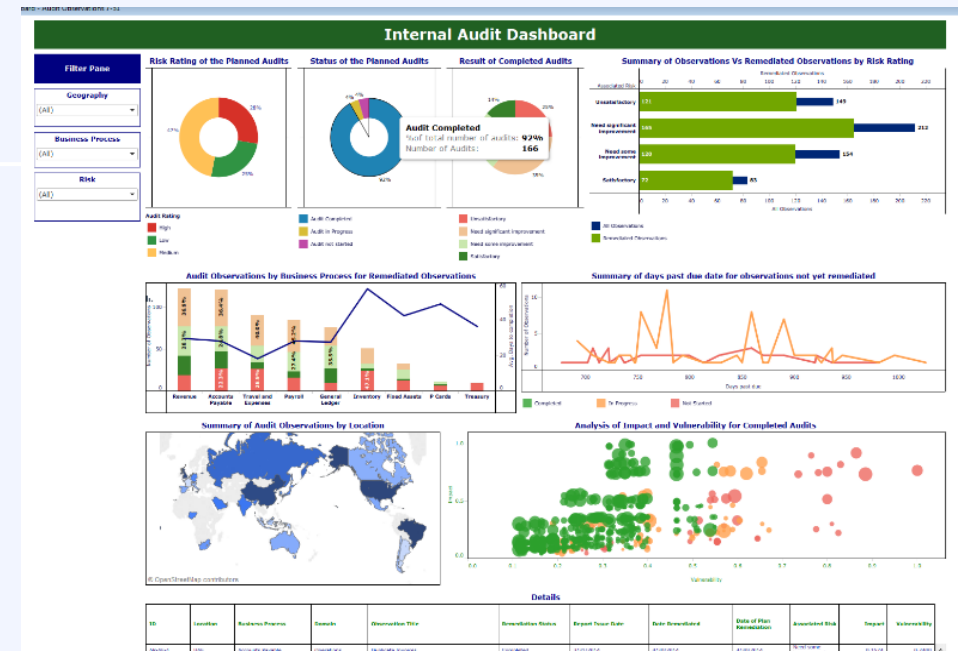
Exploratória (Visual Analytics)

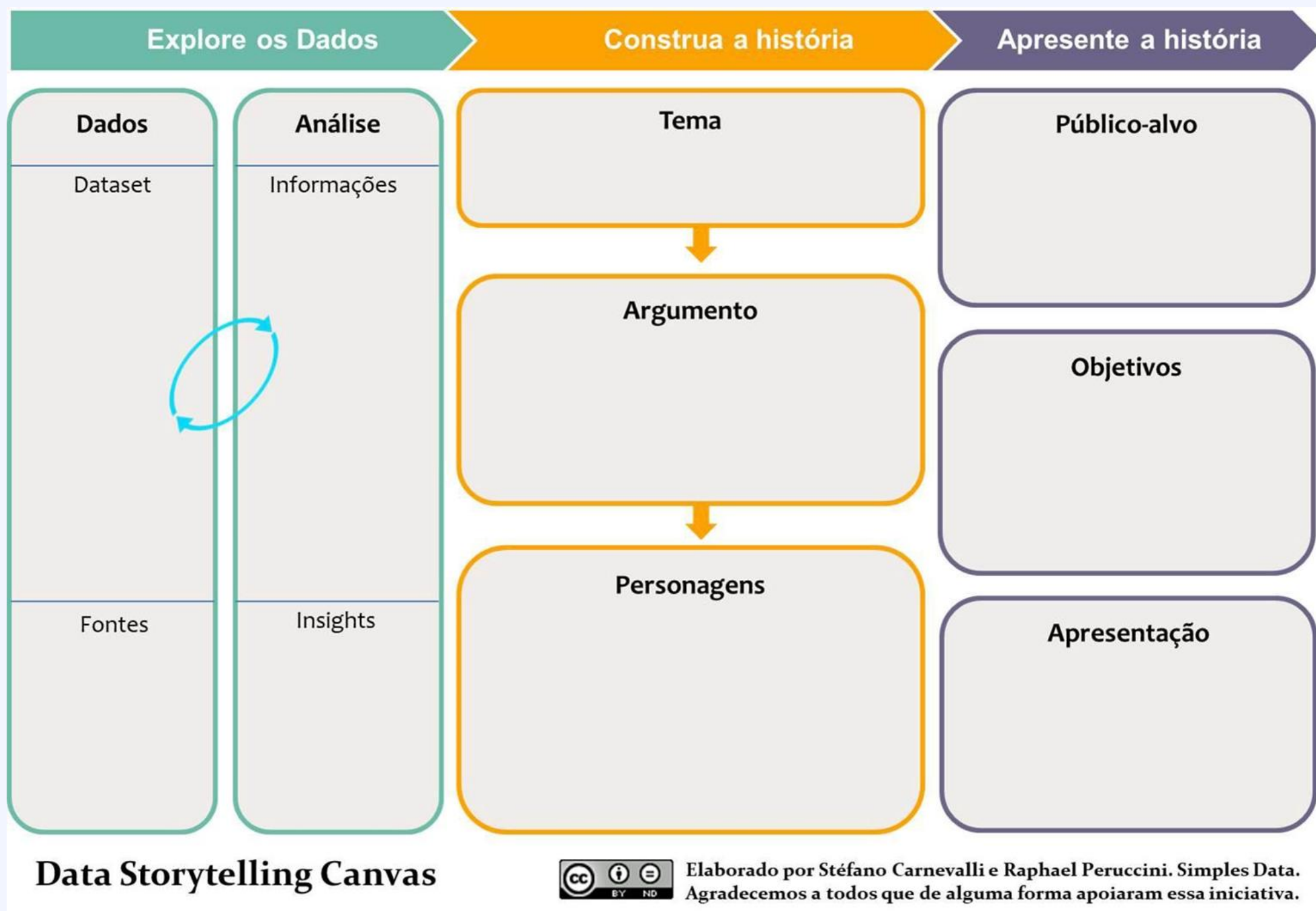
- Análise de dados para *revelar* dados e padrões DESCONHECIDOS, outliers e a história que os dados tem a dizer.



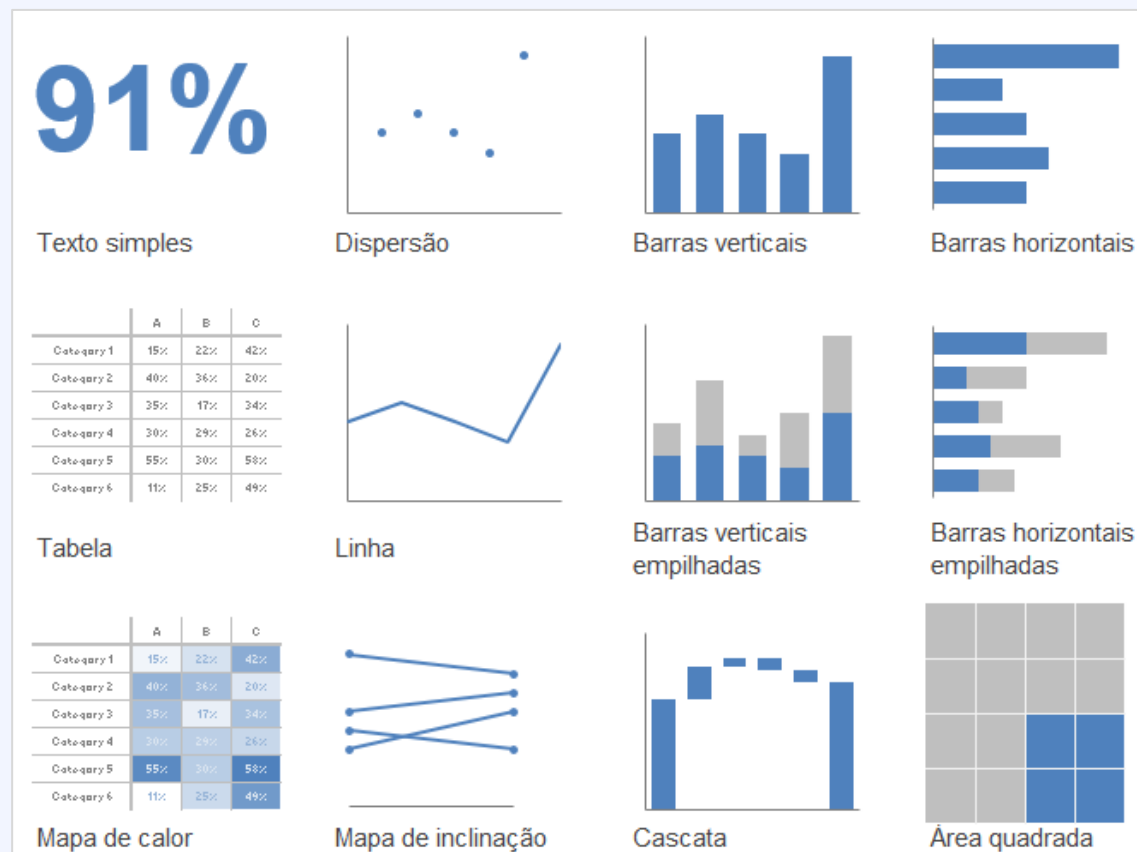
Explanatória (Data Visualization)

- Comunicar uma história CONHECIDA para os Executivos, mostrar padrões e tendências.



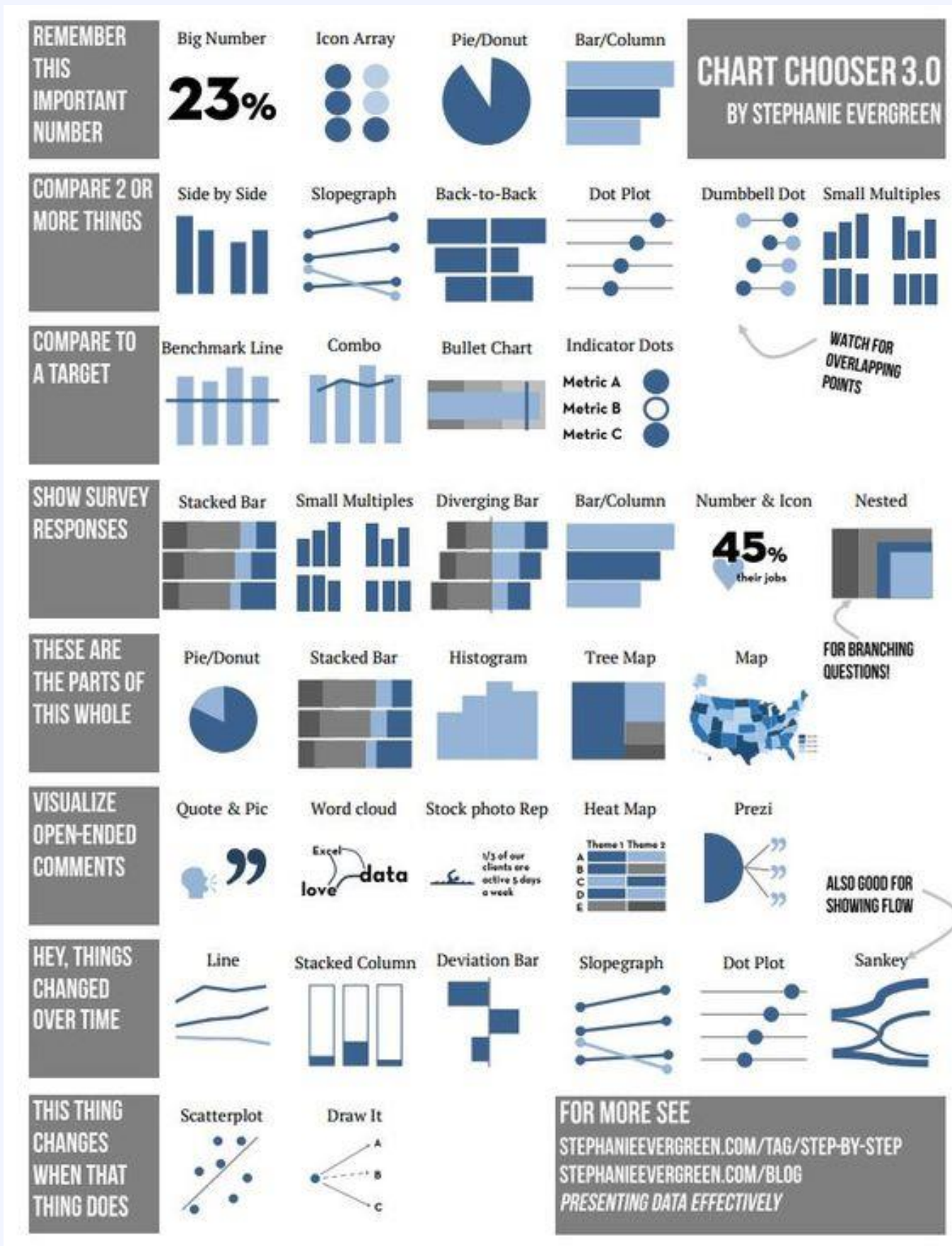


Principais modelos visuais utilizados...



E que devem ser evitados....





GUIA VISUAL PARA ESCOLHER O GRÁFICO PERFEITO

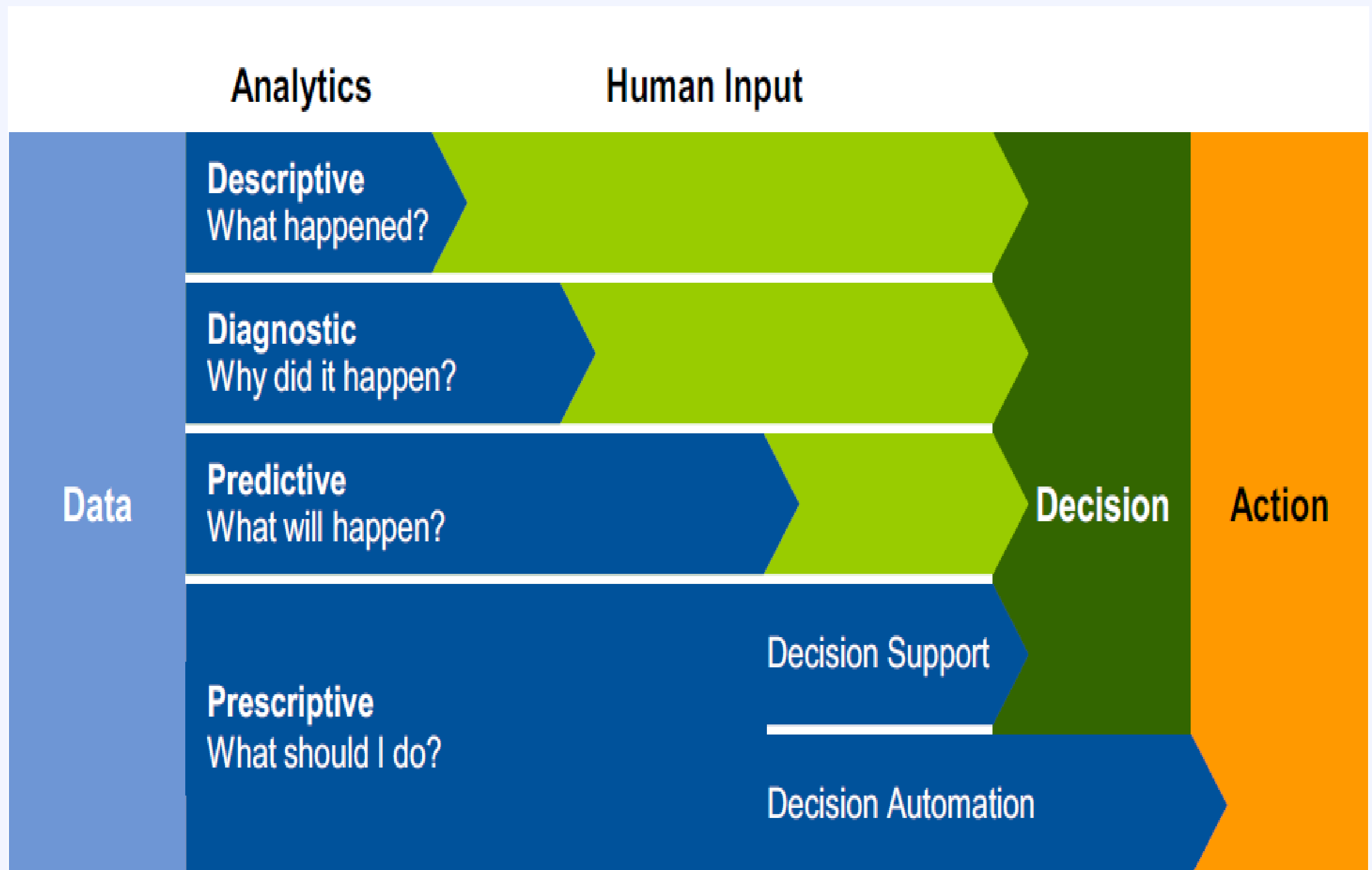


EXEMPLO PRÁTICO 1:N

TIPOS DE ANÁLISES

CONCEITOS FUNDAMENTAIS

- **Baseado em regras** - Tudo que você puder pensar em termos de if-then e yes/no.
- **Aprendizagem Supervisionada** - Você encontra um padrão e "fala" para a máquina - "Agora dá uma olhada no restante da base e me diga se tem algo parecido". Define-se um score e busca-se por similaridades. Buscamos responder um target, ou seja, há uma variável explícita a ser respondida.
- **Aprendizagem Não Supervisionada**, em que se busca identificar grupos ou padrões a partir dos dados, sem um objetivo específico a ser alcançado.



Source: Gartner (September 2013)

EXEMPLO PRÁTICO 1:N

OVERVIEW DE PROJETOS

OVERVIEW DE PROJETOS



CAPSTONES DO CURSO:

- Melhores partes
- Como extrair o melhor
- Life-long learner



CRIAÇÃO DE PORTFÓLIO:

- Importância
- Github



IMPORTÂNCIA DO PORTFÓLIO

- Ideias mais claras
- Histórico de atividades
- “Cartão de visitas”

RESUMO

OBRIGADO



OSVALDO ARANHA



(11) 97120-1719



osvaldoaranha@live.com

LinkedIn. 

