

The logo features the word "arkhe" in a bold, dark grey sans-serif font. Above the letter 'a' is a yellow arc that starts above the 'a' and ends above the 'k'. To the right of "arkhe" is the phrase "Data Analysis" in a yellow, italicized sans-serif font.

**arkhe** *Data Analysis*

- ✓ Introdução
- ✓ Solução Arkhe Data Analysis (ADA)
- ✓ DataViz
- ✓ Exemplos
- ✓ Case Mobilidade Regional



The screenshot shows the login interface for Arkhe Data Analysis. At the top, the Arkhe logo is displayed next to a yellow button labeled "Data Analysis". Below the logo, there are two input fields: "Nome de usuário:" with a subtext "Digite o nome de usuário" and "Senha:" with a subtext "Digite a senha". A checkbox labeled "Mantenha-me conectado" is positioned below the password field, and a link "Esqueceu a senha?" is to its right. A dark grey "Login" button is centered below the form. At the bottom, there is a link "Não tem conta? Registre-se".

The logo for 'arkhe Data Analysis' features the word 'arkhe' in a bold, dark grey sans-serif font. Above the 'a' in 'arkhe' is a yellow arc that curves over the top of the letter. To the right of 'arkhe' is the phrase 'Data Analysis' in a yellow, italicized sans-serif font.

**arkhe** *Data Analysis*

**Introdução**

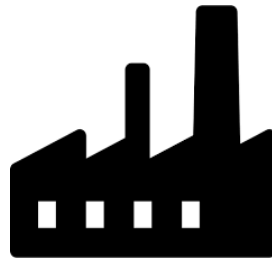


## ✓ ARKHE Intelligence

- ✓ **ARKHE – Data Analysis**
- ✓ ARKHE – COMINT
- ✓ ARKHE – ELINT

**arkhe** *Data Analysis*

**MAIS DADOS...**



**... MAIS  
CONHECIMENTO?**



**Sistemas de  
Informação**



## ARKHE DATA ANALYSIS

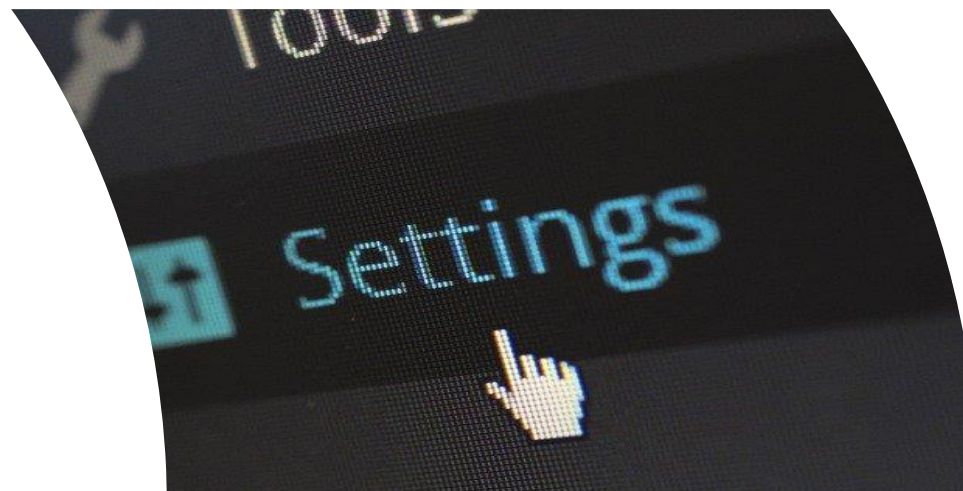
- Melhorar a governança dos dados
- Ampliar a tomada de decisão baseada em dados
- Ampliar a análise de dados e o uso de IA
- Expandir a compreensão dos dados que uma organização geral
- Manter custos escalonáveis no uso de dados



**Software,  
Consultoria e  
Treinamento**

# Solução ARKHE Data Analysis – Como?

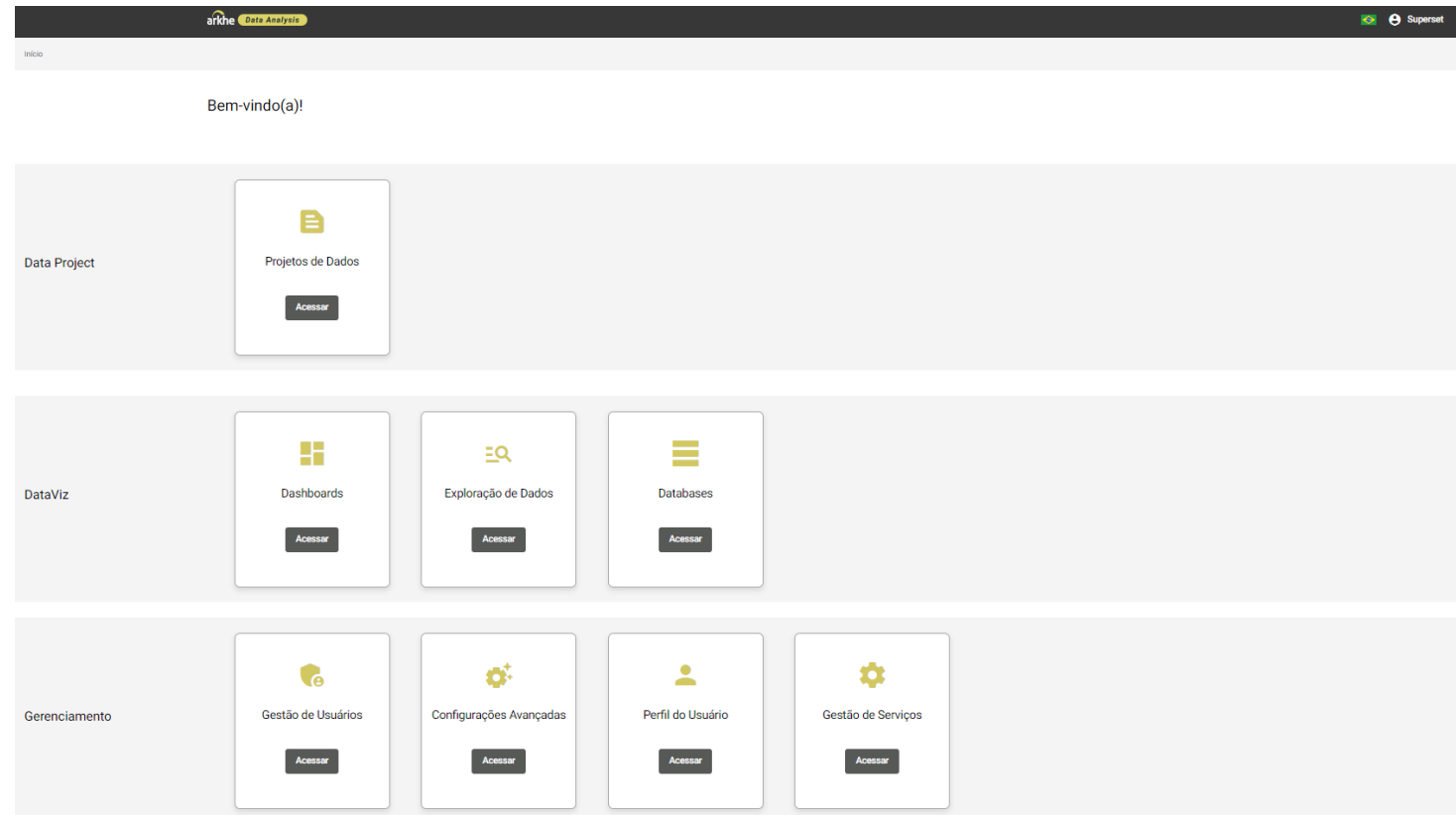
- Software
- Consultoria
- Treinamento



**arkhe** *Data Analysis*



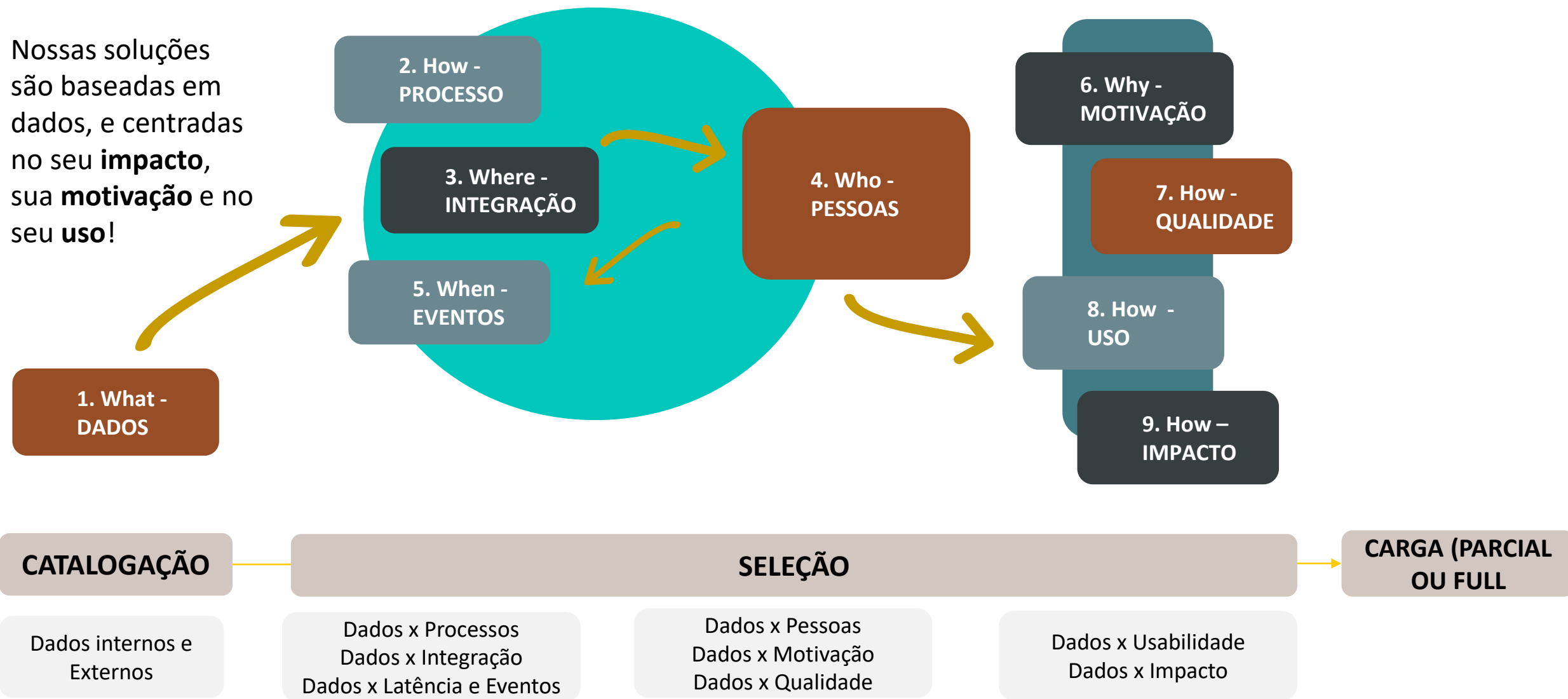
# ADA – Software 1.1



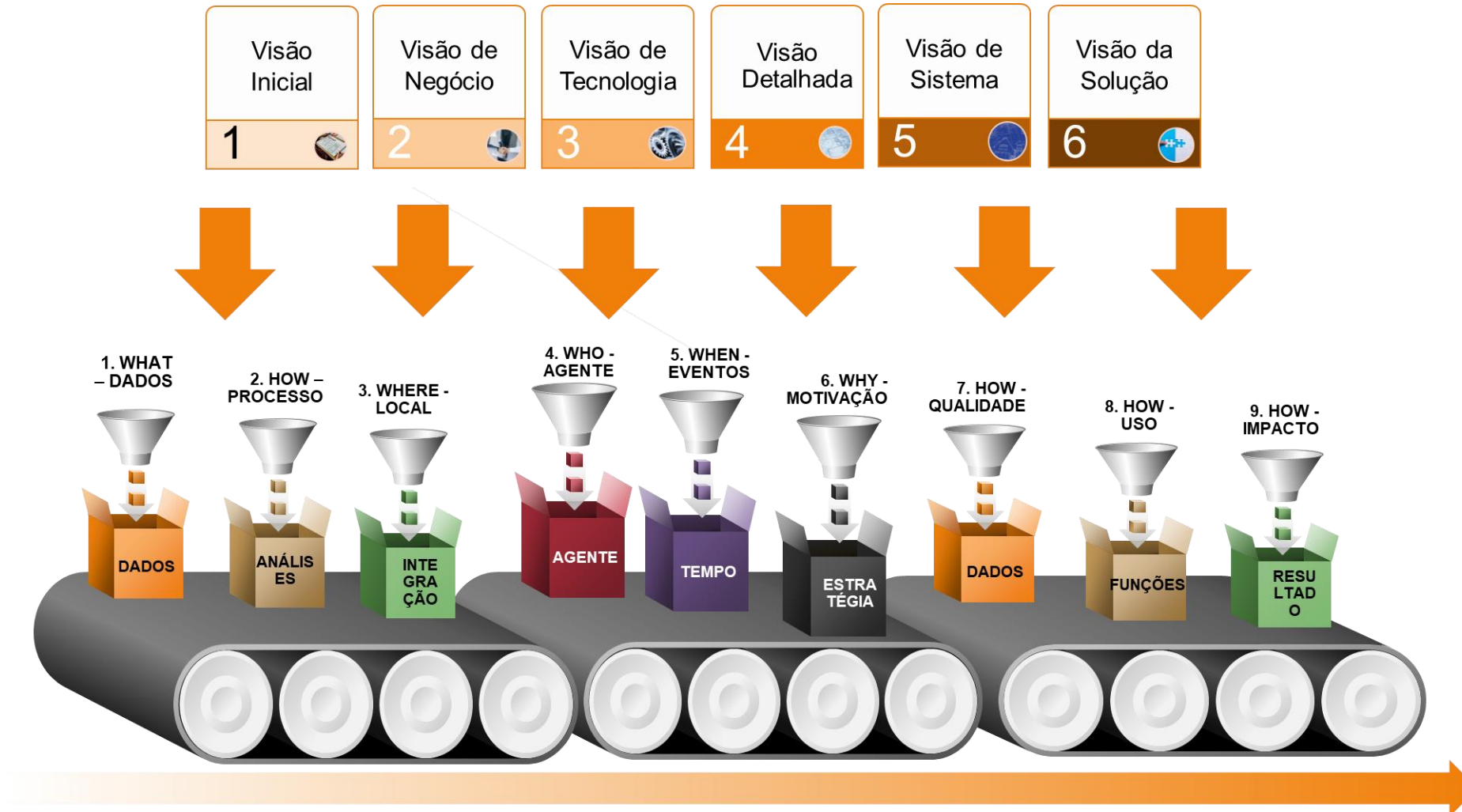


# ADA – Consultoria e Serviços

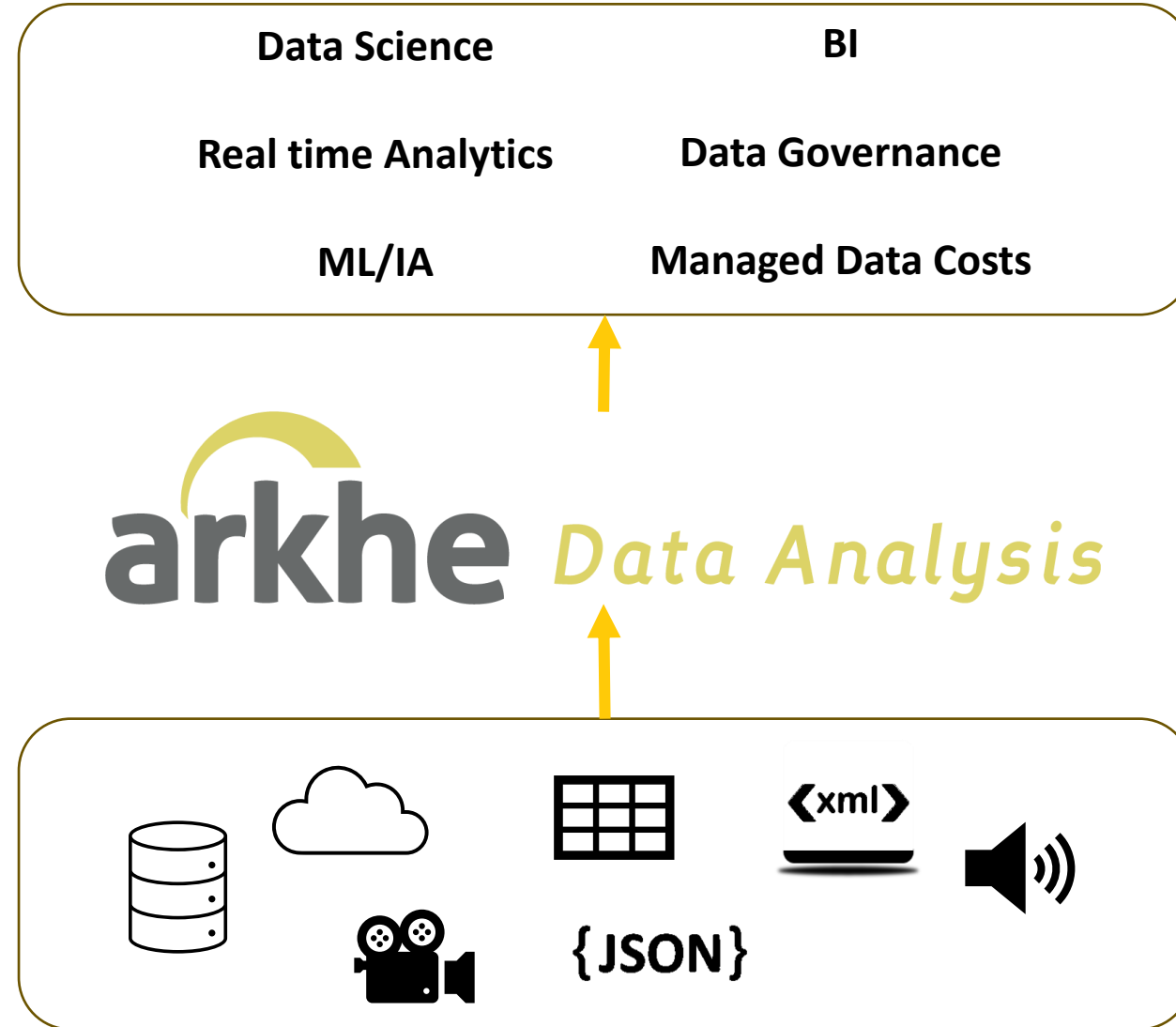
Nossas soluções são baseadas em dados, e centradas no seu **impacto**, sua **motivação** e no seu **uso!**



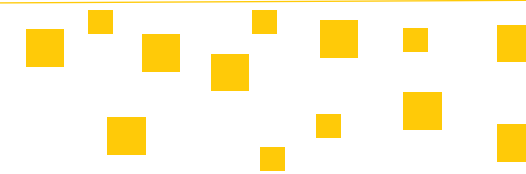
# ADA – Consultoria e Serviços



# ADA – Consultoria e Serviços - Resultados



# ADA – Treinamentos



## Curso Básico

- Introdução E Objetivo
- Conceitos Básicos
- Conhecendo o DataViz
- Operações Básicas
- Campos calculados e Métricas
- Exploração Visual de Dados

## SQL Avançado no SQL Lab

- Introdução E Objetivo
- Apresentando O Sql Lab
- Ciclo De Vida Dos Dados
- Uso Dos Dados
- Caracterização Dos Dados
- Introdução À Linguagem Sql
- Análise Inicial (Profiling) De Dados Com Sql No Dataviz
- Preparação De Dados Com Sql No Dataviz
- Enriquecimento De Dados Com Sql

## Gráficos de Mapas

- Introdução E Objetivo
- Dados Geoespaciais
- Tipos De Dados De Geometria
- Tipo De Mapas
- Configurações Javascript

## Dashboards em tempo real

- Introdução E Objetivo
- Bancos De Dados Oltp
- Bancos De Dados Olap
- Data Lakes
- Gráficos Da Série Time
- Definir Intervalo De Atualização Automática

**Hadoop**  
**Spark**  
**Pentaho**

## Gráfico De Séries Temporais

- Introdução E Objetivo
- Gráficos Temporais
- Configurações De Dados Temporais Em Gráficos
- Camadas De Anotação
- Comparação De Dados Temporais
- Análises Avançadas
- Análises Preditiva

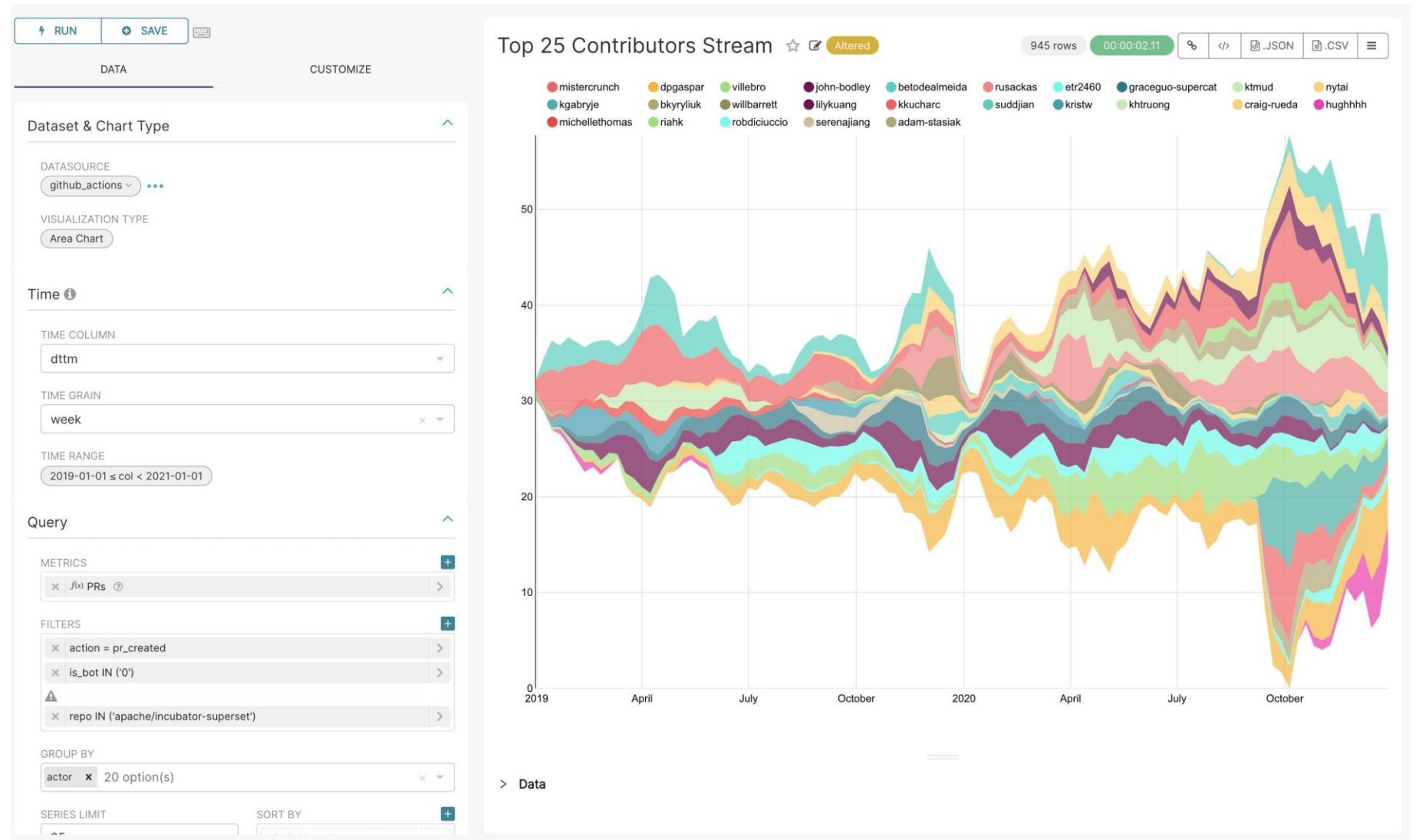
The logo for 'arkhe Data Analysis' features the word 'arkhe' in a bold, dark grey sans-serif font. Above the 'a' in 'arkhe' is a yellow arc that curves over the top of the letter. To the right of 'arkhe' is the phrase 'Data Analysis' in a yellow, italicized sans-serif font.

**arkhe** *Data Analysis*

**DATAVIZ**

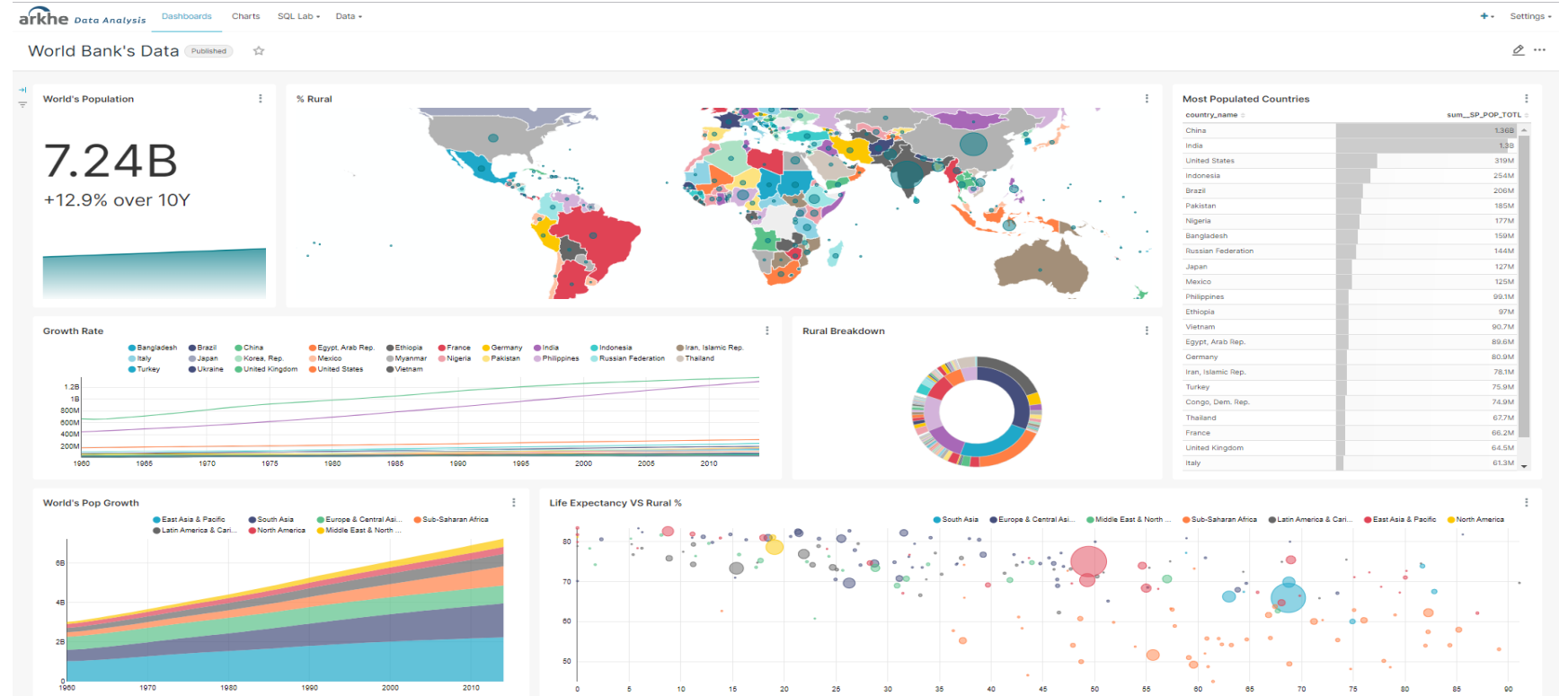
# DataViz

*Exploração visual  
e monitoramento  
de grandes  
volumes de dados  
em tempo real!*



# DataViz

*Exploração visual  
e monitoramento  
de grandes  
volumes de dados  
em tempo real!*



# ADA

*Interface intuitiva,  
dashboards interativos*

*Vários tipos de gráficos*

*Criação quase sem código  
(SQL, MD, JavaScript)*

*Interface SQL para  
preparação e exploração  
visual, e metadados*

*Simplicidade para definir  
dimensões e métricas  
personalizadas*

*Suporte para a maioria  
dos bancos de dados SQL*

*Cache assíncrono e  
consultas assíncronas e  
in-memory*

*Segurança extensível com  
regras sobre quem pode  
acessar quais recursos do  
produto e conjuntos de  
dados*

*Integração com  
autenticação (banco de  
dados, OpenID, LDAP,  
OAuth, REMOTE\_USER,  
etc.)*

*A capacidade de adicionar  
plugins de visualização  
personalizados*

*Uma API para  
personalização  
programática*

*Uma arquitetura nativa  
da nuvem projetada para  
escala horizontal*

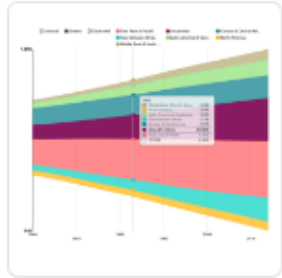


# ADA – Formatos de Dados

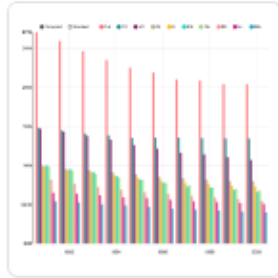
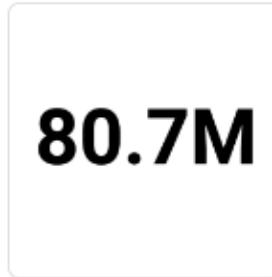
1. Amazon Athena
2. Amazon Redshift
3. Apache Drill
4. **Apache Druid**
5. Apache Hive
6. Apache Impala
7. Apache Kylin
8. Apache Pinot
9. Apache Solr
10. Apache Spark SQL
11. Ascend.io
12. Azure MS SQL
13. Big Query
14. ClickHouse
15. CockroachDB
16. Dremio
17. Elasticsearch
18. Exasol
19. Google Sheets
20. Firebolt
21. Hologres
22. IBM Db2
23. IBM Netezza
24. MySQL
25. Oracle
26. PostgreSQL
27. Trino
28. Presto
29. SAP Hana
30. Snowflake
31. SQLite
32. SQL Server
33. Teradata
34. TimescaleDB
35. Vertica
36. YugabyteDB
37. CSV



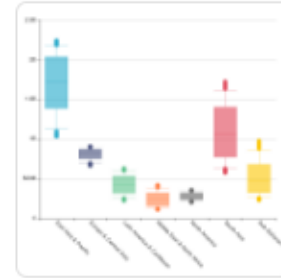
# ADA – Gráficos



Area Chart

Time-series  
Bar ChartBig Number  
with Trendline

Big Number



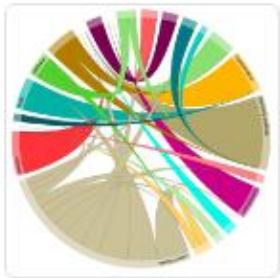
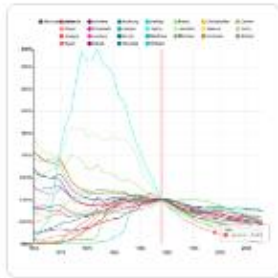
Box Plot



Bubble Chart



Bullet Chart

Calendar  
HeatmapChord  
DiagramTime-series  
Percent  
Change

Country Map



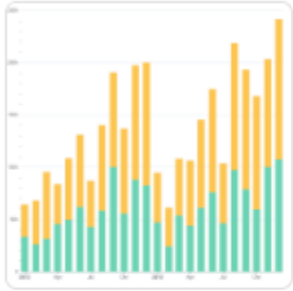
deck.gl Arc

deck.gl  
Geojson

deck.gl Grid

deck.gl 3D  
Hexagondeck.gl  
Multiple  
Layers

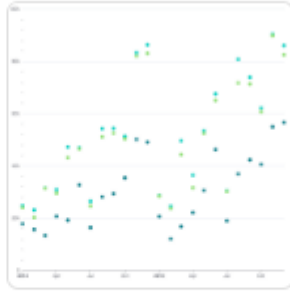
# ADA – Gráficos



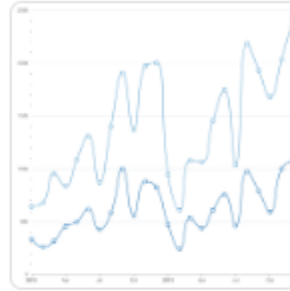
Time-series  
Bar Chart v2



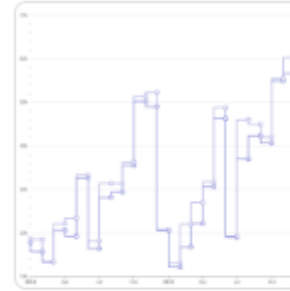
Time-series  
Line Chart



Time-series  
Scatter Plot



Time-series  
Smooth Line



Time-series  
Stepped Line



Event Flow



Filter box



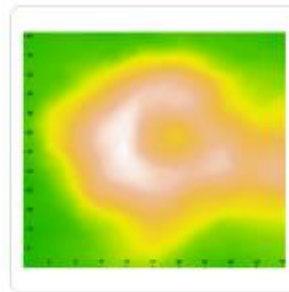
Funnel Chart



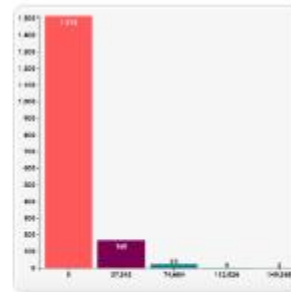
Gauge Chart



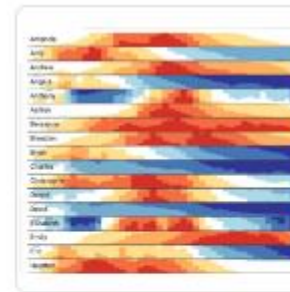
Graph Chart



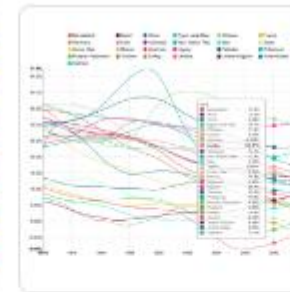
Heatmap



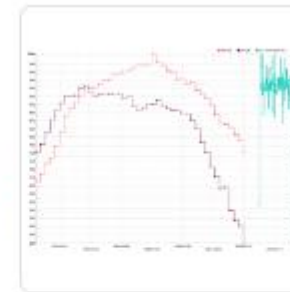
Histogram



Horizon Chart



Line Chart



Multiple Line  
Charts



MapBox

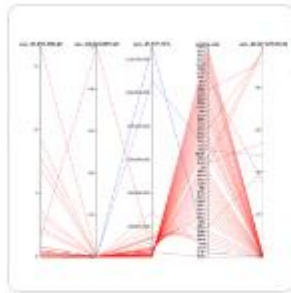
# ADA – Gráficos



Mixed Time-Series



Paired t-test Table



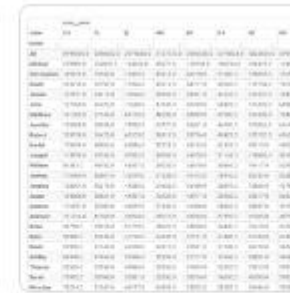
Parallel Coordinates



Partition Chart



Pie Chart



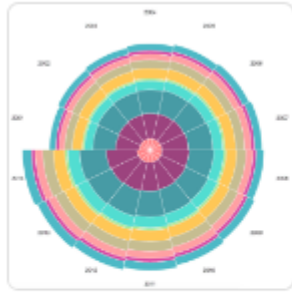
Pivot Table



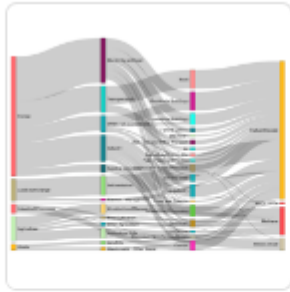
Pivot Table v2



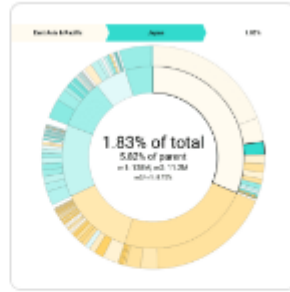
Radar Chart



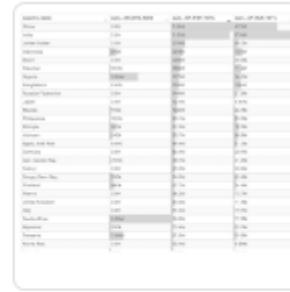
Nightingale Rose Chart



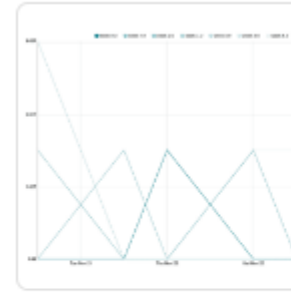
Sankey Diagram



Sunburst Chart



Table



Time-series Period Pivot



Time-series Table



Tree Chart



Treemap

# ADA – Roadmap

## Versão 1.0 – Programa BigData – SLT (jul/22)

- ✓ Real Time Analytics (BI in memory + Big Data)
- ✓ Desenvolvimento do núcleo de Analytics (Time, capacitação, desenvolvimento de projeto)
- ✓ Infraestrutura - servidores

## Versão 1.1 – Consolidação do RealTimeAnalytics (jul/23)

- ✓ Módulo Access Manager – Segurança e governança
- ✓ Plataforma ARHKE – interface revista, com foco em uma plataforma única para os produtos de defesa

## Versão 1.2 – Consolidação do RealTimeAnalytics (dez/23)

- ✓ Arquitetura modular – buliding blocks, Docker-Image, Cluster Kubernetes
- ✓ Gerenciamento de Serviços – Monitoramento, Log da aplicação e backup

## Versão 2.0 – DADOS - 2024/2025

- ✓ ChatBotAI para consulta de Dados
- ✓ Evolução da Gestão de Dados
- ✓ Evolução da Ingestão de dados
- ✓ Adição da integração de dados

## Versão 3.0 – IA/MLOPs - 2024/2025

- ✓ Criação de modelos e API's de IA
- ✓ Monitoramento e Evolução de modelos de IA

Quem vem primeiro?  
Dados ou IA

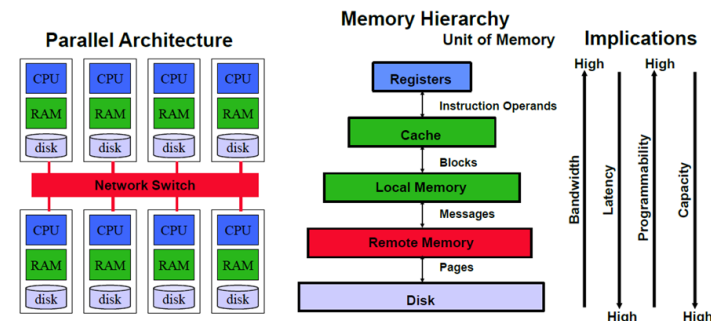
**TECH PUSH X  
MARKETING  
DEMAND**

# ADA – v2 – Dados

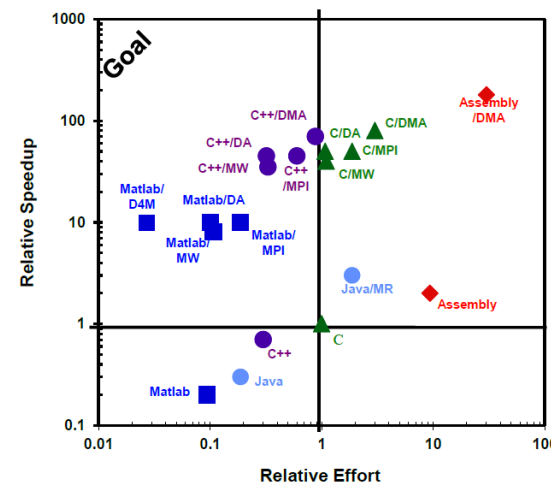
## PARADIGMAS DE OTIMIZAÇÃO DE PERFORMANCE EM DADOS

1. Processamento distribuído
2. Modelos de programação em paralelo
3. Modelagem para big data

### 1. Processamento distribuído



### 2. Modelos de programação em paralelo



### PARALLEL PROGRAMMING MODELS:

- ✓ DMA – DIRECT MEMORY ACCESS
- ✓ MPI – MESSAGE PASSING
- ✓ DA – DISTRIBUTED ARRAYS
- ✓ MW – MANAGER/WORKER
- ✓ MR – MAP/REDUCE
- ✓ D4M – DYNAMIC DISTRIBUTED DIMENSIONAL DATA MODEL

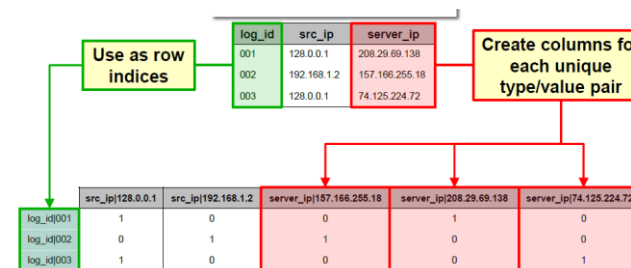
### 3. Modelagem para Big Data



Usando métodos tradicionais de big data – 240 horas  
 Usando modelo orientado a big data (D4M) – 5 horas

- Aumento de 9x no tamanho do dado (95% de compactação (ORC))
- 48x mais rápido!

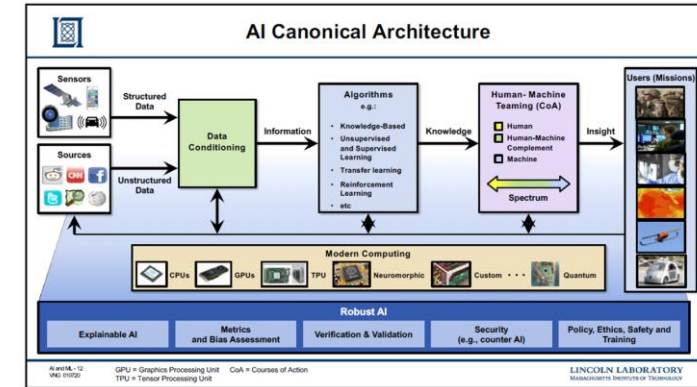
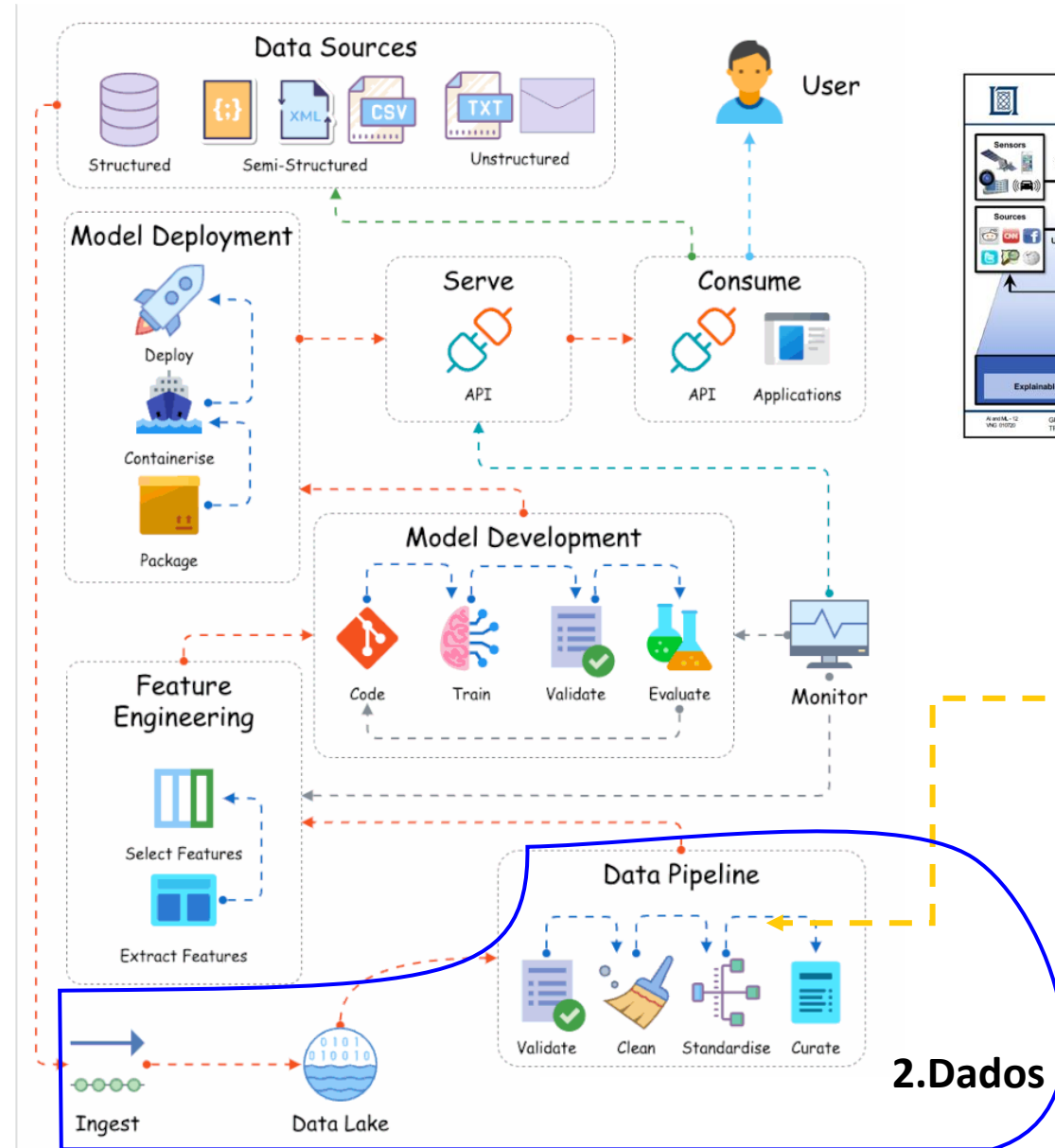
Matrizes Associativas



3 linhas x 4 colunas = (12) → 3x6 + 6x3 = 36, ou 3x mais dados que o input inicial!

# ADA – v3 – MLOps

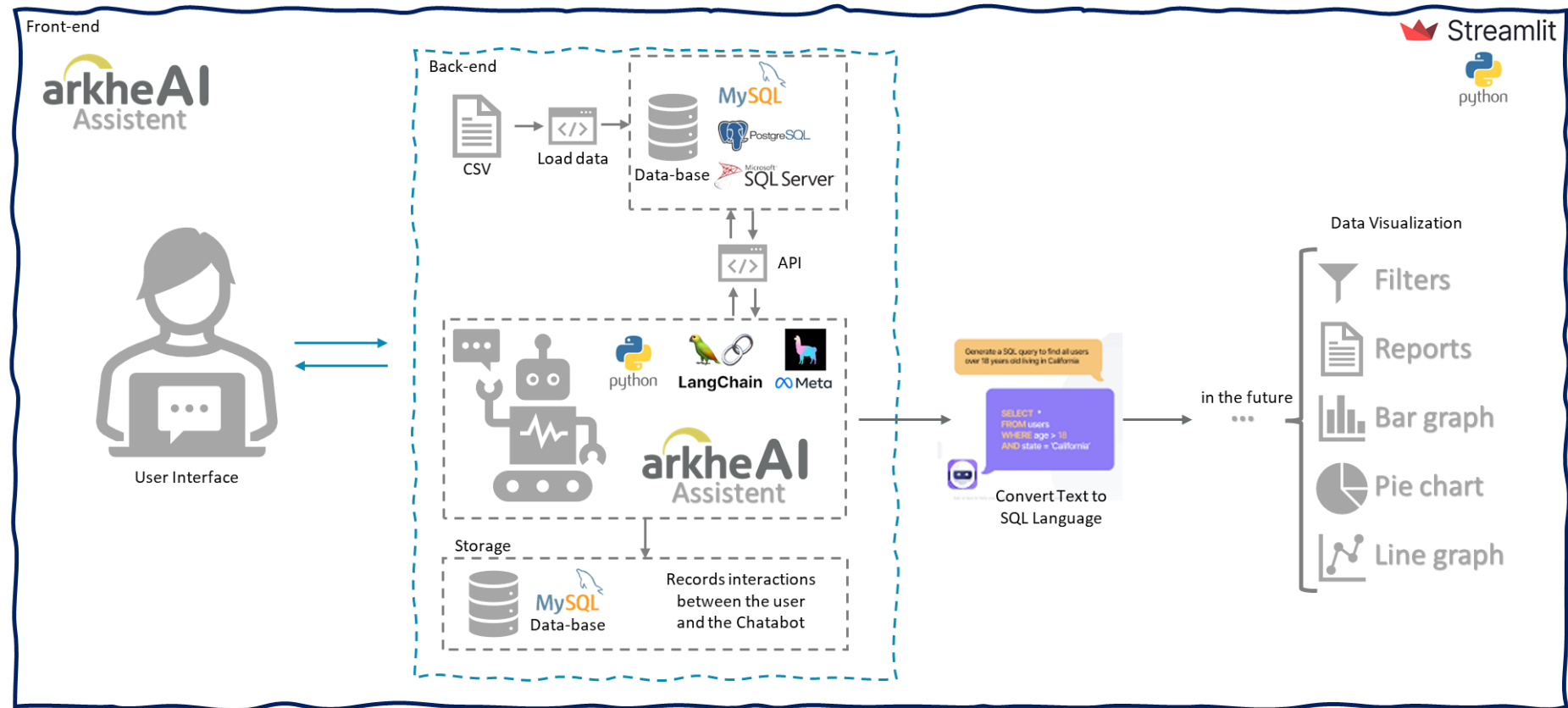
- Estudos de arquitetura
- ChatBotAI para SQL



**1. DataViz - EDA**

# ADA – v3 – MLOps

- Estudos de arquitetura
- ChatBotAI para SQL





# ADA – Why/Why not

## ✓ IMPLANTAÇÃO DE PROJETOS CORPORATIVOS DE REAL TIME ANALYTICS, ESCALÁVEIS E SEGUROS!

- ✓ Licenciamento perpétuo com instalação on Premises ou em nuvem
- ✓ Suporte customizado
- ✓ Módulo Access Manager – Governança e Segurança, gestão de serviços e clusters
- ✓ Orientado a Real Time Analytics (OLAP NoSQL + BI in memory)
- ✓ Transferência de conhecimento – trilha com 7 cursos, e evoluindo!
- ✓ Interface rica em gráficos e conexão de dados, ambiente de exploração visual de dados

### ✓ GRÁFICOS

### ✓ FONTES DE DADOS

- ✓ *Consultas SQL através de consultas em linguagem natural usando IA! (em breve!!!)*

# ADA – Comparação

Característica	ADA	MicroStrategy	Tableau	QlickView	MS PowerBI
Escalonamento Horizontal sem custos	✓	✗	✗	✗	✗
Customização de interface	✓	✗	✗	✗	✗
Customização de regras de login	✓	✗	✗	✗	✗
Criação de Novos Módulos	✓	✗	✗	✗	✗
Governança e Segurança	✓	✗	✗	✗	✗
Licenciamento perpétuo, SaaS e On Premisses	✓	✗	✗	✗	✗
Gráficos e interatividade	✗	✓	✓	✓	✓
Mapas	✗	✗	✓	✗	✗
Interoperável com sistemas C4I e Inteligência	✓	✗	✗	✗	✗

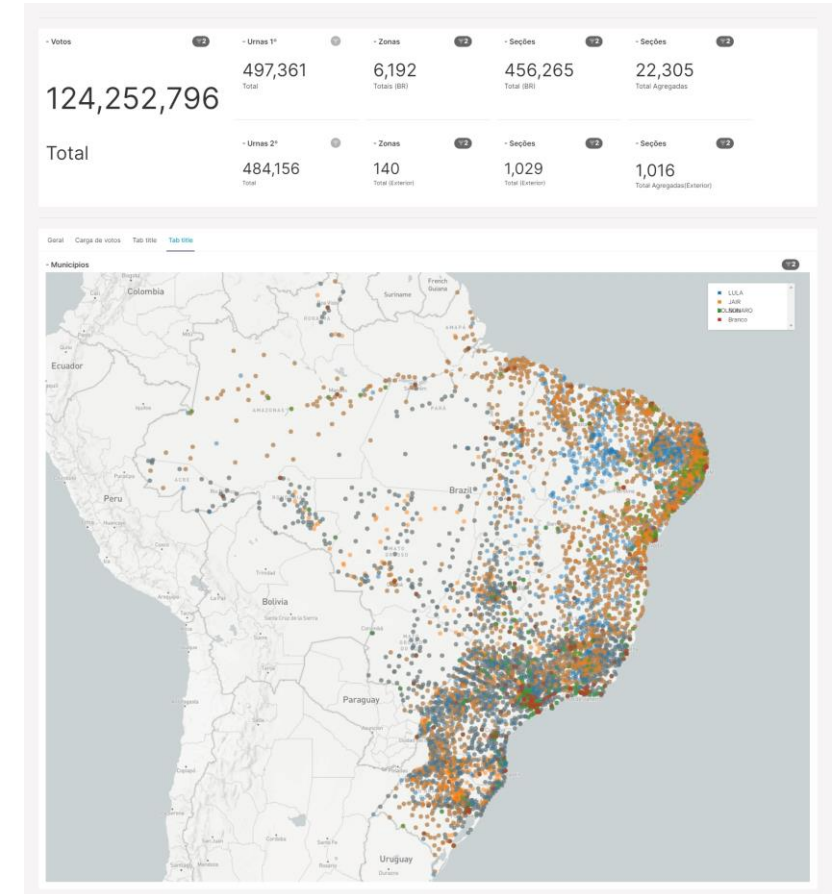
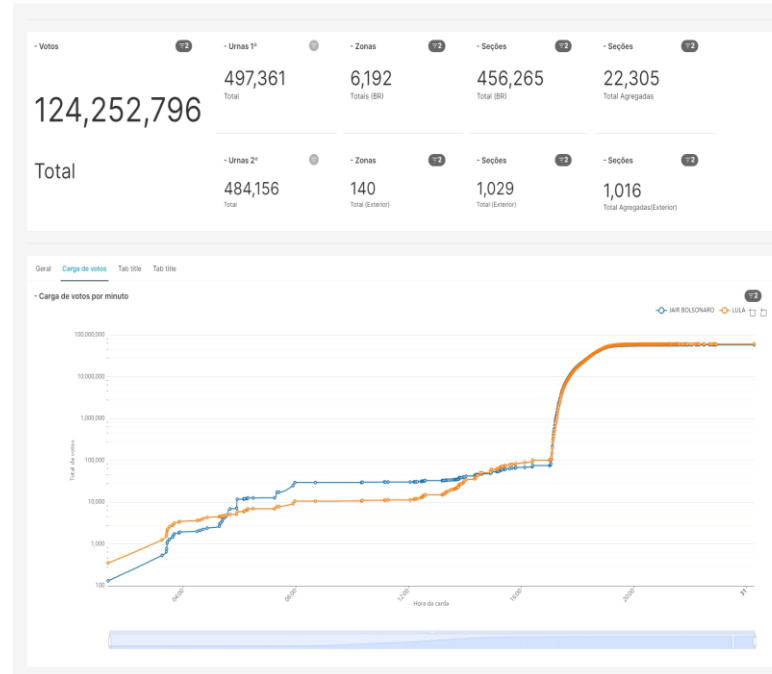
**Governança e Segurança: Módulo Access Manager** – assegura o gerenciamento centralizado de usuários, autenticação e autorização seguras, integração com LDAP, AD e outros, configurações avançadas de defesa de ataques de força bruta, autenticação de dois fatores, Ele suporta vários protocolos de segurança, como OAuth2, OpenID Connect e SAML, e oferece recursos avançados de autenticação e autorização, como autenticação multifator e autenticação baseada em certificado

The logo for 'arkhe Data Analysis' features the word 'arkhe' in a dark grey, lowercase, sans-serif font. Above the 'a' is a yellow arc that curves over the top of the letter. To the right of 'arkhe' is the phrase 'Data Analysis' in a yellow, italicized, sans-serif font.

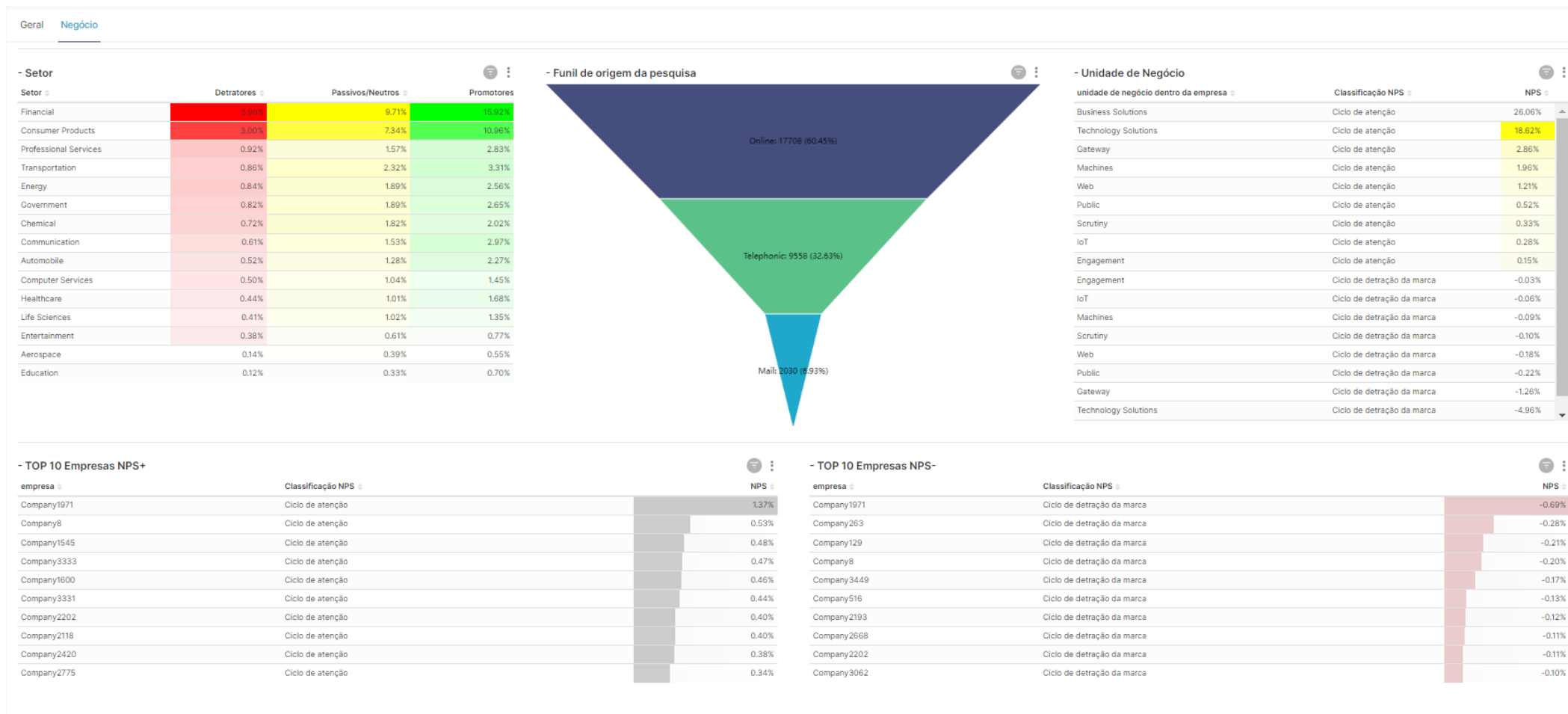
**arkhe** *Data Analysis*

**Exemplos**

# ADA – Exemplos - Eleições



# ADA – Exemplos - Vendas



# ADA – Exemplos - Consumidores

NPS Draft ☆



Geral Negócio

Net promoter Score

- Detratores

14.24%



- Passivos/Neutros

33.76%



- Promotores

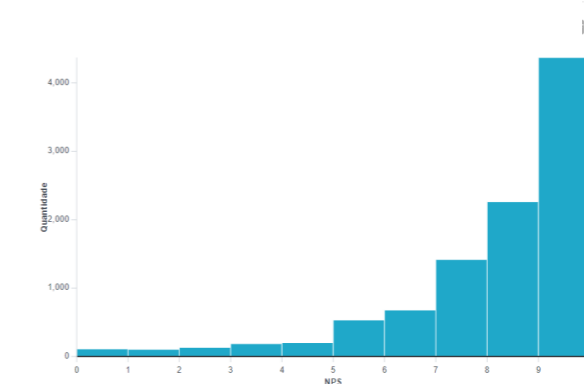
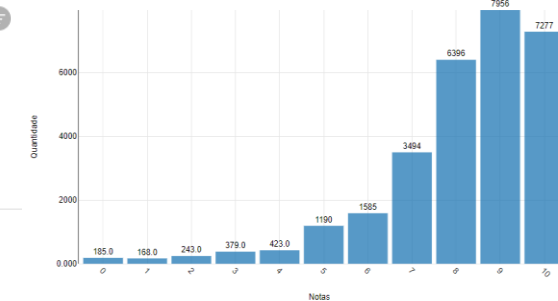
52.00%



- Net promoter Score Geral

37.75%

- Quantidade de respostas

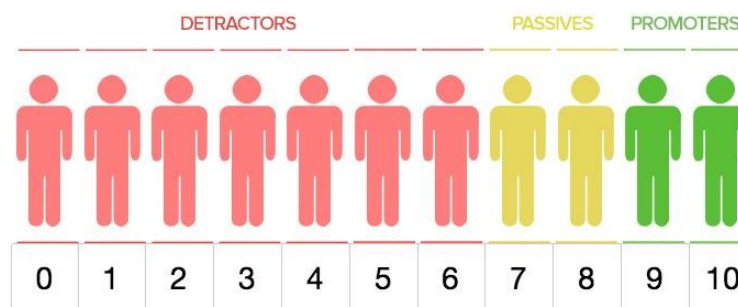


O NPS é o índice de satisfação que mede a lealdade dos seus consumidores através de uma única pergunta.

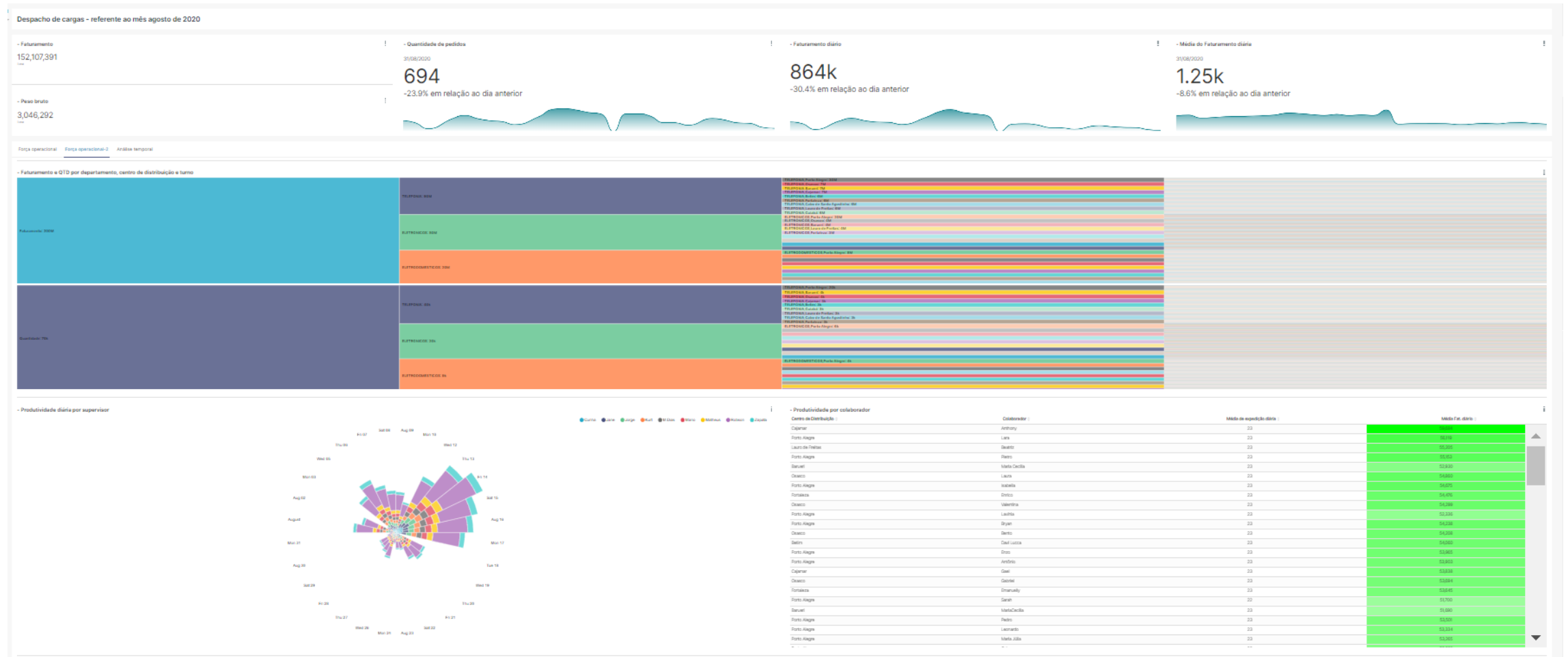
NPS é a abreviação do termo net promoter score, uma pontuação que tem por objetivo entender o nível de satisfação dos clientes.

A pergunta NPS consiste em uma escala de 0 a 10, onde o respondente avalia o grau de satisfação e a probabilidade de recomendar os produtos e serviços recebidos a alguém.

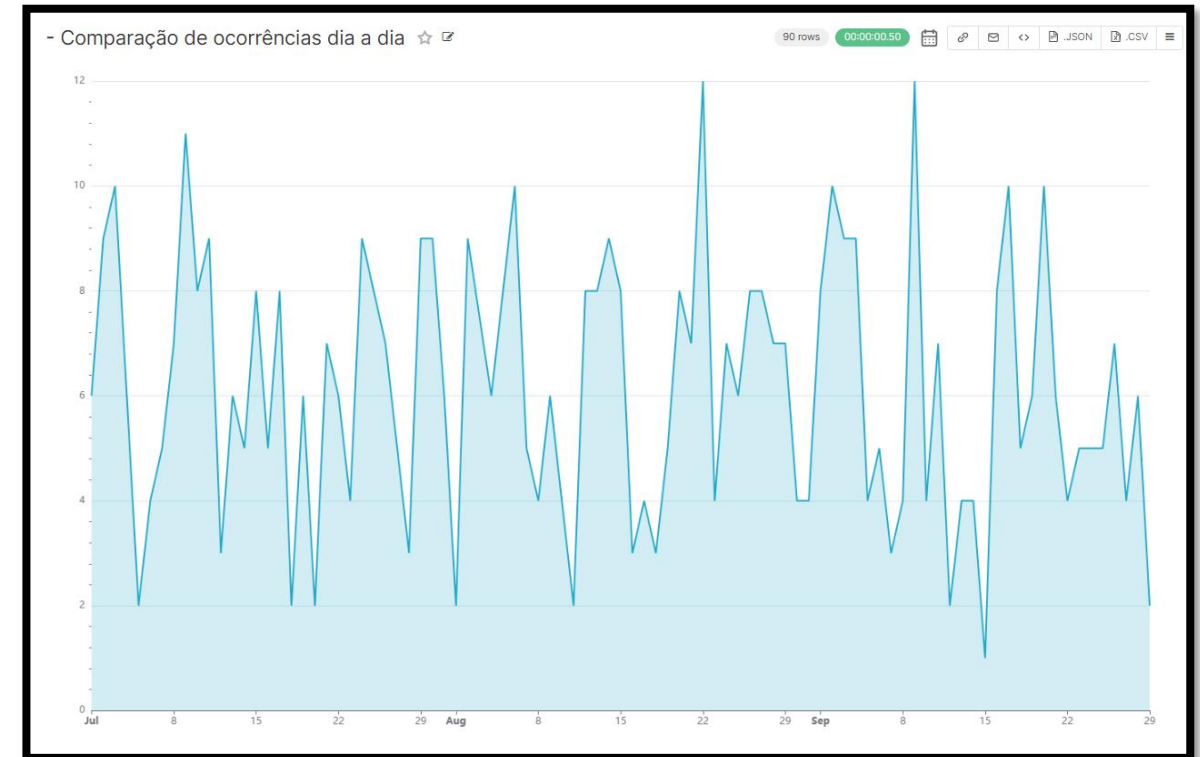
Então, de acordo com a avaliação NPS que dão, os entrevistados são classificados entre: [clientes detratores](#), [clientes passivos](#) e [clientes promotores](#).



# ADA – Exemplos – Logística



# ADA – Exemplos – Análise Preditiva







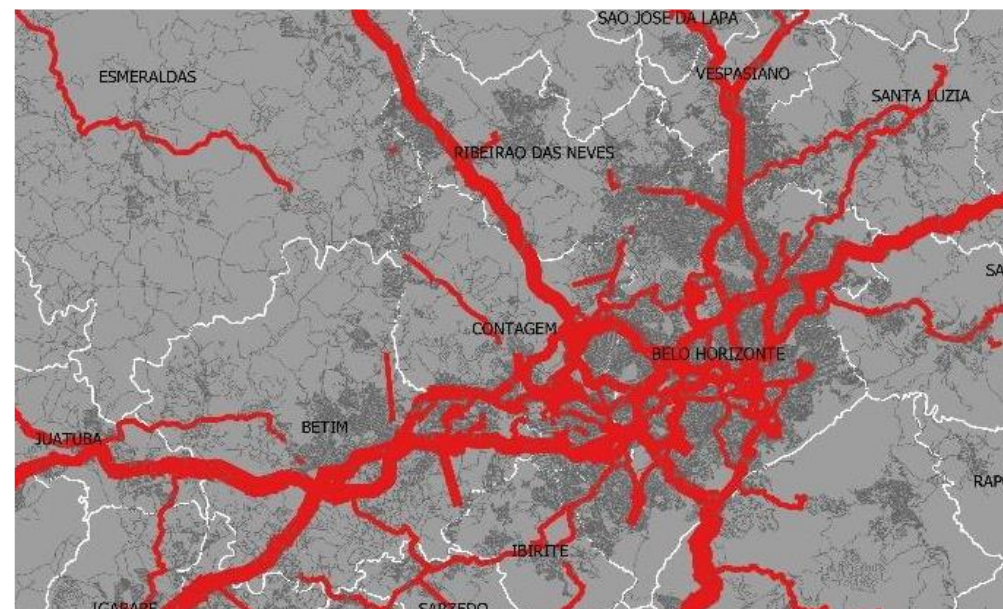
**Case**  
**Mobilidade**  
**Regional**

## RESUMO DO PROJETO

### Modelagem de simulações de redes de transportes e BI em tempo real

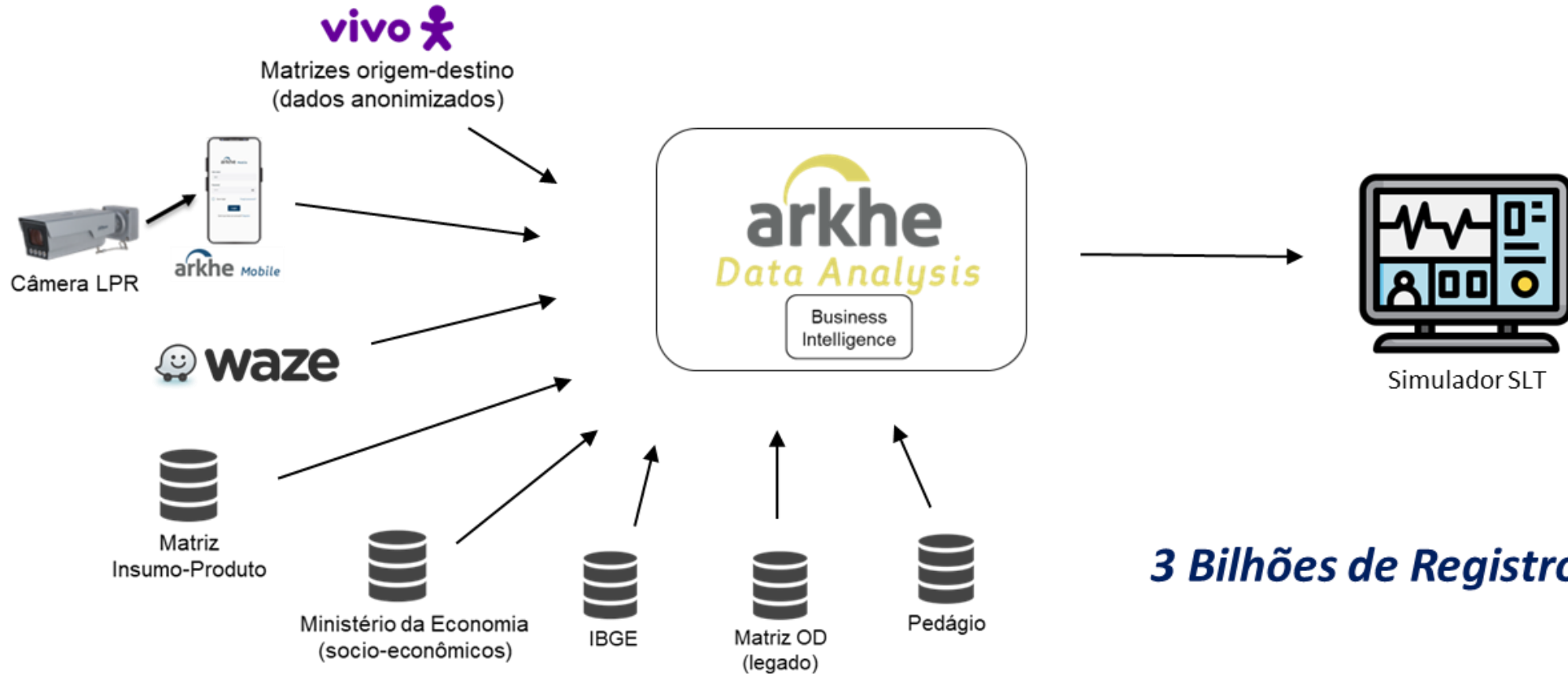
- A ATECH desenvolveu um modelo de simulação de redes de transportes utilizando dados de deslocamento de pessoas (telefonia móvel), dados de setores censitários (IBGE), dados socioeconômicos de renda e valores transportados de cargas (Min. da Economia), pedágios e radares (ARTESP) além de dados de modelos anteriores, infraestrutura de portos, aeroportos, hidrovias, ferrovias e estradas. Foram utilizados 12 meses de dados de todo o estado de São Paulo gerando **3 bilhões de registros!!!** Este modelo derivou modelos matemáticos, dezenas de dashboards em tempo real, integração com o Waze para monitoramento de congestionamento, integração com câmeras LPR e aplicativos móveis.

# arkhe *Data Analysis*



# Case Mobilidade Regional

SLT – Secretaria de Logística e Transportes do Estado de São Paulo



**3 Bilhões de Registros !!**

# Case Mobilidade Regional – SLT/DER

3

**BILHÕES DE  
REGISTROS!!!**

3000

**MILHÕES DE  
VIAGENS Origem -  
Destino**

9

**MESES**

150

**USUÁRIOS**

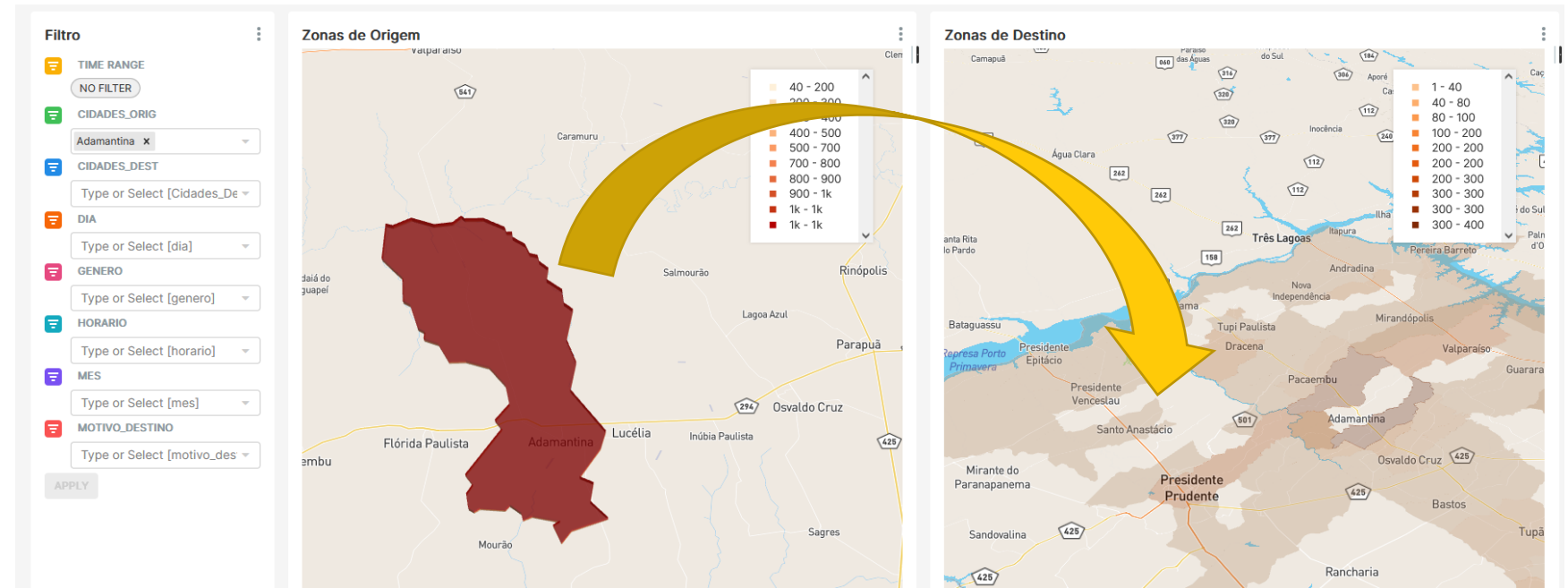
10

**PESSOAS**

# Case Mobilidade Regional – SLT/DER

## VIAGEM DE VEÍCULO:

- Viagem de veículo é uma transação que **POSSUI ORIGEM, DESTINO, TEMPO, DISTÂNCIA E MOTIVOS** definidos realizada por um veículo de um modo transportando **pessoa ou carga** identificados tradicionalmente por entrevistas ou rastreo associado a informações anexas transmitidas on-line.
- O conjunto de viagens entre uma origem e destino consiste em fluxo que somado a outros que compartilham um ponto comum em seus caminhos mínimos determinam o **volume que solicita um trecho rodoviário ou link**.

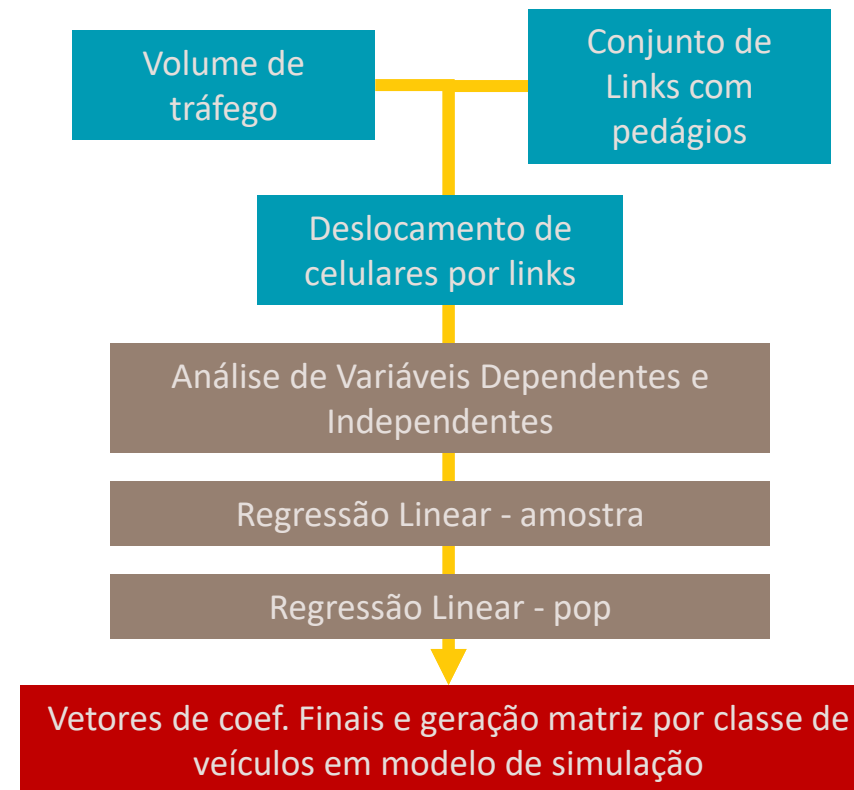


*Brasil dividido em 910 zonas, considerando viagens internas, de passagem, de e para SP*

## Case Mobilidade Regional – SLT/DER

- Determinação do tráfego de veículos nos pedágios por Regime de Atividade e classe de veículo ( $VV_P, CV, RA, ANO, MES, PD$ )**
- Determinação das matrizes de links selecionados de pedágio ( $ML_P$ ).**
- Determinação das quantidades de transações por atributos de usuários ( $QT_{RA, ANO, MES, PD, CR, SU, FE}$ ).**
- Processamento das transações de deslocamento de celulares e matrizes de links selecionados de cada pedágio para identificação do Traço da Matriz ( $Tr [ML_P, QT_{RA, ANO, MES, PD, CR, SU, FE}]$ ) referente a cada período, regime de atividade diária.**
- Identificação das variáveis dependentes (volumes de pedágio) e independentes (volumes de transações por atributo explicativo de usuário).**
- Aplicação de regressão sobre para a amostra de março e obtenção do vetor de coeficientes ( $CA_{CR, SU, FE}$ ).**
- Treinamento da regressão para o universo de dados dos demais meses de 2019 e 2020 e obtenção dos vetores finais de coeficientes ( $Cf_{CR, SU, FE}$ ).**
- Aplicação dos vetores  $Cf_s$  sobre os dados totais de telefonia de cada segmento de transações para geração das matrizes de veículos por segmento.**
- Obtenção da Matrizes por Classe de Veículos.**

## Metodologia de Geração de Matriz OD a partir de deslocamentos de celulares detectados por informações de CDR



$VVP, CV, RA, ANO, MES, PD = Tr [MLP, QTRA, ANO, MES, PD, CR, SU, FE]. CA_{CR, SU, FE}$

$$y_1 = \alpha_1 \cdot y_2 + \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_k \cdot x_k + u$$

Onde:

$$y_1 = VVP, RA, ANO, MES, PD$$

$$y_2 = QTRA, ANO, MES, PD, CR, SU, FE$$

$x_n =$  variáveis explicativas independentes

# Case Mobilidade Regional – SLT/DER

