SUCESU RS

GRUPO DE USUÁRIOS DE BI (GUBI)

DESMISTIFICANDO OS CONCEITOS DE BI: ANALYTICS, PREDITIVA, PRESCRITIVA, HISTÓRICA E MACHINE LEARNING

Evolução do analytics e do perfil dos profissionais de TI

Cassio Cervo

Cássio Socal Cervo

Administração de Empresas - PUCRS

Ciências Contábeis - PUCRS

MBA Gerenciamento de Projetos - FGV/Decision

MBA Big Data e Analytics – FGV-SP

20 anos de experiência em TI

Sicredi Gerente de Sistemas de BI, Risco e Modelagem

GUBI Grupo de Usuários Business Intelligence

Integrante do grupo de coordenadores do GUBI na SUCESU RS

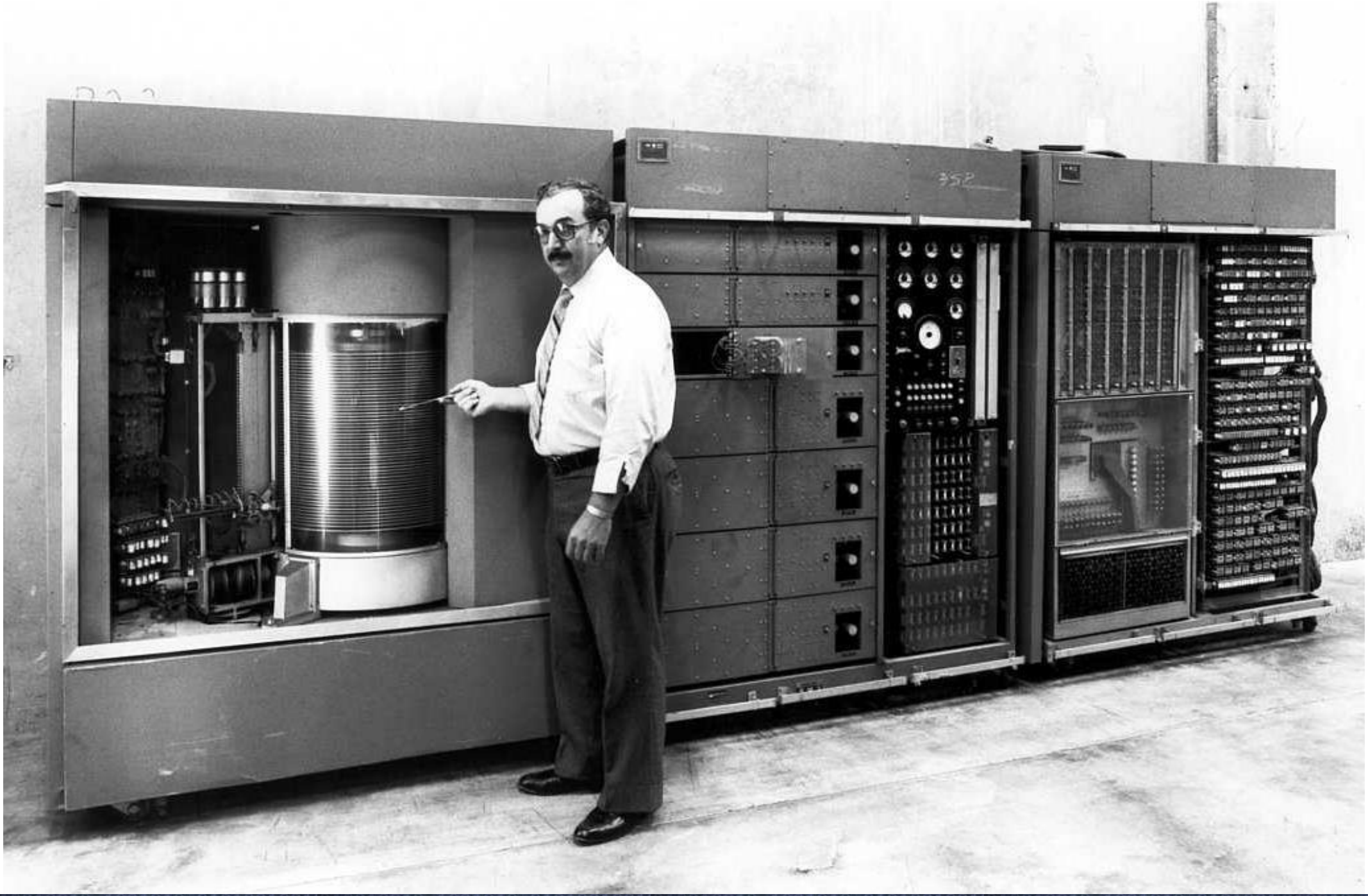
(Grupo de Usuários de Business Intelligence)

IBM
Data Processing
Division

PAA

IBM 305 RAMAC Disk System

Ele foi lançado em 1956 como parte do IBM 305 RAMAC, um computador desenvolvido pela IBM para o mercado corporativo, e podia armazenar impressionantes 5MB em 50 discos com 24 polegadas (60 cm) de diâmetro.



IBM
305 RAMAC

1956

Evolução do Analytics

1970

Mainframe OLAP

Express

Anos 70

Soluções de Armazenamento físicos e os Bancos de Dados.

Evolução do Analytics

Anos 70

Soluções de Armazenamento físicos e os Bancos de Dados.

Anos 80

Linguagens de programação,
ERPs,
termo BI,
analise de dados

Evolução do Analytics

Anos 70

Soluções de Armazenamento físicos e os Bancos de Dados.

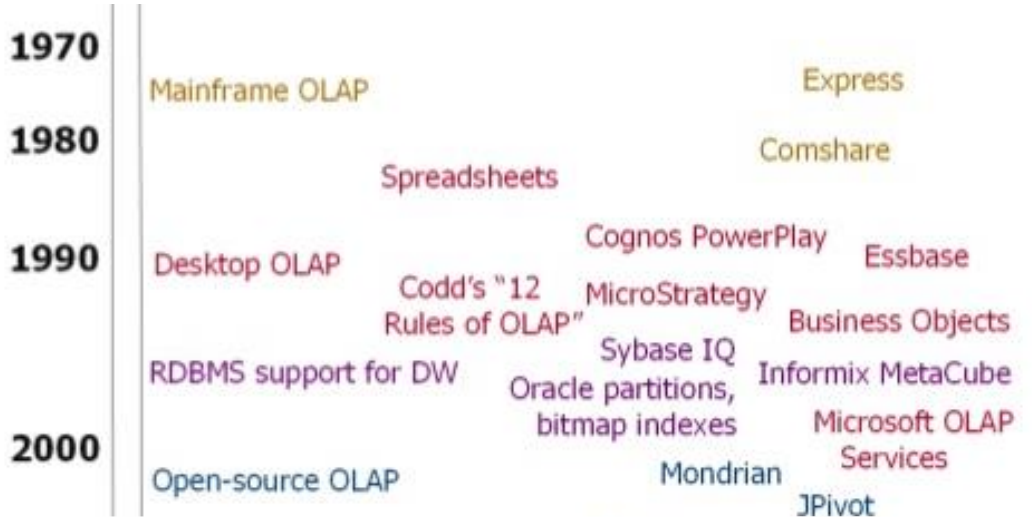
Anos 80

Linguagens de programação,
ERPs,
termo BI,
analise de dados

Anos 90

Construção dos Data Warehouse,
softwares de BI

Evolução do Analytics



Anos 70

Soluções de Armazenamento físicos e os Bancos de Dados.

Anos 80

Linguagens de programação,
ERPs,
termo BI,
analise de dados

Anos 90

Construção dos Data Warehouse,
softwares de BI

Anos 2000

DW consolidado
Desafio da fonte única da Informação
Grande volume de dados.
Evolução dos sistemas de BI
Ferramentas de ETL

Evolução do Analytics

Anos 70

Soluções de Armazenamento físicos e os Bancos de Dados.

Anos 80

Linguagens de programação,
ERPs,
termo BI,
analise de dados

Anos 90

Construção dos Data Warehouse,
softwares de BI

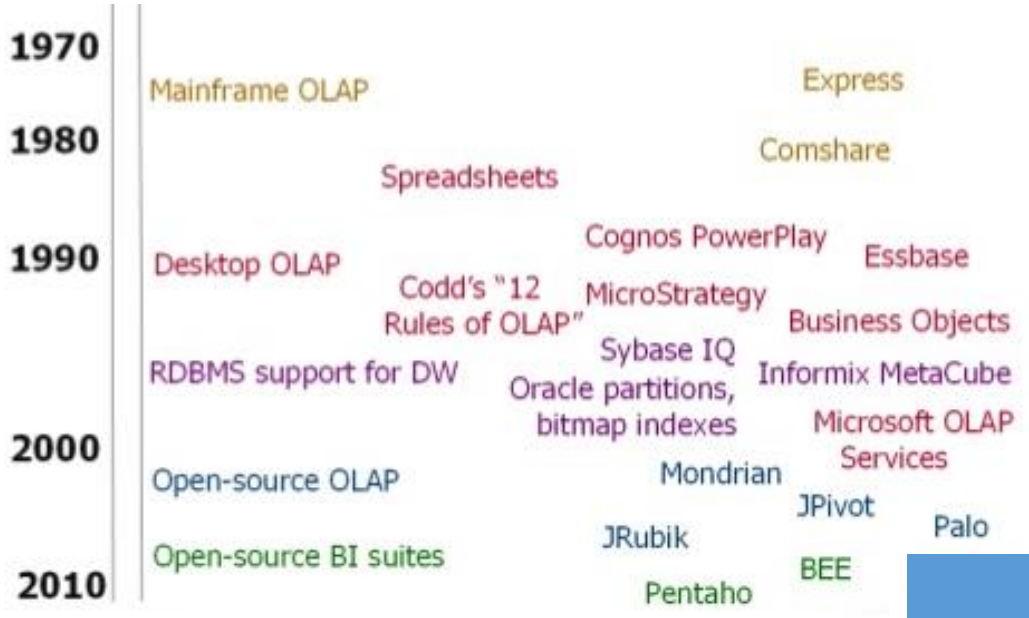
Anos 2000

DW consolidado
Desafio da fonte única da Informação
Grande volume de dados.
Evolução dos sistemas de BI
Ferramentas de ETL

Atualmente

Selfservice BI
Dados não estruturados
Text Mining
Hadoop
Spark
Mapeduce
Machine Learning
Redes Neurais
IoT
...

Evolução do Analytics



Anos 70

Soluções de Armazenamento físicos e os Bancos de Dados.

Anos 80

Linguagens de programação,
ERPs,
termo BI,
analise de dados

Anos 90

Construção dos Data Warehouse,
softwares de BI

Anos 2000

DW consolidado
Desafio da fonte única da Informação
Grande volume de dados.
Evolução dos sistemas de BI
Ferramentas de ETL

Atualmente

Selfservice BI
Dados não estruturados
Text Mining
Hadoop
Spark
Mapeduce
Machine Learning
Redes Neurais
IoT
...

Big Data

Business Intelligence = Big Data

Business Intelligence = Big Data ?

Business Intelligence = Big Data ?

"BI é o processo de ***coleta, organização, análise, compartilhamento*** *e monitoramento* ***de informações estruturadas*** *que oferecem* ***suporte a gestão*** *de negócios."*

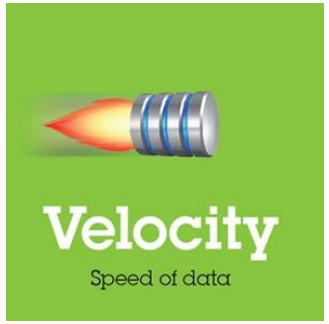
"Big Data é o termo que descreve o ***imenso volume de dados – estruturados e não estruturados*** *– que impactam os negócios no dia a dia."*

Definição da SAS

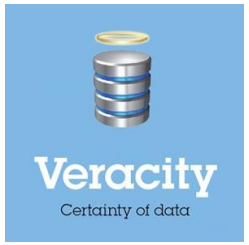
Big Data

"Big Data é o termo que descreve o ***imenso volume de dados*** *–* ***estruturados e não estruturados*** *– que impactam os negócios no dia a dia."* *Definição da SAS*

3Vs



4V –???

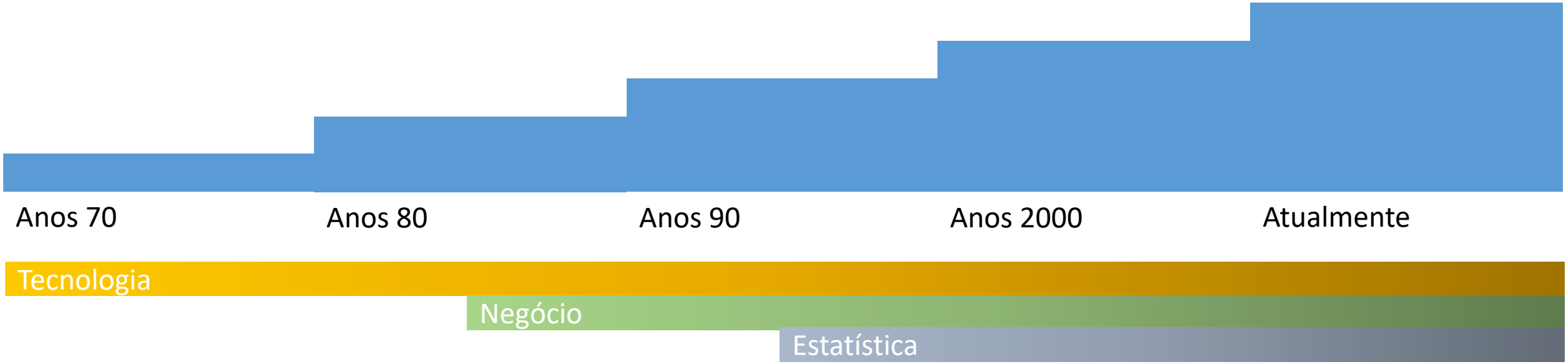


5V

2017 This Is What Happens In An Internet Minute

2018 This Is What Happens In An Internet Minute

Evolução do Analytics – Perfil



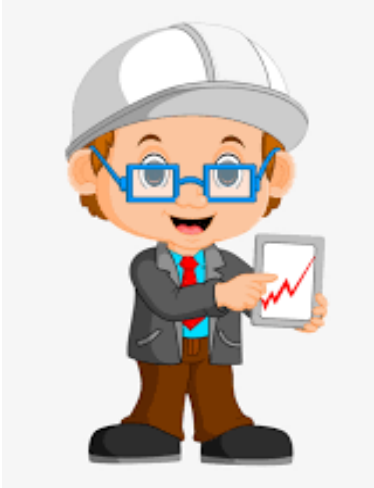
Evolução do Analytics – Perfil

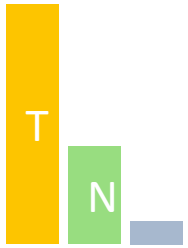
Tecnologia

Negócio

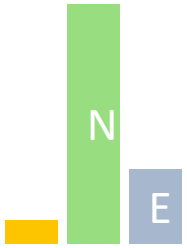
Estatística

Engenheiro de Dados

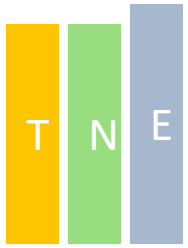




Analista de BI



Cientistas de Dados



"Analista de Sistemas, Desenvolvedor de ETL"

Isso é natural para você?

Microsoft SQL Server

Spark

hadoop

twitter

ORACLE PL/SQL

```
INSERT xml_document(?)

INTO Northwind.dbo.Employees (Employe    ,LastName,

xml_row_pattern('/root/row')

VALUES(
    xml_xpath('EmployeeID/text()','INTEGER'),
    xml_xpath('LastName/text()','VARCHAR'),
    xml_xpath('FirstName/text()','VARCHAR'),
    xml_xpath('Title/text()','VARCHAR'),
    xml_xpath('TitleOfCourtesy/text()','VARCHAR'),
    xml_xpath('BirthDate/text()','TIMESTAMP'),
    xml_xpath('HireDate/text()','TIMESTAMP'),
    xml_xpath('Address/text()','VARCHAR'),
    xml_xpath('City/text()','VARCHAR'),
    xml_xpath('Region/text()','VARCHAR'),
    xml_xpath('PostalCode/text()','VARCHAR'),
    xml_xpath('Country/text()','VARCHAR'),
    xml_xpath('HomePhone/text()','VARCHAR'),
    xml_xpath('Extension/text()','VARCHAR'),
    xml_xpath('Notes/text()','LONGVARCHAR'),
    xml_xpath('ReportsTo/text()','INTEGER'),
    xml_xpath('PhotoPath/text()','VARCHAR'))
```

SELECT | UPDATE

Name
northwind:1433
_test
Books
demo
FactBook
master
msdb
Northwind
dbo
Categories
CustomerCustomerDemo
CustomerDemographics
Customers
Employees
EmployeeTerritories
Order Details
Orders
Products
Region
Shippers
Suppliers
Territories
Alphabetical list of pro...
Category Sales for 1997
Current Product List
Customer and Supplier...
Invoices
Order Details Extended
Order Subtotals
Orders Qry
Product Sales for 1997

Oracle Data Integrator Studio 12c : EDWINTCAD_MUA_IP_F2

File Edit View Search ODI Tools Window Run Help

Designer | Operator | Topology

Projects
Packages
Cópia de EDWINTCAD_MU
Copy of EDWINTCAD_MUA
Copy of EDWINTCAD_MUA
Copy (2) of EDWINTCAD_I
Copy (4) of EDWINTCAD_I
EDWINTCAD_MUA
EDWINTCAD_MUA_AR
EDWINTCAD_MUA_AR_AC
D_MUA_AR_RE
D_MUA_CL
D_MUA_CL_CO
D_MUA_CL_CO
D_MUA_CL_GR
D_MUA_CL_INI
D_MUA_CL_PES
D_MUA_CL_PR
D_MUA_IP
D_MUA_IP_BAC
D_MUA_IP_CLI
D_MUA_IP_CO
D_MUA_IP_F1
D_MUA_IP_F2
D_MUA_IP_F3
D_MUA_IP_F3_
D_MUA_IP_OR

EDWINTCAD_MUA_IP_F2

Toolbox
All
Changed Data Capture
Event Detection
Files
Internet
Metadata
Oracle Data Integrator Objects
Plugins
SAP
Utilities
OdiAnt
OdiBeep
OdiEnterpriseDataQuality
OdiKillAgent
OdiOSCommand
OdiPingAgent
OdiPurgeLog
OdiReinitializeSeq
OdiRemoveTemporaryObjects
OdiStartLoadPlan
OdiStartOwbJob

prMOSTRA_VARIA...
vsIDV_VW_CL_IDV...
vsIDV_VW_CL_IDV...
vsIDV_VW_CL_CT...
vsIDV_VW_CL_GN...
vsIDV_VW_CL_CN...
vgTIMESTAMP_W...
vsIDV_VW_CL_MN...
vsCL_CGY_BNK_I...
vsCL_GND_INV
vsSICREDI_UNT_V...
vsSICREDI_UNT_V...
vsSZ_TP_DTL_VW...
vsIDV_VW_CL_MN...
vsCL_CGY_BNK_I...
vsCL_GND_INX
prMOSTRA_VARIA...
vsSICREDI_UNT_V...
vsSICREDI_UNT_V...
vsSZ_TP_DTL_VW...
mpIDV_040_T$TO...
mpIDV_030_INCR...
mpIDV_010_MAIN...
vsSZ_TP_DTL_VW...
vsSZ_TP_DTL_VW...
prMOSTRA_VARIA...
mpFI_CCROAGBA
mpSICREDI_UNT_...
mpSICREDI_UNT_...
mpSICREDI_UNT_...
mpSICREDI_UNT_...

Properties
Select an element in the diagram to display its properties

Overview | Diagram

Engenheiro de Dados

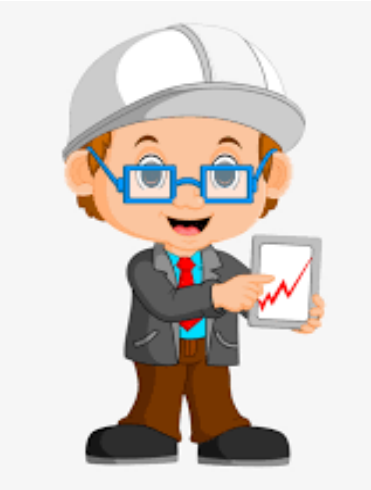
*O **Data Engineer** é então **responsável por garantir que os dados estejam disponíveis** para a análise de forma segura.*

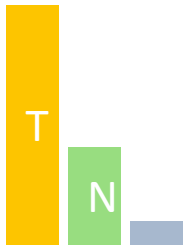
Arquitetura de Soluções

Infra estrutura

Ferramentas

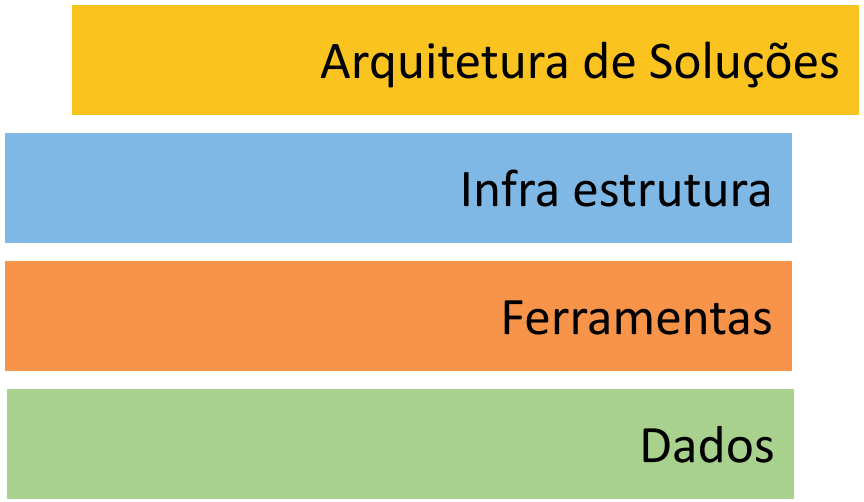
Dados



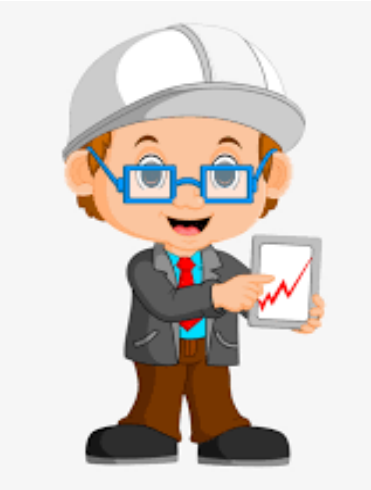


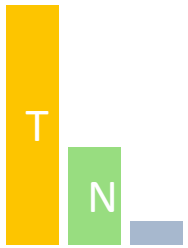
Engenheiro de Dados

*O **Data Engineer** é então **responsável por garantir que os dados estejam disponíveis** para a análise de forma segura.*



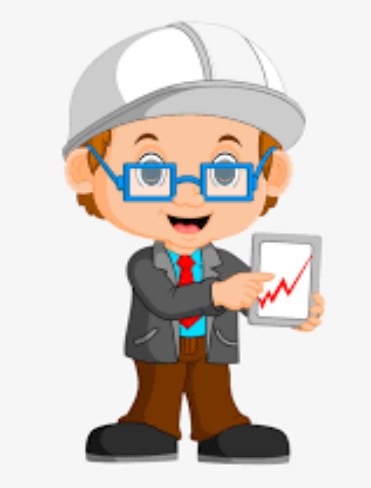
- *Modelagem de Serviços*
- Containeres de micro-serviços
- Jobs
- APIs

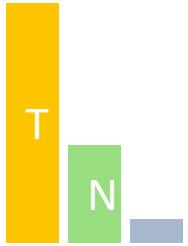




Engenheiro de Dados

*O **Data Engineer** é então **responsável por garantir que os dados estejam disponíveis** para a análise de forma segura.*





Arquitetura de Soluções

Infra estrutura

Ferramentas

Dados

- Manter ambiente de alta disponibilidade
- Clouding computer
- Hadoop/Spark
- Data lakes
- DW/DM

Engenheiro de Dados

*O **Data Engineer** é então **responsável por garantir que os dados estejam disponíveis** para a análise de forma segura.*

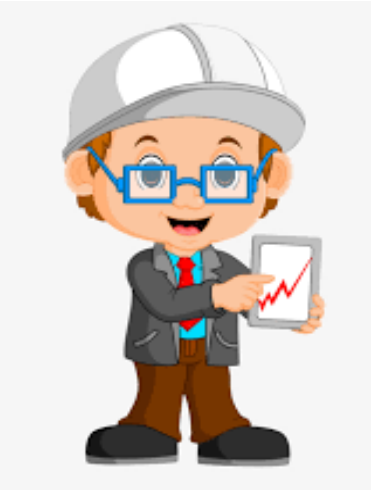
Arquitetura de Soluções

Infra estrutura

Ferramentas

Dados

- Bancos de Dados
- Ferramentas de Extração
- Ferramentas de Qualidade
- Ferramentas de Distribuição

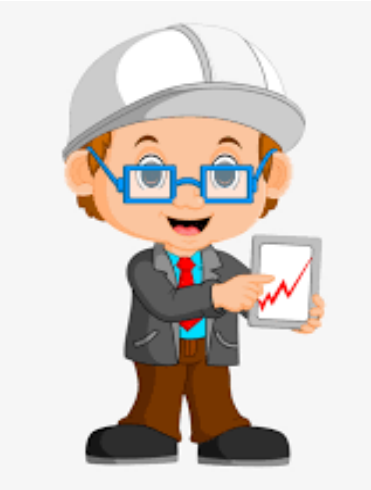


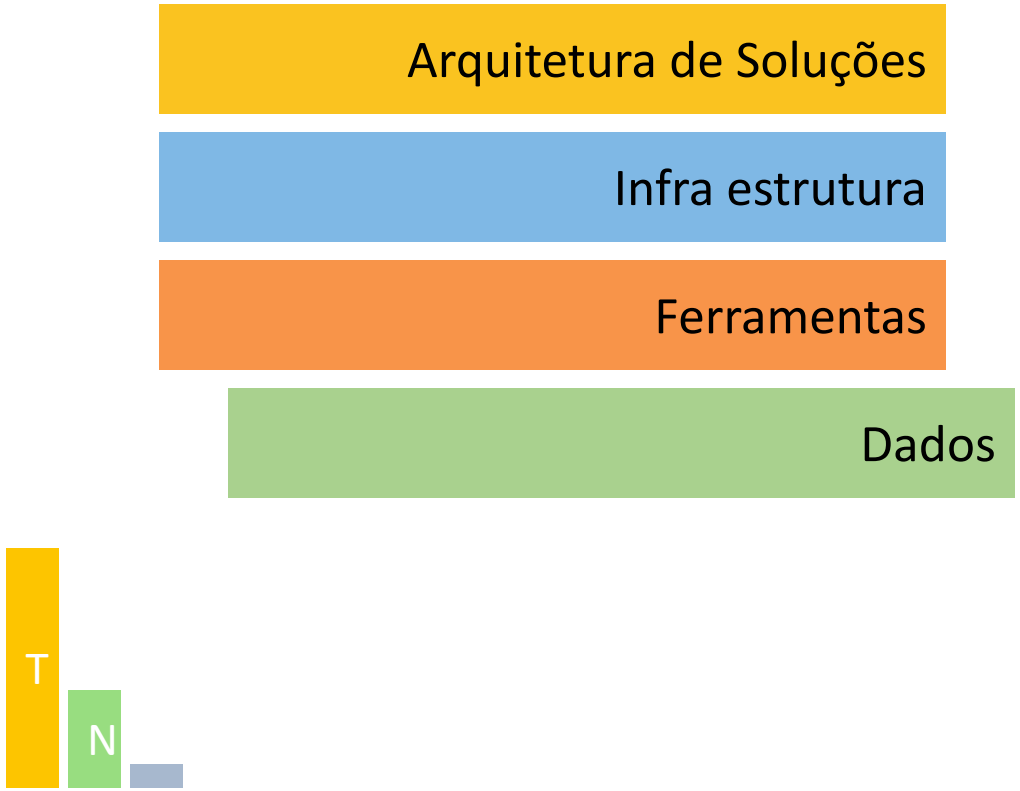
T

N

Engenheiro de Dados

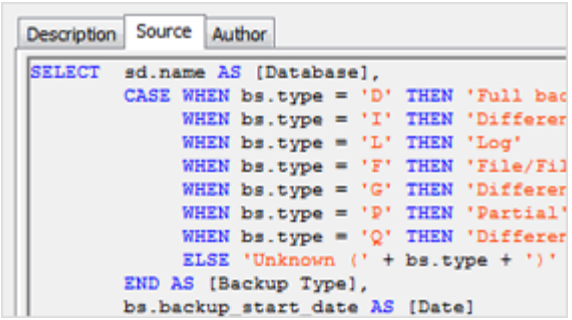
*O **Data Engineer** é então **responsável por garantir que os dados estejam disponíveis** para a análise de forma segura.*

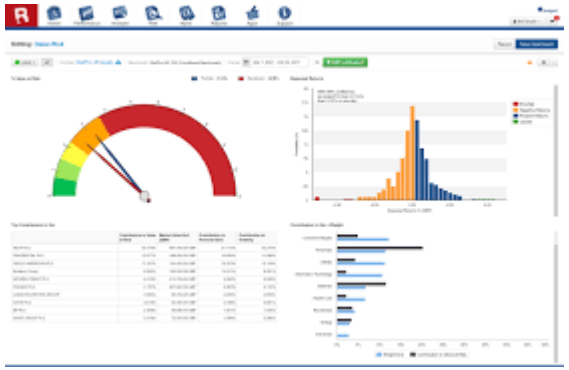




- Coletar os dados
- Modelar os dados
- Transformar os dados
- Armazenar os dados
- Disponibilizar os dados

Isso é natural para você?



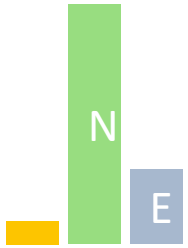


Analista de BI

*O **Analista de BI** é quem pensa nos dados com o viés **de distribuição das informação** de forma a **auxiliar a tomada de decisão** e **acompanhar a operação da empresa.***

Visão de Negócio

Tecnologia



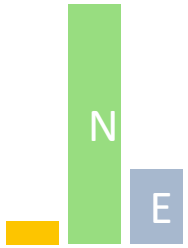
Analista de BI

*O **Analista de BI** é quem pensa nos dados com o viés **de distribuição das informação** de forma a **auxiliar a tomada de decisão** e **acompanhar a operação da empresa.***

Visão de Negócio

Tecnologia

- KPIs de negócio
- Conhecimento do planejamento estratégico
- Conhecimento do segmento da empresa
- Conhecimento da finanças da empresa
- Conhecimentos mercadológicos
- Conhecimentos detalhado dos produtos e serviços



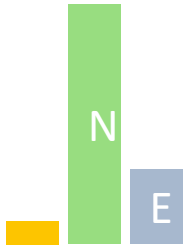
Analista de BI

*O **Analista de BI** é quem pensa nos dados com o viés **de distribuição das informação** de forma a **auxiliar a tomada de decisão** e **acompanhar a operação da empresa.***

Visão de Negócio

Tecnologia

- Fontes de dados disponíveis
- Qualidade das informações disponíveis
- Ferramentas e suas capacidade
- Selfservice BI



Isso é natural para você?

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

θ_0

$$h_\Theta(x) \in \mathbb{R}^K \quad (h_\Theta(x))_i = i^{th} \text{ output}$$

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log(h_\Theta(x^{(i)}))_k + (1-y_k^{(i)}) \log(1-(h_\Theta(x^{(i)}))_k)\right]$$

$$\boxed{+\frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{ji}^{(l)})^2}$$

$\Theta_{ji}^{(l)}$

$\Theta_{i0}^{(2)} x_0 + \Theta_{i1}^{(2)} x_1 + \dots$

a_0

$i = 0 \dots s_l$

y_k

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\vec{V}_2^1 = A_1 . C_o . \vec{V}_1^1 + A_2 . C_1 . \left[\frac{\overrightarrow{MP}^1 - \vec{S}_1^1}{\Delta t}\right] + A_3 . C_2 . \left[\frac{\overrightarrow{MB}_1 - \vec{S}_1^1}{\Delta t}\right]$$

$$\vec{V}_2^1 = 0{,}4 . 1 . \begin{bmatrix} 1{,}39 \\ 1{,}31 \end{bmatrix} + 0{,}3 . 2{,}5 . \begin{bmatrix} 2{,}39 & 2{,}39 \\ & - \\ 2{,}31 & 2{,}31 \end{bmatrix} + 0{,}7 . 1{,}5 . \begin{bmatrix} 2{,}39 & 2{,}39 \\ & - \\ 2{,}31 & 2{,}31 \end{bmatrix}$$

$$\vec{V}_2^1 = \begin{bmatrix} 0{,}55 \\ 0{,}52 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\boxed{\vec{V}_2^1 = \begin{bmatrix} 0{,}55 \\ 0{,}52 \end{bmatrix}}$$

$$\vec{S}_2^1 = \vec{S}_1^1 + \vec{V}_2^1 . \Delta t$$

$$\vec{S}_2^1 = \begin{bmatrix} 2{,}39 \\ 2{,}31 \end{bmatrix} + \begin{bmatrix} 0{,}55 \\ 0{,}52 \end{bmatrix} . 1$$

$$\boxed{S_2^1 = \begin{bmatrix} 2{,}94 \\ 2{,}83 \end{bmatrix}}$$

$$f_2^1(x,y) = x^2 + y^2 - 6 . x - 6 . y + 20$$

$$f_2^1(x,y) = 8{,}64 + 8{,}00 - 17{,}64 - 16{,}98 + 20$$

$$\boxed{f_2^1 = 2{,}02}$$

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad m = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$m \times v = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_{11} \cdot x + a_{12} \cdot y + a_{13} \cdot z \\ a_{21} \cdot x + a_{22} \cdot y + a_{23} \cdot z \\ a_{31} \cdot x + a_{32} \cdot y + a_{33} \cdot z \end{pmatrix}$$

$$m = \frac{n\sum_{i=1}^{n} iD_i - (\sum_{i=1}^{n} i)(\sum_{i=1}^{n} D_i)}{n\sum_{i=1}^{n} i^2 - (\sum_{i=1}^{n} i)^2}$$

$$b = \frac{(\sum_{i=1}^{n} D_i)(\sum_{i=1}^{n} i^2) - (\sum_{i=1}^{n} i)(\sum_{i=1}^{n} iD_i)}{n\sum_{i=1}^{n} i^2 - (\sum_{i=1}^{n} i)^2}$$

$$21\left(\frac{m-2}{3}+1\right) = \left(\frac{2m}{7}\right)21$$

$$21\left(\frac{m-2}{3}\right) + 21(1) = \left(\frac{2m}{7}\right)21$$

$$7(m-2) + 21 = (2m)(3)$$

Cientista de Dados

*"Uma **nova geração** de **especialistas analíticos** que têm as habilidades técnicas para **resolver problemas complexos** – e a **curiosidade** de explorar quais são os problemas que precisam ser resolvidos."*

Definição da SAS

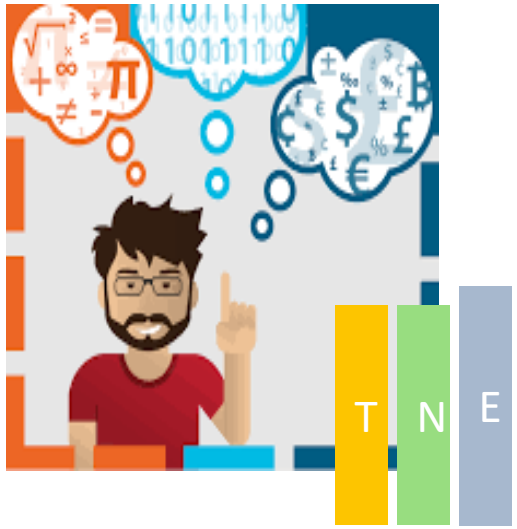
Funções típicas dos cientistas de dados

Trabalhar grandes quantidades de **dados**, buscado **padrões e tendências**, para **solucionar problemas de negócio**

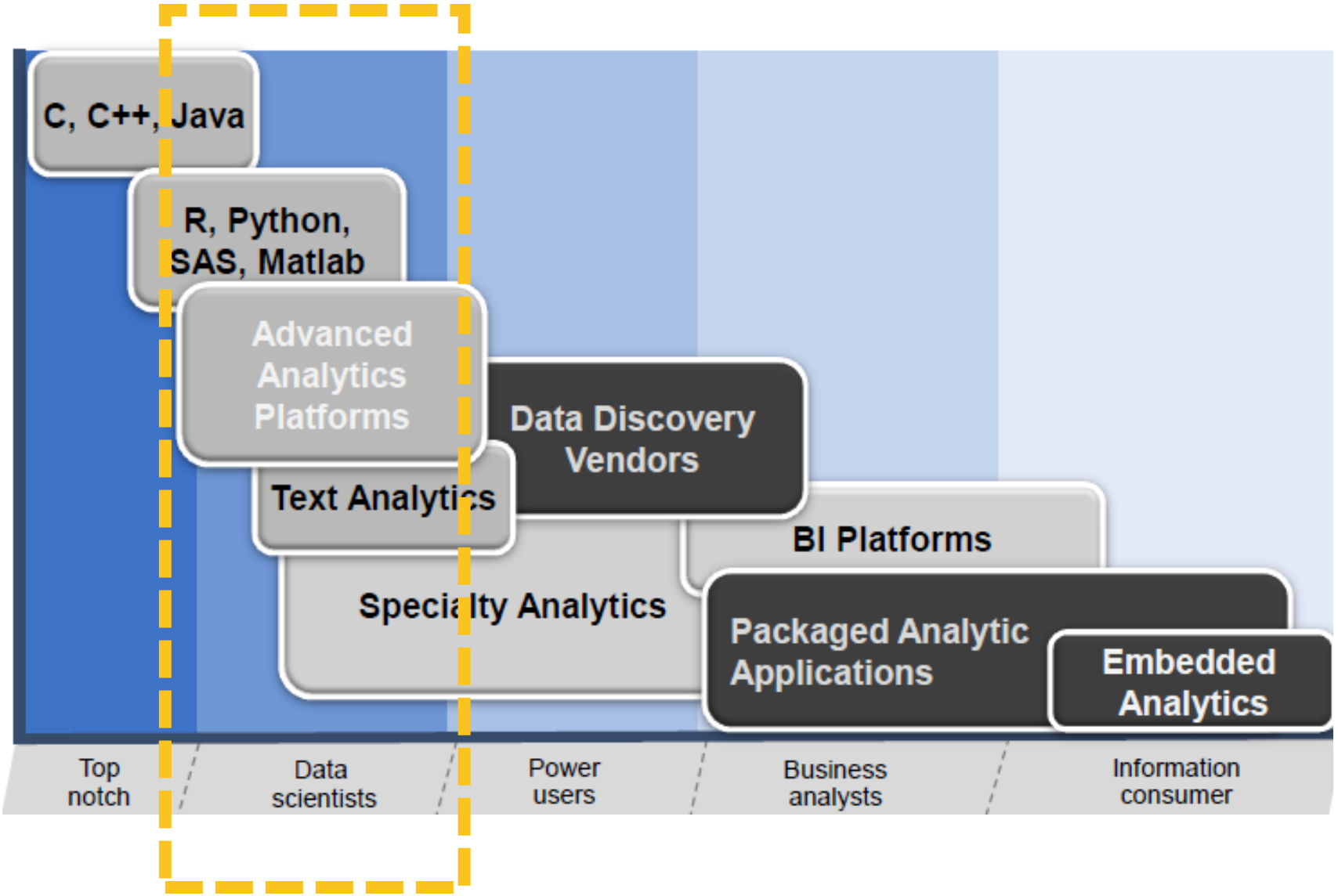
Trabalhar com uma variedade de **linguagens de programação** e ferramentas **tecnológicas**

Conhecer **o negócio e comunicar-se** com toda a empresa

Ter uma **sólida** compreensão de **estatísticas**, incluindo testes estatísticos e distribuições.



Cientista de Dados



Cientista de Dados

Negócio

Tecnologia

Cientistas de Dados

Estatística

Cientista de Dados - Estatística

Inferência e Estatística

Séries Temporais

Geoanálise e Estatística Espacial

Cientista de Dados - Estatística

*Inferência Estatística é um **conjunto de técnicas** que objetiva estudar uma **população** através de **evidências fornecidas por uma amostra**.*

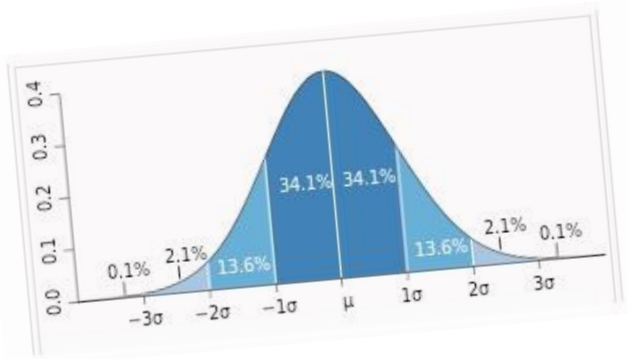
Inferência e Estatística

Séries Temporais

Geoanálise e Estatística Espacial

- Estatística básica (Variância, Desvio Padrão, Média, Moda, Mediana, ...)
- Diagrama de Dispersão
- Regra empírica
- Distribuição Qui-Quadrado
- Tipos de testes de hipóteses
- Regressão linear simples
- Regressão linear múltipla
- Analise de resíduos
- Cadeia de Markov
- Page Rank
- Algoritmos de buscas otimizadas
- Simulated Annealing
- Algoritmos Genéticos
- Particle Swarm Optimization PSO
- Matrizes Esparsas
- Sistemas de recomendação

$$Y_i = \alpha + \beta X_i + \mu_i$$



Cientista de Dados - Estatística

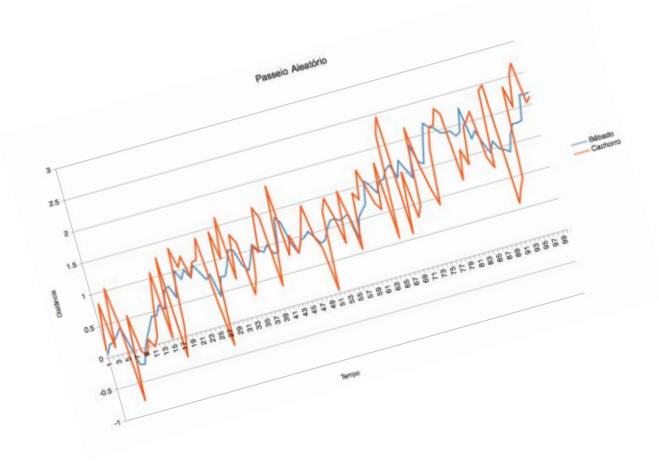
*Uma série temporal é um **conjunto de observações** ordenadas no tempo, não necessariamente igualmente espaçadas, que **apresentam dependência serial**, isto é, dependência entre instantes de tempo.*

Inferência e Estatística

Séries Temporais

Geoanálise e Estatística Espacial

- Procedimentos estatísticos de previsão
- Análise de autocorrelação
- Redes neurais na previsão de séries temporais
- Estacionariedade de uma série
- Avaliação do desempenho das previsões



Cientista de Dados - Estatística

*O objetivo das técnicas de Análise Espacial é descrever os **padrões existentes** nos **dados espaciais** e estabelecer, preferencialmente de forma quantitativa, os relacionamentos entre as diferentes **variáveis geográficas**.*

Inferência e Estatística

Séries Temporais

Geoanálise e Estatística Espacial

- Análise Espacial e Geoprocessamento
- Análise de Eventos Pontuais
- Análise de Superfícies por Geoestatística Linear
- Análise de Superfícies por Geoestatística Indicadora
- Análise de Dados de Área

Cientista de Dados - Tecnologia

- Bancos de Dados
- Modelagem de dados
- Programação
- Ferramenta de Visualização
- Machine Learning
- Big Data

Cientista de Dados - Tecnologia

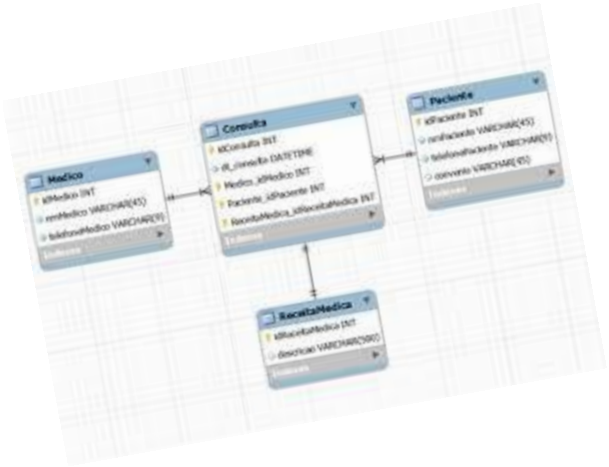
- Bancos de Dados
- Modelagem de dados
- Programação
- Ferramenta de Visualização
- Machine Learning
- Big Data

- Funcionamento dos bancos de dados mais comuns
  (Oracle, SQL Server, Postgree, MySQL, Teradata, ...)
- Banco de dados colunares
- Banco de dados in memory
- Estrutura de Data Warehouse
- Banco de Dados noSQL (MongoDB, Cassandra)

Cientista de Dados - Tecnologia

- Bancos de Dados
- Modelagem de dados
- Programação
- Ferramenta de Visualização
- Machine Learning
- Big Data

- Modelos conceitual
- Modelos Lógico
- Modelos Físicos
- Modelagem Multidimensional (Star Schema, Snow Flake)
- Atributos
- Relacionamentos
- Data Mining
- Versionamento
- Historização
- Expurgo de dados



Cientista de Dados - Tecnologia

- Bancos de Dados
- Modelagem de dados
- Programação
- Ferramenta de Visualização
- Machine Learning
- Big Data

- Lógica de Programação
- R (R studio)
- Python
- SQL
- Dplyr
- SAS
- Java
- C, C++, C#

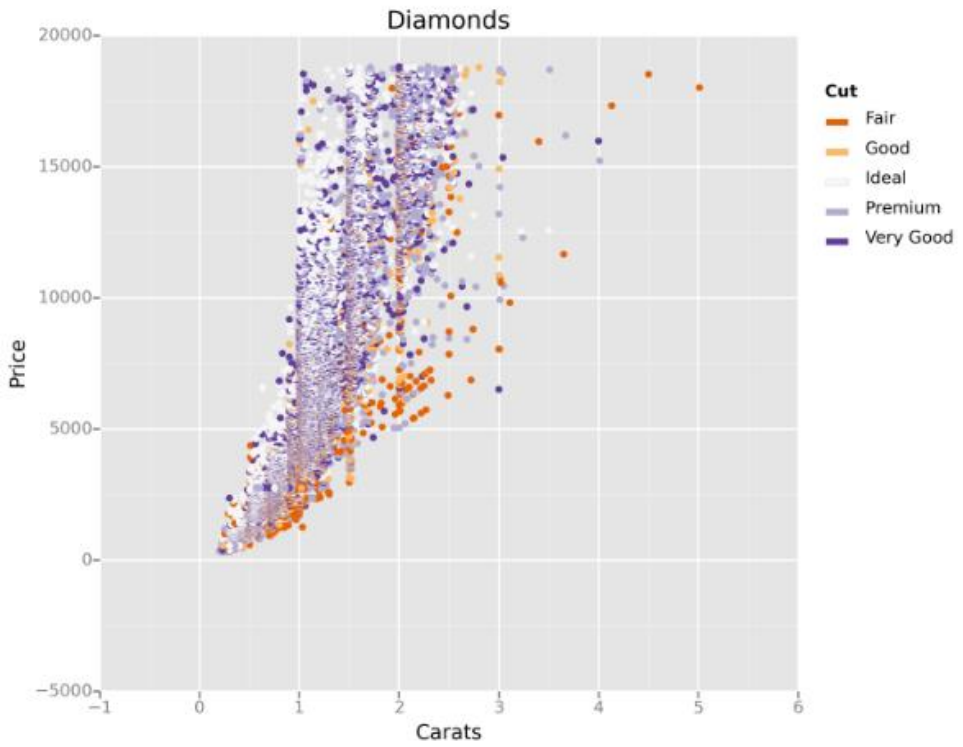
Cientista de Dados - Tecnologia

- Bancos de Dados
- Modelagem de dados
- Programação
- Ferramenta de Visualização
- Machine Learning
- Big Data

- GGPlot
- Tableau
- QlikView
- Pentaho
- Excel Power Pivot
- Microstrategy
- Oracle
- IBM
- PowerBI

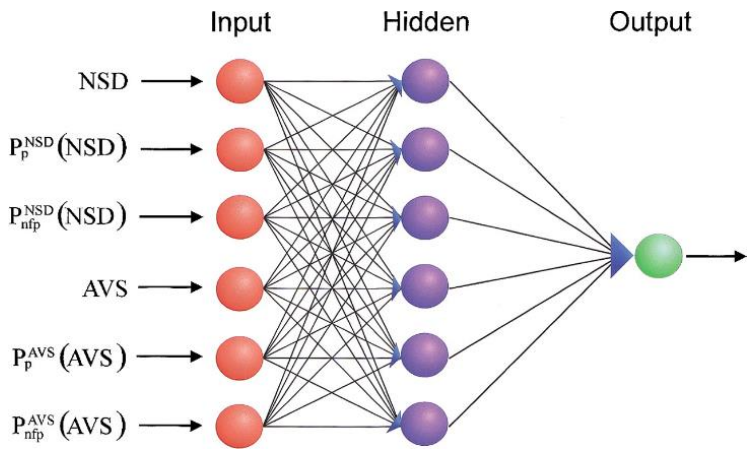
ggplot is fun

```
ggplot(diamonds, aes(x='carat', y='price', color='cut')) +\
    geom_point() +\
    scale_color_brewer(type='diverging', palette=4) +\
    xlab("Carats") + ylab("Price") + ggtitle("Diamonds")
```



Cientista de Dados - Tecnologia

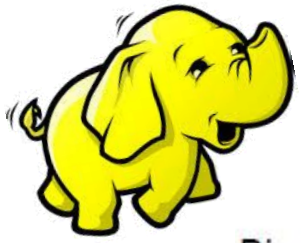
- Bancos de Dados
- Modelagem de dados
- Programação
- Ferramenta de Visualização
- Machine Learning
- Big Data



- Redes Neurais Artificiais (RNAs)
- Inteligência artificial
- Internet das Coisas (IoT)
- IBM Watson Analytics

Cientista de Dados - Tecnologia

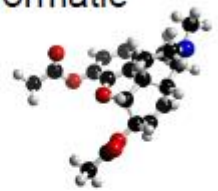
- Bancos de Dados
- Modelagem de dados
- Programação
- Ferramenta de Visualização
- Machine Learning
- Big Data
  - Hadoop
  - Spark
  - MapReduce
  - Hive



Biggest Neural Net So Far

kaggle

Merck's Cheminformatic



IBM

Google

word2vec

Android Speech Recognition

AlchemyAPI™

blackcloud

Cientista de Dados – Negócio

Análise Setorial

Análise Financeira

Análise Mercadológica

Pessoas

Decisões Empresariais e Raciocínio Analítico

Cientista de Dados – Negócio

Análise Setorial

Análise Financeira

Análise Mercadológica

Pessoas

Decisões Empresariais e Raciocínio Analítico

- Conhecer o setor
- Macroeconomia e Cenário político
- Leis e Normas
- Conhecer a cadeia produtiva
- Produtos, serviços e tecnologias
- Ciclo de vida setorial

Cientista de Dados – Negócio

Análise Setorial

Análise Financeira

Análise Mercadológica

Pessoas

Decisões Empresariais e Raciocínio Analítico

- Estrutura de Capital das empresas
- CMV
- Controladoria Gerencial
- Gestão de Custos
- Margem de contribuição dos produtos
- Alavancagem financeira
- Mercado de Capitais
- Matemática Financeira

Cientista de Dados – Negócio

Análise Setorial

Análise Financeira

Análise Mercadológica

Pessoas

Decisões Empresariais e Raciocínio Analítico

- Segmentação
- Os produtos e o mercado
- Pontos de vendas
- Perfil dos consumidores
- Análise SWOT
- Preços
- Campanhas publicitárias
- Direcionadores Estratégicos

Cientista de Dados – Negócio

Análise Setorial

Análise Financeira

Análise Mercadológica

Pessoas

Decisões Empresariais e Raciocínio Analítico

- *User Experience*
- Praxeologia*
- Psicologia do consumo

- Gestão de Pessoas nas empresas

Cientista de Dados – Negócio

Análise Setorial

Análise Financeira

Análise Mercadológica

Pessoas

Decisões Empresariais e Raciocínio Analítico

- Teoria dos Jogos
- Teoria de jogos combinatórios
- Teoria das Restrições
- Processo decisórios nas organizações

Cientista de Dados

Tecnologia

- Bancos de Dados
- Modelagem de dados
- Programação
- Ferramenta de Visualização
- Machine Learning
- Big Data

Negócio

- Análise Setorial
- Análise Financeira
- Análise Mercadológica
- Pessoas
- Decisões Empresariais e Raciocínio Analítico

Estatística

- Inferência Estatística
- Séries Temporais
- Geoanálise e Estatística Espacial
- Modelagem Estatística Avançada e Preditivas

Landscape de Big Data 2016

Infrastructure

Hadoop On-Premise
cloudera, Hortonworks, MAPR, Pivotal, IBM InfoSphere, splice machine, bluedata, jethro

Hadoop in the Cloud
amazon web services, Microsoft Azure, Google Cloud Platform, IBM InfoSphere, CAZENA, TREASURE DATA, altiscale, Qubole, xplenty

Spark
databricks, GridGain, TACHYON NEXUS

Cluster Services
amazon web services, kubernetes, HPCC SYSTEMS, docker, MESOSPHERE, CoreOS, pepperdata, StackIQ

NoSQL Databases
amazon DynamoDB, Google Cloud Platform, Microsoft Azure, ORACLE, MarkLogic, mongoDB, DATASTAX, AEROSPIKE, Couchbase, SequoiaDB, redislabs, influxdata

NewSQL Databases
SAP HANA, Clustrix, Pivotal, paradigm4, memsql, NUODB, MariaDB, VOLTDB, citusdata, deepdb, Trafodion, Cockroach LABS

Graph Databases
neo4j, GIRAPH, OrientDB, InfiniteGraph

MPP Databases
TERADATA, VERTICA, NETEZZA, kognitio, dremio

Cloud EDW
amazon web services, Google Cloud Platform, Microsoft Azure, Pivotal, snowflake, WATERLINE DATA, Infoworks

Data Transformation
alteryx, TRIFACTA, tamr, Paxata, StreamSets, Alation

Data Integration
informatica, MuleSoft, snapLogic, BedrockData

Management / Monitoring
New Relic, APPDYNAMICS, amazon web services, actifio, Numerify, splunk, DATADOG, rocana, Anodot

Security
TANIUM, illumio, CODE42, DataGravity, CipherCloud, VECTRA, sqrrl, BlueTalon

Storage
amazon web services, Google Cloud Platform, Microsoft Azure, panasas, nimblestorage, Qumulo

App Dev
apigee, CASK, Keen IO, Typesafe, CONCURRENT

Crowd-sourcing
amazon mechanical turk, CrowdFlower, WorkFusion

Analytics

Analyst Platforms
Palantir, AYASDI, Quid, enigma, Digital Reasoning, ORBITAL INSIGHT

Analytics Platforms
Microsoft, guavus, Datameer, inter|ana

Data Science Platforms
context relevant, CONTINUUM ANALYTICS, DataRobot, Alpine DATA LABS, plotly, ADATAO, MODE, dataiku, nutonian, DOMINO, sense, yhat, ALGORITHMIA

Visualization
tableau, Google Cloud Platform, Roambi, ZOOMDATA, Qlik, CHARTIO

BI Platforms
Power BI, amazon web services, DOMO, salesforce, Wave Analytics, GoodData, birst, platfora, kyvos insights, looker, atscale, ARCADIA, SISENSE

Statistical Computing
sas, SPSS, MATLAB

Log Analytics
splunk, sumologic, kibana, CLOUD PHYSICS, loggly

Social Analytics
NETBASE, DATASIFT, tracx, bitly, synthesio, bottlenose, simplereach

Real-Time
amazon web services, METAMARKETS, confluent, DATATORRENT, dataArtisans

Machine Learning
Azure Machine Learning, amazon web services, H2O.ai, Dato, SKYTREE, DATARPM, rapidminer, deepsense.io, ViSENZE, PredictionIO, glowfi.sh

Speech & NLP
NarrativeScience, api.ai, NUANCE, Gridspace, semanticmachines, cortical.io, maluuba, MindMeld, iDIBON, yseop

Horizontal AI
IBM Watson, Cortana, sentient technologies, VIV, vicarious, nara, Numenta, SI SCALED INFERENCE, HyperScience, MetaMind, clarifai, DEXTRO, Geometric Intelligence

Search
hp, Autonomy, ORACLE ENDECA, EXALEAD, Lucidworks, elastic, ThoughtSpot, MAANA, swiftype, Algolia, SINEQUA

Data Services
uo, OPERA SOLUTIONS, Mu Sigma DO THE MATH, DATASCIENCE, SILICON VALLEY DATA SCIENCE, kaggle, datascope, DataKind

For Business Analysts
OrigamiLogic, ClearStory, CIRRO, import.io

SMB / Commerce
Google Analytics, AMPLITUDE, RJMetrics, BLUECORE, sumall, granify, Airtable, retention SCIENCE, custora

Applications

Sales & Marketing
RADIUS, Gainsight, bloomreach, Zeta, livefyre, blueyonder, kahuna, Lattice, SAILTHRU, persado, infer, 6sense, AVISO, ACTIONIQ, QUANTIFIND, ENGAGIO

Customer Service
MEDALLIA, ATTENSITY, CLARABRIDGE, STELLAService, Preact, NGDATA, wise.io, DigitalGenius, appuri, fuse|machines

Human Capital
gild, Connectifier, textio, entelo, hiQ

Legal
RAVEL, JUDICATA, Everlaw, Brevia, PREMONITION

Ad Optimization
MediaMath, Integral Ad Science, rocketfuel, OpenX, theTradeDesk, Adgorithms, LiveIntent, dstillery, DataXu, Appier, TAPAD

Security
CYLANCE, CounterTack, cybereason, AREA 1 SECURITY, SentinelOne, ThreatMetrix, Recorded Future, Guardian Analytics, FORTSCALE, siftscience, Keybase, feedzai, SIGNIFYD

Vertical AI Applications
facebook, Clara, KASISTO, lumiata

Publisher Tools
outbrain, mixpanel, Chartbeat, yieldbot, Yieldmo

Govt/ Regulation
Socrata, OPENGOV, FiscalNote, enigma, PREDPOL, mark43, OpenDataSoft

Finance
affirm, LendingClub, OnDeck, Kreditech, zestfinance, LendUp, Kabbage, tidemark, Payoff, INSIKT, zuora, Dataminr, Lenddo, KENSHO, AIDYIA, iSENTIUM, Quantopian, sentient technologies

Education/ Learning
KNEWTON, Clever, declara, PANORAMA EDUCATION, knowre

Life Sciences
23andMe, PATHWAY GENOMICS, Counsyl, deep genomics, Recombine, KYRUUS, FLATIRON, zymergen, HealthTap, METABIOTA, ZEPHYR HEALTH, ovia, Ginger.io, transcriptic, Glow, enlitic, AiCure, Atomwise

Industries
OPOWER, eHarmony, RetailNext, duetto, STITCH FIX, WorkFusion, BLUE RIVER, TACHYUS, SwiftKey, Seeq, FarmLogs, HowGood, celect, SIGHT MACHINE, statmuse, BOXEVER

Cross-Infrastructure/Analytics

amazon web services, Google, Microsoft, IBM, SAP, sas, hp, Autonomy, vmware, talend, TIBCO, TERADATA, ORACLE, NetApp

Open Source

Framework
hadoop HDFS, hadoop map reduce, YARN, MESOS, Spark, TEZ, Apache Kylin, Flink, CDAP

Query / Data Flow
SLAMDATA, HIVE, APACHE DRILL, Google Cloud Dataflow

Data Access
cassandra, APACHE HBASE, accumulo, mongoDB, couchbase, SciDB, kafka, CouchDB, riak, OPENTSDB, nifi

Coordination
talend, OOZIE, Apache Zookeeper, Apache Ambari

Real-Time
STORM, Spark, APEX, Flink, TACHYON, druid

Stat Tools
R, Scala, NumPy, SciPy

Machine Learning
mllib, mahout, Apache SINGA, Aerosolve, MADlib, Caffe, torch, CNTK, TensorFlow, learn, FeatureFu, WEKA, VELES, DIMSUM, jupyter, DL4J

Search
elasticsearch, Solr, lucene

Security
Apache Ranger

Visualization
Zeppelin

Data Sources & APIs

Health
JAWBONE, GARMIN, practice fusion, fitbit, Withings, VALIDIC, netatmo, kinsa, Human API

IOT
UPTAKE, ThingWorx, helium, samsara, AUGURY, estimote

Financial & Economic Data
Bloomberg, DOW JONES, YODLEE, PREMISE, S&P CAPITAL IQ, quandl, xignite, CB INSIGHTS, mattermark, estimize, PLAID

Air / Space / Sea
PLANET LABS, spire, WINDWARD, CRUISE, SKYCATCH, Airware, DroneDeploy

Location/People/Entities
GARMIN, foursquare, InsideView, esri, STREETLINE, CARTODB, factual, PlaceIQ, Crimson Hexagon, placemeter, BASIS TECHNOLOGY, Sense 360

Other
qualtrics, panjiva, DATA.GOV

Incubators & Schools
GA, DataCamp, INSIGHT, DataElite, METIS, The Data Incubator

SUCESU RS

GRUPO DE USUÁRIOS DE BI (GUBI)

OBRIGADO

Cassio Socal Cervo
cassiocervo@gmail.com
51 99455-4260