

Analise de Dados com a Linguagem R



Minha Apresentação

- Doutor em Informática e Gestão do Conhecimento pelo PPGI - Programa de Pós Graduação em Informática e Gestão do Conhecimento da UNINOVE (Conceito CAPES 4) realizado com Bolsa CAPES/PROSUP e UNINOVE.
- Possui mestrado em Administração Gestão de Projetos pela UNINOVE (Conceito CAPES 5), especialização em Energia pela escola politécnica da USP (PECE), Administração Industrial pela Escola Politécnica USP (Fundação Vanzolini), Engenharia de Refrigeração e Ar Condicionado pela FEI (Faculdade de Engenharia Industrial).
- Atualmente é pesquisador e docente permanente do Programa de Mestrado e Doutorado em Informática e Gestão do Conhecimento na Universidade Nove de Julho e docente nos cursos de graduação na mesma universidade.
- Possui certificações PMP Project Management Professional, Segurança da Informação e Cloud Computing.
- Atualmente suas pesquisas envolvem o emprego de técnicas de Inteligência Computacional, estudando o emprego de técnicas como Lógica Fuzzy, Redes Neurais e Máquinas de Vetor de Suporte (SVM) e Processamento de Linguagem Natural em áreas **como detecção de fraudes, na gestão de riscos , bem como no apoio na tomada decisão em condições de riscos e na análise automática de textos educacionais. .**

Nossa Disciplina

- Principais Tópicos
 - O que é Big Data
 - O que é Ciência de Dados
 - Linguagem R (Programação Básica)
 - Noções de Estatística
 - Análise Exploratória de Dados (EDA)
 - Mineração de Dados (Framework)
 - Manipulação de Bases de Dados para Análise
 - Visualização de Dados
 - Modelagem
 - Análise de Resultados

Nossos porque...

- Essa disciplina é importante porque...
- Cada vez mais dados são produzidos
- Cada vez mais empresas entendem que podem obter conhecimentos a partir desses dados
- Cada vez mais analisar dados é uma competência demandada pelo mercado
- Para lidar com esses dados e aproveitar essas você precisa dominar alguma linguagem focada em análise de dados, particularmente (R, Python ou Julia)

O QUE É R?

- O R é uma linguagem e ambiente voltados para estatística computacional e gráficos. Aberto e gratuito, é compatível com Windows, UNIX e MacOS. Atualmente o R se encontra em sua versão 4.0.2.
- Apenas para evitar a confusão, o R pode se referir às duas coisas: ao *software* utilizado para executar o código escrito na linguagem R e à linguagem em si. Para mais informações, podemos visitar o [site do Projeto R](#)



[Home]

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Developer Pages](#)

[R Blog](#)

R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

Help With R

[Getting Help](#)

Documentation

[Manuals](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- [R version 4.0.2 \(Taking Off Again\)](#) has been released on 2020-06-22.
- [useR! 2020 in Saint Louis has been cancelled](#). The European hub planned in Munich will not be an in-person conference. Both organizing committees are working on the best course of action.
- [R version 3.6.3 \(Holding the Windsock\)](#) has been released on 2020-02-29.
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

News via Twitter

The R Foundation Retweeted



useR2020muc

@useR2020muc

Please let us know how you liked #useR2020

Even if you joined only 1 event or watched 1 talk, your feedback is helpful to us and will help plan future #stats events
https://twitter.com/R_Forwards/status/1290193576169803776

Instalando o R

- Para instalar o R, basta seguir os seguintes passos:
- Visite um dos *CRAN Mirrors* para fazer o download do software.
- Sugerido: (<https://vps.fmvz.usp.br/CRAN/>);
- Escolha a versão para o seu sistema operacional (Linux, Mac OS ou Windows);
- Na página seguinte, escolha a versão **base**;
- Baixe o arquivo disponível no link e instale!

Introdução ao R x CRAN - Mirrors

cran.r-project.org/mirrors.html

Apps Sites Sugeridos Importado do IE Google Inteligência Artificia... Rafael Santos Como instalar uma... Index of /~oda/con... Resultado de image... Sci-Hub: removing... » Outros favoritos

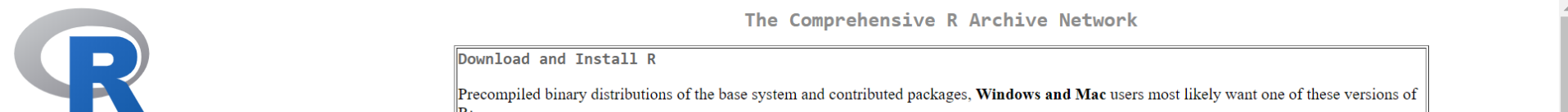
CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: [main page](#), [windows release](#), [windows old release](#).

If you want to host a new mirror at your institution, please have a look at the [CRAN Mirror HOWTO](#).

0-Cloud	https://cloud.r-project.org/	Automatic redirection to servers worldwide, currently sponsored by Rstudio
Algeria	https://cran.usthb.dz/	University of Science and Technology Houari Boumediene
Argentina	http://mirror.fcaglp.unlp.edu.ar/CRAN/	Universidad Nacional de La Plata
Australia	https://cran.csiro.au/ https://mirror.aarnet.edu.au/pub/CRAN/ https://cran.ms.unimelb.edu.au/ https://cran.curtin.edu.au/	CSIRO AARNET School of Mathematics and Statistics, University of Melbourne Curtin University
Austria	https://cran.wu.ac.at/	Wirtschaftsuniversität Wien
Belgium	https://www.freeststatistics.org/cran/ https://lib.ugent.be/CRAN/	Patrick Wessa Ghent University Library
Brazil	https://nbcgib.uesc.br/mirrors/cran/ https://cran-r.c3sl.ufpr.br/ https://cran.fiocruz.br/ https://vps.fmvz.usp.br/CRAN/ https://brieger.esalq.usp.br/CRAN/	Computational Biology Center at Universidade Estadual de Santa Cruz Universidade Federal do Parana Oswaldo Cruz Foundation, Rio de Janeiro University of Sao Paulo, Sao Paulo University of Sao Paulo, Piracicaba
Bulgaria	https://ftp.uni-sofia.bg/CRAN/	Sofia University
Canada	https://mirror.rcg.sfu.ca/mirror/CRAN/ https://muug.ca/mirror/cran/ https://mirror.its.dal.ca/cran/ http://cran.utstat.utoronto.ca/	Simon Fraser University, Burnaby Manitoba Unix User Group Dalhousie University, Halifax University of Toronto

Windows taskbar: Digite aqui para pesquisar, 16:34, 17/08/2020



The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2020-06-22, Taking Off Again) [R-4.0.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to

Instalando o RStudio

- O RStudio é um ambiente de interface gráfica para o R, ou melhor, um IDE (*Integrated Development Environment*).
- Atualmente o RStudio é considerado o melhor IDE para quem programa em R.
- Além de uma interface mais amigável, possui diversas funcionalidades que facilitam o aprendizado e a produtividade.
- O RStudio possui uma versão para desktop grátis e para baixar e instalar, basta seguir os seguintes passos:
- Ir até a [página de download do RStudio](#);
- Clicar em download para a versão grátis;
- Escolher o instalador de acordo com o seu sistema operacional (Windows, Mac OS, Linux);
- Após o download, realizar a instalação.

Download RStudio

Choose Your Version

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace.

[LEARN MORE ABOUT RSTUDIO FEATURES](#)



RStudio's new solution for every professional data science team. RStudio Team includes RStudio Server Pro, RStudio Connect and RStudio Package Manager.

[LEARN MORE](#)

RStudio Desktop

Open Source License

Free

[DOWNLOAD](#)

[Learn more](#)

RStudio Desktop

Commercial License

\$995

/year

[BUY](#)

[Learn more](#)

RStudio Server

Open Source License

Free

[DOWNLOAD](#)

[Learn more](#)

RStudio Server Pro

Commercial License

\$4,975

/year

(5 Named Users)

[BUY](#)

[Evaluation](#) | [Learn more](#)

Integrated Tools for R



RStudio Desktop 1.3.1073 - Release Notes

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



Requires Windows 10/8/7 (64-bit)



All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an [older version of RStudio](#).

OS	Download	Size	SHA-256
Windows 10/8/7	RStudio-1.3.1073.exe	171.62 MB	2fe472a
macOS 10.13+	RStudio-1.3.1073.dmg	148.66 MB	0878b305
Ubuntu 16	rstudio-1.3.1073-amd64.deb	124.07 MB	6d71c5ff
Ubuntu 18/Debian 10	rstudio-1.3.1073-amd64.deb	126.78 MB	86be9352
Fedora 19/Red Hat 7	rstudio-1.3.1073-x86_64.rpm	146.95 MB	01abb3d8
Fedora 28/Red Hat 8	rstudio-1.3.1073-x86_64.rpm	151.04 MB	4b4e4878
Debian 9	rstudio-1.3.1073-amd64.deb	126.98 MB	0226bbc2
SLES/OpenSUSE 12	rstudio-1.3.1073-x86_64.rpm	119.43 MB	7c1a6f2c
OpenSUSE 15	rstudio-1.3.1073-x86_64.rpm	128.39 MB	29078f11

<https://download1.rstudio.org/desktop/windows/RStudio-1.3.1073.exe>

Algumas Referências em R

- R in a Nutshell: A Desktop Quick Reference (Adler [2012](#))
- Learning R: A Step-by-Step Function Guide to Data Analysis (Cotton [2013](#))
- **R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (Wickham and Grolemund [2016](#))**
- 25 Recipes for Getting Started with R (Teetor [2011a](#))
- R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics (Teetor [2011b](#))
- The Essential R Reference (Gardener [2012](#))
- **Introdução a Mineração de Dados com Aplicações em R – (Silva, Peres e Boscaroli)**

Materiais da Disciplina

- O pdf com as aulas será enviado pelo link material de apoio (verifique como acessar esses dados, assim que estiver disponível)
- Há um repositório da disciplina com mais materiais de apoio, códigos etc...
- Esse repositório fica no Github
- https://github.com/DeepFluxion/Analise_Dados_Linguagem_R

Learn Git and GitHub without any code!

Using the Hello World guide, you'll start a branch, write comments, and open a pull request.

[Read the guide](#)


Projects Wiki Security Insights Settings

master 1 branch 0 tags

[Go to file](#)

[Add file](#)

[Code](#)

	DeepFluxion Add files via upload	83faeb2 15 minutes ago 8 commits
Adicionais	Add files via upload	15 minutes ago
Apresentacoes	Criado Diretório de Apresnetações	32 minutes ago
Notebooks	Criação do diretório	33 minutes ago
Scripts	Create .gitkeep	30 minutes ago
README.md	Update README.md	31 minutes ago

README.md

Analise_Dados_Linguagem_R

Repositório de Arquivos e Matreiais de Aula da Disciplina Análise de dados coma Linguagem R 2o. Semestre 2020

- Na pasta **Notebooks** estão os arquivos no formato Jupyter Notebooks
- Na pasta **Scripts** estão os arquivos no formato RScript para serem usados no RStudio

About

Repositório de Arquivos e Matreiais de Aula da Disciplina Análise de dados coma Linguagem R 2o. Semestre 2020

[Readme](#)

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Science....pdf

Using R and RStud....pdf

R Cookbook Pro....epub

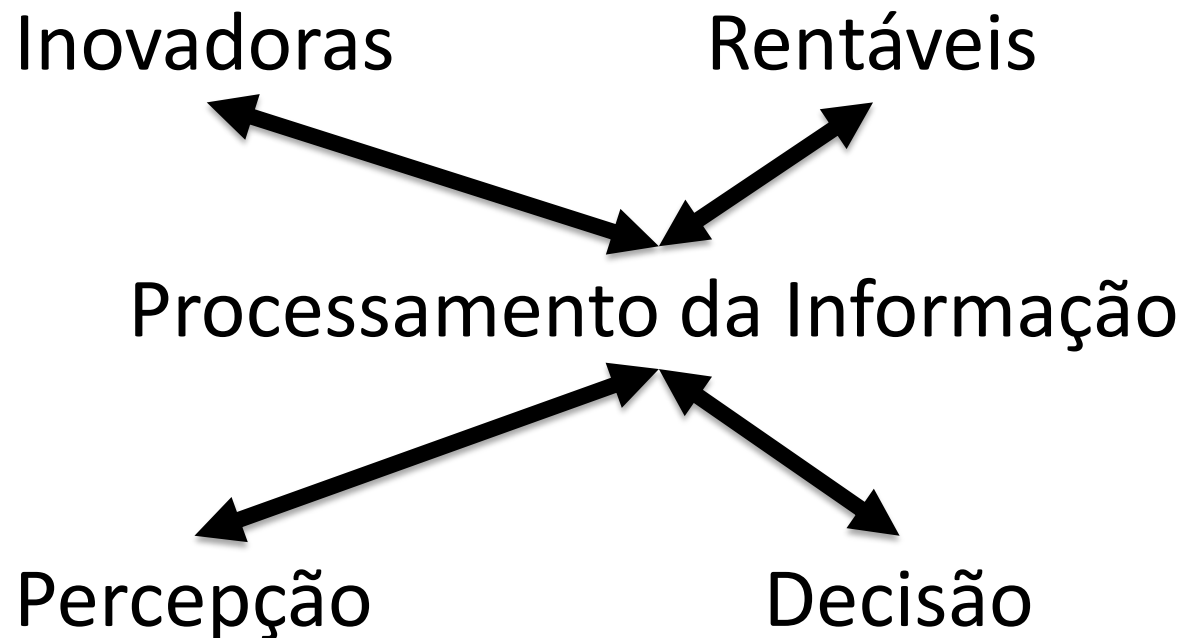
Hands-On Progra....pdf

R for Data

INTRODUÇÃO AO BIG DATA

Introdução ao Big Data

- “Big Data faz referência ao grande **volume**, **variedade** e **velocidade** de dados que demandam formas inovadoras e rentáveis de processamento da informação, para melhor percepção e tomada de decisão.” (*Gartner*)



Sites

- Introdução ao R *Curso Básico de Linguagem R* Eduardo Ferreira / Weverton Lisboa (https://bookdown.org/wevsena/curso_r_tce/curso_r_tce.html#introducao)
- **Hands-On Programming with R** Garrett Grolemund (<https://rstudio-education.github.io/hopr/>)
- **R for Data Science** Garrett Grolemund Hadley Wickham (<https://r4ds.had.co.nz/>)
- *Blog - Revolutions* (<https://blog.revolutionanalytics.com/>)
- *Blog – Rbloggers* (<https://www.r-bloggers.com/>)
- Portal – kdd nuggets (<https://www.kdnuggets.com/>)
- Ajuda:
- Documentação dos pacotes;
- Ajuda das funções (Digite ? seguido do nome da função);
 - [Comunidade do RStudio](#);
 - [StackOverflow](#);
 - [GitHub](#).
 - Google

Introdução ao Big Data

- “Big Data faz referência ao grande **volume, variedade e velocidade** de dados que demandam formas inovadoras e rentáveis de processamento da informação, para melhor percepção e tomada de decisão.” (*Gartner*)



01100101011100100
0000011000010111000
1010011000010111000
1011011100110011100
11100100111011101
0100000011100000110
1000001101001011011
0110000101110000011

DATA
IS THE NEW OIL

Data Monetization

Capitalizing on Your Data Assets



Data Driven Decisions



Data
Processing



Data
Analytics



Data Driven
Decisions

Introdução ao Big Data

- “Big Data faz referência ao grande **volume**, **variedade** e **velocidade** de dados que demandam formas inovadoras e rentáveis de processamento da informação, para melhor percepção e tomada de decisão.” (*Gartner*)

Os 3 Vs de Big Data

- **Os 3 Vs de Big Data**
 - Volume
 - Variedade
 - Velocidade



BIG DATA



VOLUME

DATA SIZE



VELOCITY

SPEED OF CHANGE



VARIETY

DIFFERENT FORMS
OF DATA SOURCES



VERACITY

UNCERTAINTY OF
DATA

Os 3 Vs de Big Data

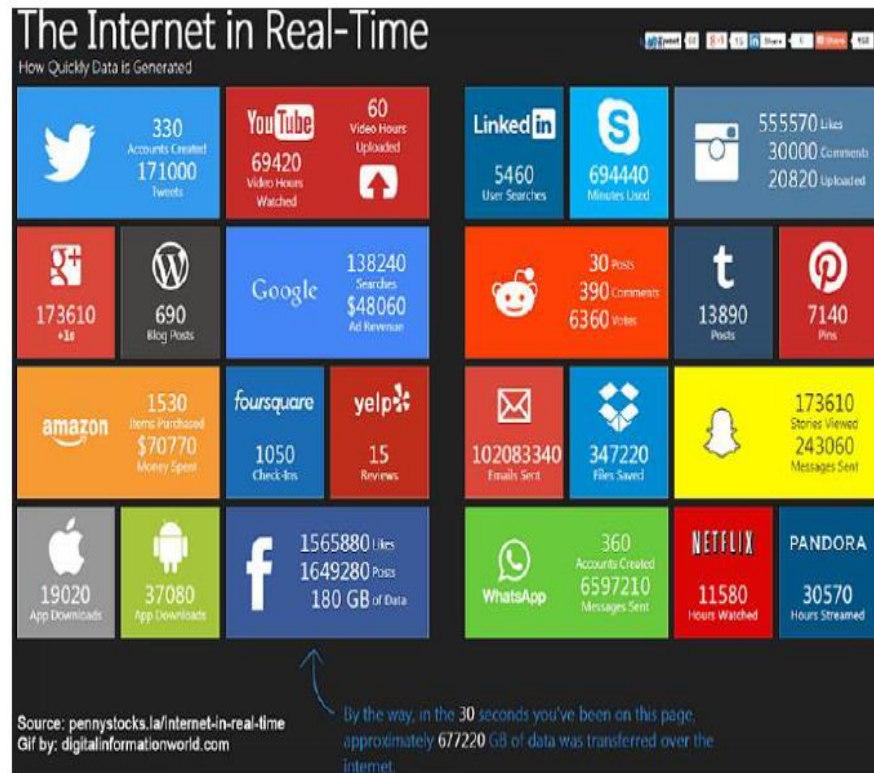
- **Volume**

- Cerca de 2,5 bilhões de gigabytes de dados são criados diariamente;
- De toda a quantidade de dados disponível no mundo, aproximadamente 90% foi criado nos últimos 2 anos;
- 1,8 bilhão de usuários ativos no Facebook;
- 1 bilhão de usuários ativos no WhatsApp;
- 95 milhões de fotos e vídeos por dia no Instagram;
- 44 milhões de artigos na Wikipedia;
- 4 bilhões de visualizações por dia no YouTube;
- 300 horas de vídeos são carregados a cada 1 minuto no Youtube;

Os 3 Vs de Big Data

- **Velocidade**

- Processamento em tempo real e Streaming Data (dados em streaming).
- O que acontece em 30 segundos na Internet.



Os 3 Vs de Big Data

- **Variedade**

- Dados são criados em diferentes formatos - como e-mails, comentários no Facebook, fotos publicadas em redes sociais e transações.
- Big Data inclui dados estruturados, semi-estruturados e não-estruturados.
 - Dados estruturados:
 - Bases de dados relacionais.
 - Dados semi-estruturados:
 - Não possuem uma estrutura pré-definida (representação estrutural heterogênea).
 - Auto-descritivos e sem esquema prévio definido.
 - Possuem esquema de representação presente (de forma explícita ou implícita).
 - Exemplo: XML (eXtensible Markup Language).
 - Dados não-estruturados:
 - Documentos, fotos, vídeos, *tweets*, comentários em redes sociais, etc.
 - Estima-se que pelo menos **80%** dos dados gerados atualmente sejam do tipo não-estruturados.

Outros Vs do Big Data

- Além dos 3 Vs, outras duas dimensões são comumente associadas à definição de Big Data. São elas:

- **Veracidade**

- Confiabilidade dos dados.

- **Valor**

- Gerar valor para o negócio;
 - Melhor entender as necessidades dos clientes;
 - Oferecer produtos e serviços que melhor atendam as necessidades dos clientes;
 - Oferecer produtos e serviços personalizados;
 - Melhorar o relacionamento com os clientes;
 - Aumentar a fidelização e satisfação dos clientes;
 - Gerar vantagem competitiva para o negócio.

Desafios do Big Data

- Onde armazenar esses dados?
- Como estruturar esses dados?
- Como consultar esses dados?
- Como extrair valor desses dados?
- Necessidade de novas tecnologias capazes de oferecer escalabilidade, disponibilidade, flexibilidade e desempenho para a manipulação de grandes volumes de dados.

Desafios do Big Data

- Big Data necessita de grande capacidade de processamento e armazenamento.
- Computação em Nuvem oferece capacidade de processamento e armazenamento conforme a necessidade do usuário.
- Computação em Nuvem é um imperativo para Big Data.
- De acordo com NIST (*National Institute of Standards and Technology*): Computação em Nuvem é um modelo que permite um acesso sob demanda via redes de computadores a um conjunto compartilhado de recursos computacionais que podem ser rapidamente provisionado e liberado com um mínimo de esforço administrativo ou interação com o provedor de serviços.

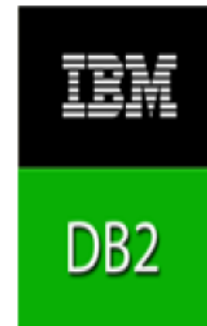
Armazenamento

- SGBDs mais utilizados:



ORACLE®

TERADATA®



Business Intelligence (BI)

- **Business Intelligence (BI)**

- Consolida os principais indicadores de uma empresa com base em dados obtidos no sistema de gestão integrado ERP e fornece visões precisas e analíticas, que apoiam a tomada de decisão.
- As ferramentas de BI geram gráficos (Dashboards) e relatórios de acordo com a necessidade do usuário.
- Por meio do BI podem ser detectadas sazonalidades, tendências e padrão de comportamento.

Business Intelligence (BI)

- Ferramentas de Business Intelligence (BI)



Armazenamento

- **Base de dados não relacionais**
 - **NoSQL (Not Only SQL)**
 - Conjunto de conceitos que permite o processamento rápido e eficiente de conjuntos de dados com foco em desempenho, confiabilidade e agilidade.
 - Diferentes formas de armazenamento
 - Orientado a documentos (o mais popular)
 - Ex.: MongoDB (<https://www.mongodb.com/>) e Apache CouchDB (<https://couchdb.apache.org/>)
 - Orientado a chave-valor (o mais simples)
 - Ex.: Amazon DynamoDB (<https://aws.amazon.com/dynamodb/>)
 - Orientado a grafos (o mais especializado)
 - Ex.: Neo4j (<https://neo4j.com/>)
 - Orientado a colunas (o mais complexo)
 - Ex.: HBase (<http://hbase.apache.org/>)
 - Características
 - Não-relacional
 - *Cluster-friendly*
 - Interface de consulta simples.

Business Intelligence



- Ajuda você a conhecer o seu negócio;
- Melhora o processo de tomada de decisão, baseando-a em evidências;
- Facilita o acesso e o compartilhamento de informações que darão suporte à gestão;
- Permite análise em tempo real com navegação rápida;
- Ajuda a identificar perdas no sistema;
- Reduz o risco de problemas e obstáculos;
- Entrega resultados corretos, no tempo certo, para a pessoa exata;
- Identifica oportunidades de vendas cruzadas e de vendas diretas;
- Possibilita respostas rápidas a qualquer consulta de negócios;
- Obtém informações valiosas sobre o comportamento de seus clientes.

Armazenamento

- **Data Lake**

- Repositório único no qual dados estruturados e não-estruturados, coletados de diferentes fontes, são armazenados em sua forma bruta, como foram coletadas na fonte, sem qualquer processamento.

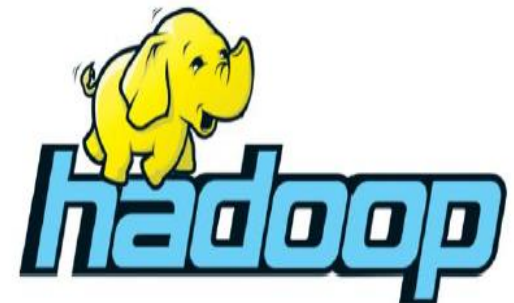
- **Enterprise Data Hub (EDH)**

- Permite que a empresa tenha uma fonte de dados centralizada e unificada que possa fornecer rapidamente informações a diversos usuários do negócio, apoiando a tomada de decisão.
 - Soluções:
 - Azure Data Lake
 - <https://azure.microsoft.com/pt-br/solutions/data-lake/>
 - Cloudera Enterprise Data Hub
 - <https://www.cloudera.com/products/enterprise-data-hub.html>
 - Enterprise Data Hub (MapR)
 - <https://mapr.com/solutions/enterprise/enterprise-data-hub/>

Processamento

- **Hadoop**

- <http://hadoop.apache.org/>
- Solução *open source* que permite a execução de aplicações de Big Data utilizando milhares de máquinas.
- Projetado para processar grandes quantidades de dados estruturados e não-estruturados.
- Oferece recursos de armazenamento, gerenciamento e processamento de dados distribuídos.
- Benefícios:
 - Redução de custo
 - Flexibilidade
 - Escalabilidade
 - Desempenho



Processamento

- Principais fornecedores de mercado

cloudera



Processamento

- **Ecosistema Hadoop**
 - Hadoop possui 2 componentes principais:



Processamento

- **Hadoop HDFS (Hadoop Distributed File System)**



- Sistema de arquivos distribuídos;
- Otimizado para processamento de **grandes volumes de dados (alta taxa de transferência)**;
- Otimizado para ler e armazenar grandes arquivos em *clusters*;
- Arquivos são divididos em blocos de 64 ou 128 MB (*tamanho default – pode ser configurado*);
- Escalável e tolerante a falhas.

- **Hadoop MapReduce**



- É uma implementação do Hadoop;
- Ferramenta utilizada para facilitar o processamento de grandes volumes de dados (multi-terabyte data-sets) de forma distribuída;
- Tolerante a falhas;
- Funções *Map* e *Reduce*.

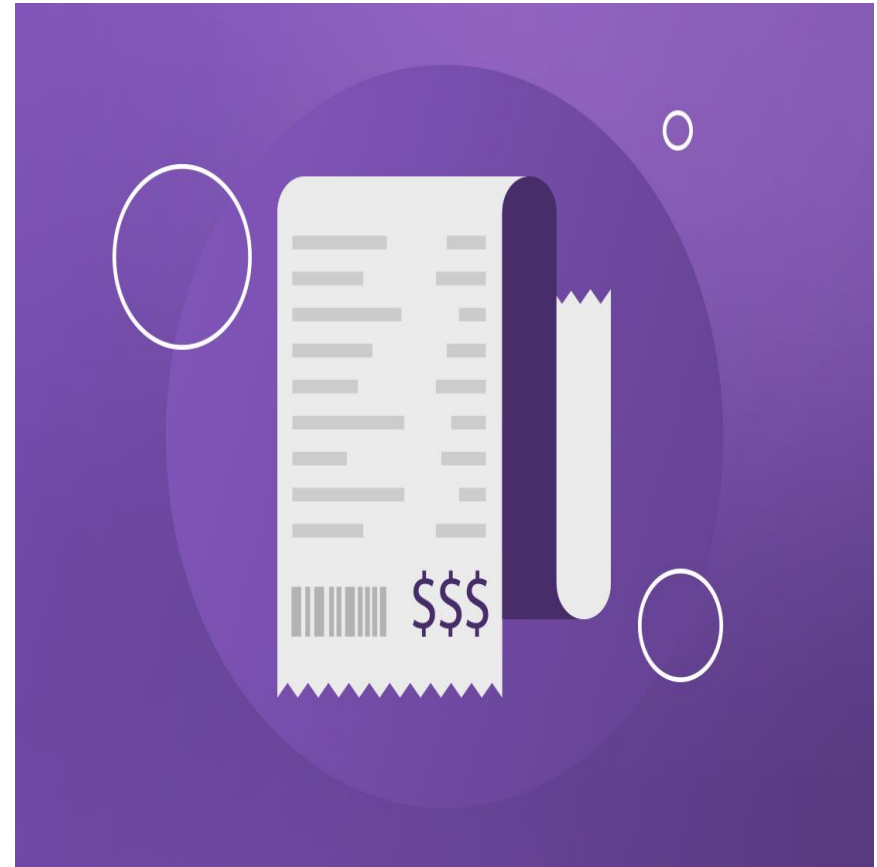
Mãos a obra

- A empresa “Magazine Maraísa” possui 125 lojas espalhadas pelo Brasil
- O sistema MaraNota concentra todas as notas fiscais emitidas pela Maraísa em todas as lojas e é atualizado de modo quase instantâneo



A Nota Fiscal

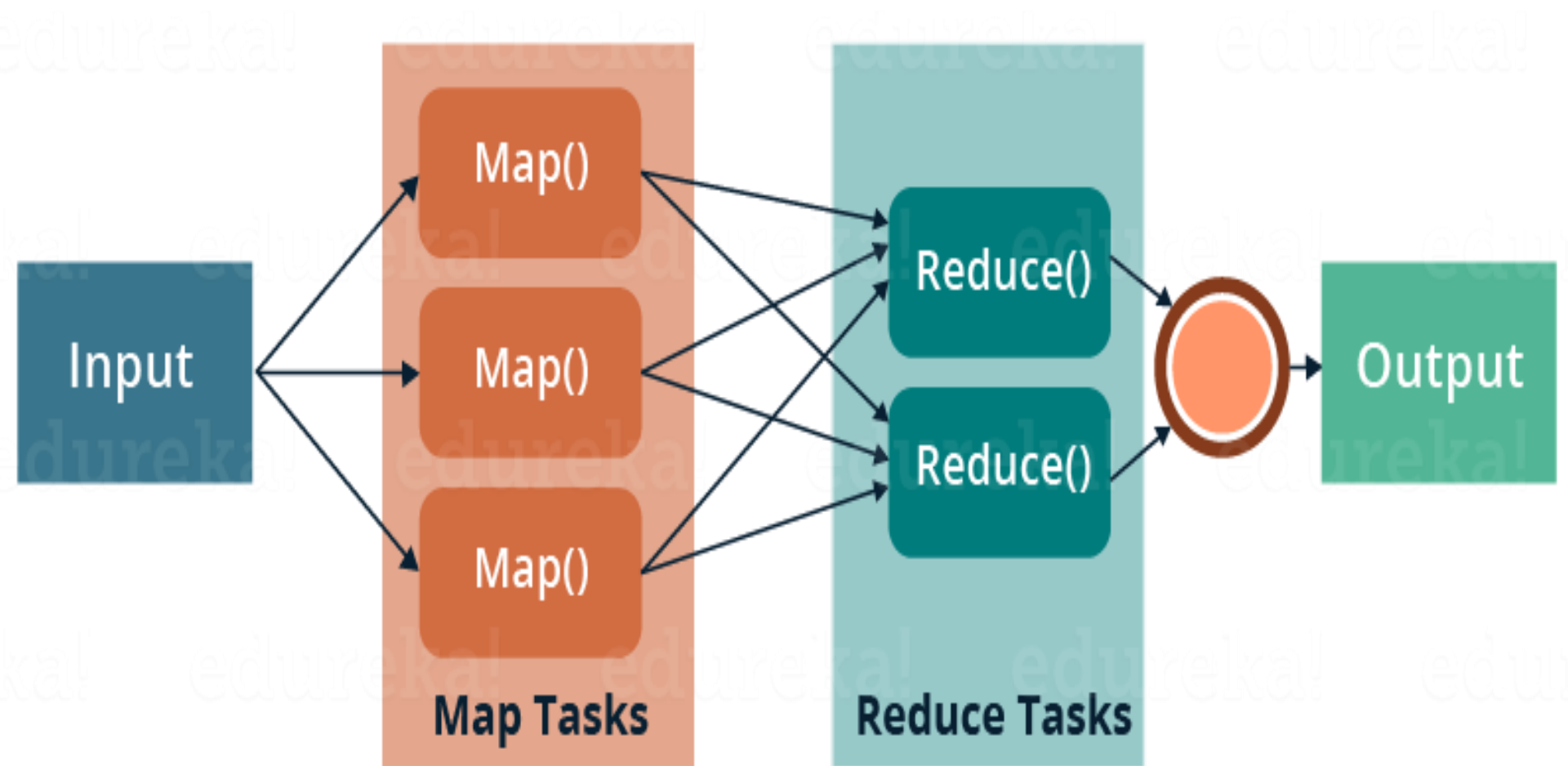
- Cada nota possui um ou mais registros que possuem os seguintes atributos:
 - Numero da Nota Fiscal
 - Nome do Cliente
 - Produto adquirido
 - Valor do Produto adquirido
 - Valor do Imposto
 - Código da Loja
 - Código do Departamento da Loja
 - Código do Vendedor
 - Forma de Pagamento
 - Parcelamento
 - Desconto



Que tipo de relatório podemos gerar e como?

- Para a Presidente da Magazine Maraísa
- Para o Diretor Comercial
- Para um Gerente de Loja
- Para um Gerente de Departamento
- Para o Diretor de RH
- Para o Diretor de Materiais e Logística





Nota
de
Venda

Loja

Produtos

MAP

Loja 1 Venda 1 Venda 2 ...

Loja 2 Venda 1 Venda 2 ...

Loja 3 Venda 1 Venda 2 ...

...

Loja 125 Venda 1 Venda 2 ...

REDUCE

Loja 1 Total Vendas 1

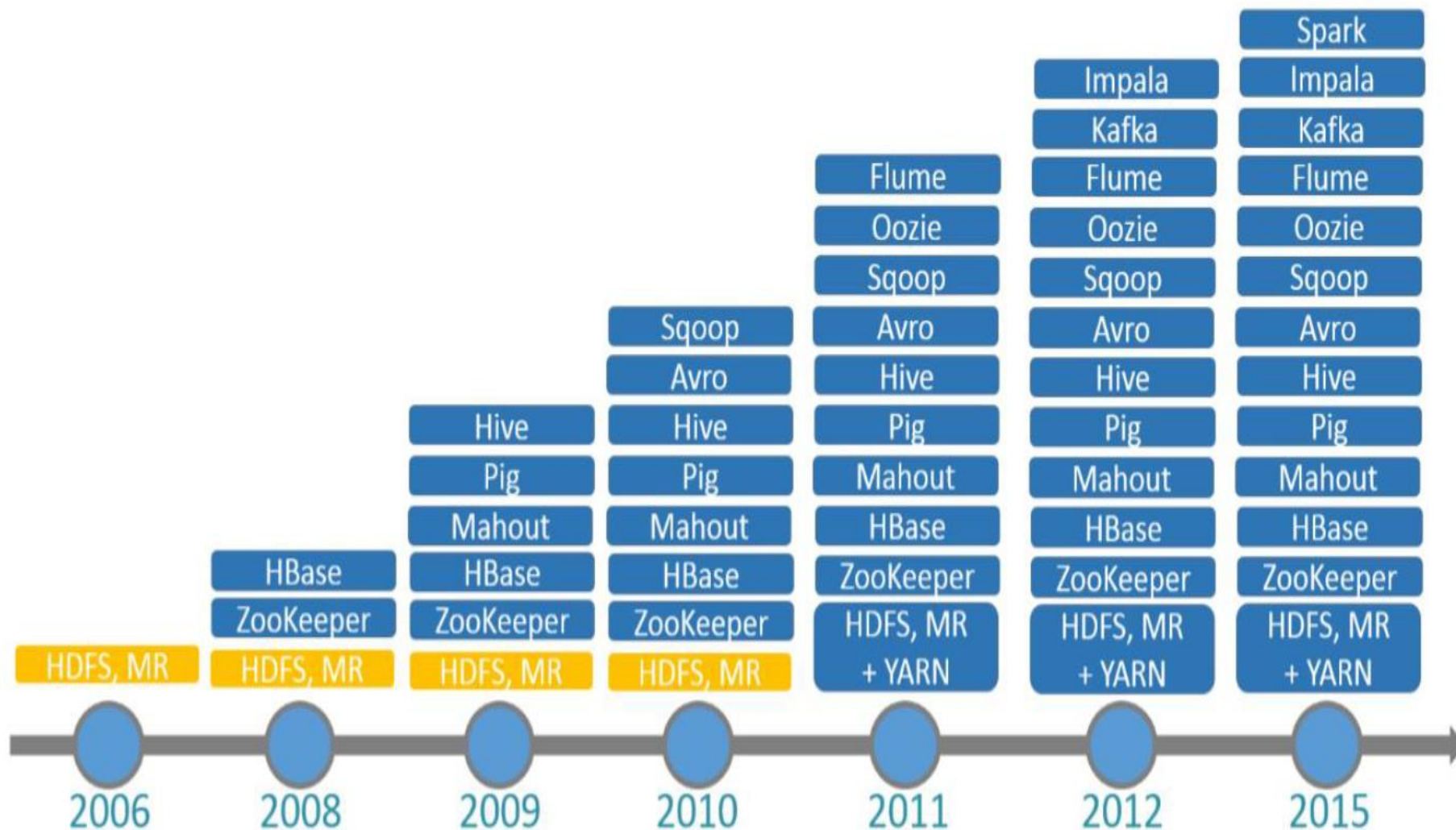
Loja 2 Total Vendas 2

Loja 3 Total Vendas 3

...

Loja 125 Total Vendas 125

Ecossistema Hadoop



Processamento

- Algumas empresas que utilizam o Hadoop



Aplicações

- **Varejo**

- Melhor segmentação de clientes;
- Propaganda personalizada;
- Melhor oferta de produtos e serviços com base no perfil, rastro digital e no histórico de compras do cliente;
- Previsão e prevenção de *Customer Churn* (identificação de clientes com alta propensão a cancelar produtos e serviços);
- *Chatbots*.

- **Setor financeiro**

- Detecção de transações fraudulentas envolvendo utilização de Internet Banking e cartões de crédito;
- Análise de crédito;
- Melhor relacionamento com os clientes.

Aplicações

- **People Analytics (HR Analytics)**

- Processo de coleta, armazenamento e análise de dados sobre o comportamento dos colaboradores em uma organização.
- Utilização de análise de dados em Gestão de Pessoas.
- Utilização de informações disponíveis em redes sociais.
- Utilização de leitores biométricos e crachás inteligentes.
- Análise de currículo utilizando *Text Analytics* (Análise de Texto).
- Utilização de rastro digital e informações de redes sociais para ajudar na identificação do perfil mais adequado para cada vaga.
- Principais benefícios:
 - Otimização do processo de Recrutamento e Seleção;
 - Avaliação de Desempenho;
 - Aumento da produtividade;
 - Desenvolvimento de programas de treinamento e capacitação;
 - Retenção de talentos;
 - Redução da Rotatividade.

Aplicações

- **Internet of Things (Internet das Coisas)**

- Rede formada por milhares de dispositivos (objetos) inteligentes conectados a Internet.
- Dispositivos capazes de capturarem grandes quantidades de dados por meio de sensores.
- Exemplos:
 - Veículos autônomos e conectados;
 - Casas inteligentes;
 - Eletrodomésticos inteligentes;
 - *Wearables* (Relógios e pulseiras inteligentes).

Exemplos de empresas com negócios centrados em dados



NETFLIX

