

IBOPE **dtm**

*Big Data Analytics e o R*

Fundada  
em

2004

a DTM é uma empresa de marketing científico que integra dados, dentro do conceito de

*big data*

para elaboração e implementação de estratégias de marketing e sistemas de tomada de decisão.

Em  
2013

consolidou-se a união da DTM com o IBOPE, grupo líder em provimento de informação e conhecimento em toda a América Latina.

Esta união soma o profundo conhecimento do consumidor brasileiro e o RIGOR TÉCNICO e METODOLÓGICO do IBOPE à administração das estratégias de relacionamento da DTM.

# Quem somos nós



## Inteligência de Marketing

Equipe formada por estatísticos, matemáticos e econométristas, responsáveis em trabalhar com os dados e identificar os principais indicadores de acompanhamento do relacionamento do cliente.

A metodologia de Análise 360 Graus permite um estudo completo do comportamento do consumidor e configura o ponto de partida para o planejamento das ações.



## Planejamento e Gestão

Equipe responsável pela gestão do projeto e interface junto ao cliente. Especialistas nas diferentes indústrias e antenados com os movimentos do mercado, os consultores movimentam as engrenagens internas para criar um projeto específico para cada cliente.

São responsáveis pelo planejamento de relacionamento, gestão das ações planejadas e análise do andamento do projeto.



## Tecnologia da Informação

Equipe responsável pela captura de dados brutos cadastrais e transacionais dos clientes e no desenvolvimento dos ambientes de tratamento e armazenamento de dados, com a melhor infraestrutura disponível e todos os requisitos de segurança.

Os especialistas também desenvolvem as interações automáticas com os clientes e os respectivos relatórios de acompanhamento online.



## Criação

Equipe de redatores e diretores de arte que tangibilizam as estratégias e o planejamento, garantindo o melhor comunicação possível com cada perfil de consumidor, utilizando as técnicas mais modernas de personalização.

As peças são desenvolvidas a partir do profundo entendimento da ação e das informações apreendidas nas análises estatísticas, sempre respeitando as melhores práticas de marketing direto.



## Produtos

Equipe responsável pelo desenvolvimento de produtos que apoiam a gestão dos programas de relacionamento administrados pelo IBOPE DTM.

Estas ferramentas são desenvolvidas de acordo com a metodologia do IBOPE DTM e são parametrizadas conforme as necessidades de cada projeto.

Com estas ferramentas à disposição, a IBOPE DTM faz **entregas faseadas** e possibilita que o retorno do investimento seja percebido pelo cliente no decorrer do projeto.

# Principais Clientes

The logo for CAAO, featuring the letters 'CAOA' in a bold, blue, sans-serif font with a green horizontal line underneath.The logo for NET, with 'NET' in blue and 'O MUNDO É DOS NETS' in smaller blue text below it.The logo for Itaú, consisting of a blue square with the word 'Itaú' in yellow.The logo for TIM, featuring a blue oval with three red wavy lines and the word 'TIM' in white.The logo for RaiaDrogasil, with 'RaiaDrogasil' in black and a stylized red and green leaf-like graphic to the right.The logo for st|marche, with 'st|marche' in green and 'SUPERMERCADO' in smaller black text below.The logo for MAC, with the letters 'MAC' in a stylized, black, outlined font.The logo for Zona Sul, with 'Zona Sul' in black, a red heart icon, and the tagline 'Tudo pra você gostar da gente.' below.The logo for WIZARD, featuring a stylized red and blue bird-like graphic above the word 'WIZARD' in black.The logo for ESTADÃO, with a blue horse head icon and the word 'ESTADÃO' in blue.The logo for Grupo RBS, with 'Grupo' in black and 'RBS' in blue.The logo for grupo boticário, with a colorful circular graphic above the text 'grupo boticário'.The logo for Beleza Natural, with a red stylized 'B' icon and the text 'Beleza Natural' in red.The logo for SulAmérica, with a wavy orange line above the text 'SulAmérica'.The logo for xp investimentos, with 'xp' in a blue square and 'investimentos' in black.The logo for Unimed, with 'Unimed' in green and a green tree icon to the right.The logo for Leo, with 'Leo' in yellow on a black background and the tagline 'MADEIRAS E MUITO MAIS.' below.The logo for sv são vicente, with 'sv' in red and 'são vicente' in black.

# A metodologia

- O IBOPE DTM entende que todas as informações geradas por consumidores e prospects são essenciais para definir um plano de ação dirigido, capaz de criar uma estratégia de relacionamento com o intuito de atingir o objetivo de negócio da marca e ser mensurável de forma estatística.
  - Para que os resultados de nossos clientes sejam maximizados, as entregas dos programas de relacionamento, réguas de comunicação, ações pontuais, monitoramento, diagnóstico, planejamento e estruturação de canais de contato entre marca e consumidor/prospect são baseadas em:



## CAPTURA DE DADOS

variados disponibilizados pelos consumidores, integrando seus perfis online e offline;

## TRANSFORMAÇÃO DE DADOS

em informação, com insights de negócio e construção de uma estratégia de relacionamento sólida e efetiva com campanhas direcionadas;



## MENSURAÇÃO

contínua dos resultados e investimentos.

# A metodologia

CAPTURA DE DADOS

TRANSFORMAÇÃO DE DADOS

MENSURAÇÃO

JUNTO A ESTES PONTOS PRINCIPAIS, A METODOLOGIA DA IBOPE DTM ENVOLVE:

**IMERSÃO**

Análise de micro e macro ambiente, impactos, desafios e apresentação de cronograma detalhado.

**DIAGNÓSTICO**

Estudo de dados, indicadores de comportamento e consumo, segmentação e construção de data mart de acordo com a visão do cliente.

**PLANEJAMENTO E PARAMETRIZAÇÃO**

Aprendizado da etapa de Imersão, que auxiliam no desenho de estratégia de relacionamento, régua de comunicação e define os demais processos até os indicadores de resultados.

**SET-UP**

Implementação dos processos definidos anteriormente, criando peças de comunicação e desenvolvendo relatórios.

**GESTÃO**

Mensuração contínua dos resultados, aprimoramentos e ajustes com acompanhamento das demandas e atividades do projeto.



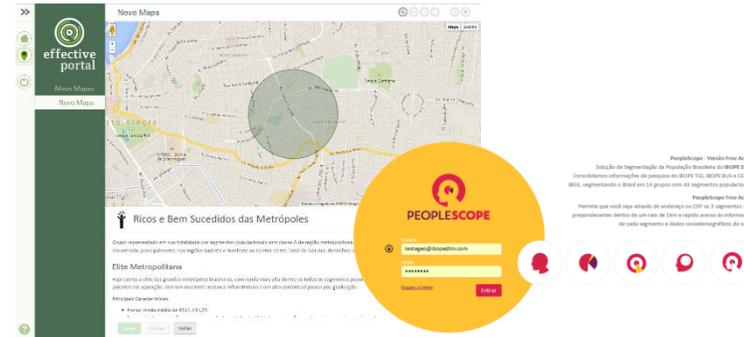
# IBOPE dtm SUÍTECRM

Nossa suite de produtos para gestão completa do relacionamento com clientes e prospects de uma marca



# PEOPLESCOPE

Segmentação da população brasileira com base em informações sócio-demográficas (CENSO) para enriquecer os nossos diagnósticos e modelos de propensão



# SOCIAL BRANDS

Portal de relatórios web e real time do monitoramento de marcas e seus universos nas redes sociais

# SISTEMA de RECOMENDAÇÃO

Sistema com os melhores modelos estatísticos para definir a oferta ideal para o consumidor ideal



**Quem comprou, também comprou:**

Adicione os produtos ao seu carrinho de compra

- Mouse Wireless Microsoft Azul Microsoft: Por: **R\$ 75,80** (10x R\$ 9,49 sem juros) [COMPRAR](#)
- Pen Drive Sandisk 8GB Cruiser Blade: Por: **R\$ 14,50** (10x R\$ 109,09 sem juros) [COMPRAR](#)
- Medida CaseLogic Para Notebook 17" - Vsb-217: Por: **R\$ 116,90** [COMPRAR](#)

**Quem viu este produto, viu também:**

- Notebook Acer ES 111 C7NE: Por: **R\$ 2.290,00** [COMPRAR](#)
- Notebook Acer ES 111 C7NE: Por: **R\$ 2.990,00** [COMPRAR](#)
- Notebook Acer ES 111 C7NE: Por: **R\$ 1.090,00** [COMPRAR](#)
- Notebook Dell ES 111 C7NE: Por: **R\$ 2.290,00** [COMPRAR](#)
- Notebook HP ES 111 C7NE: Por: **R\$ 3.290,00** [COMPRAR](#)



# O QUE É BIG DATA?

Prática de utilizar grandes volumes e variedades de dados, com velocidade, para inferir e agir sobre um determinado problema de interesse.

# O QUE É BIG DATA?

BIG  
DATA

=

Volume



+

Variedade



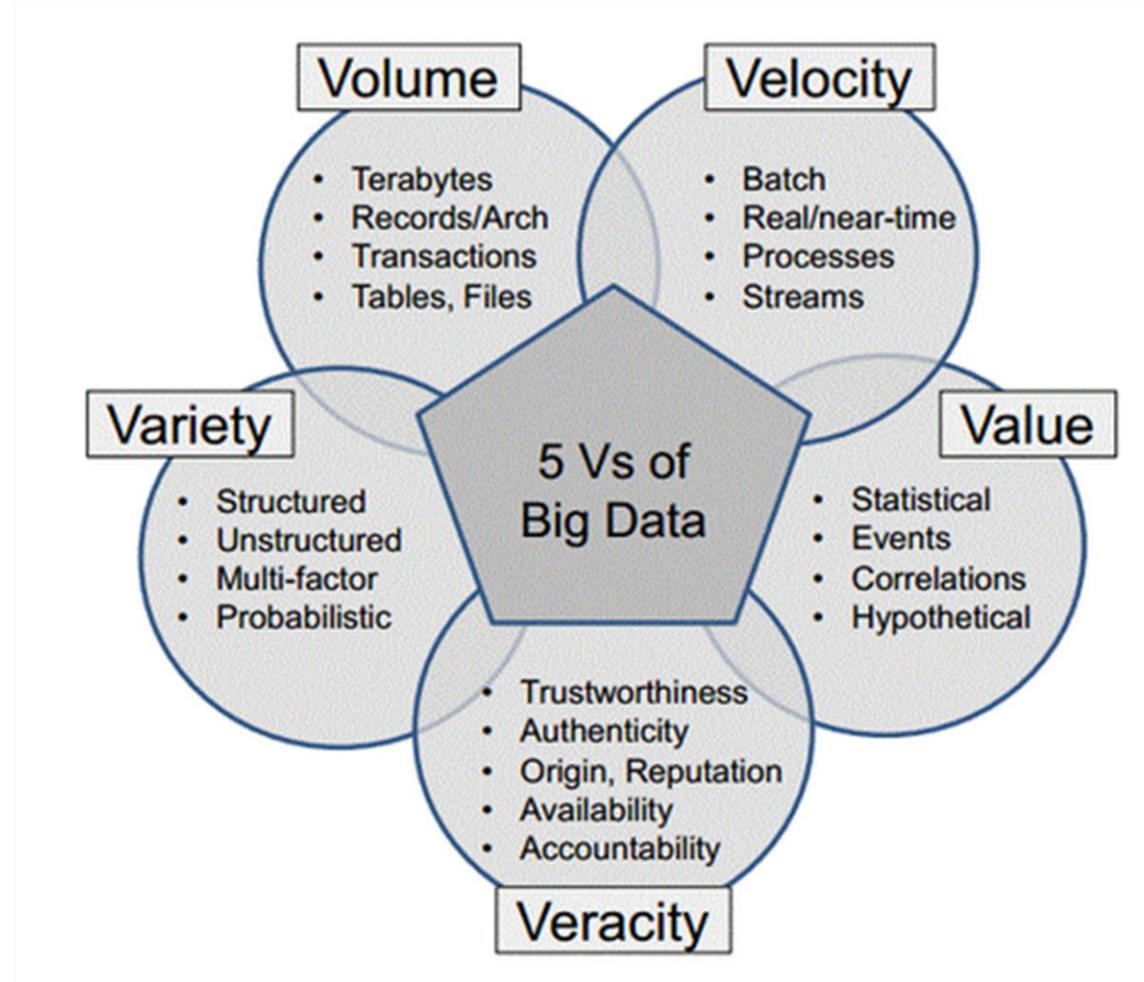
+

Velocidade



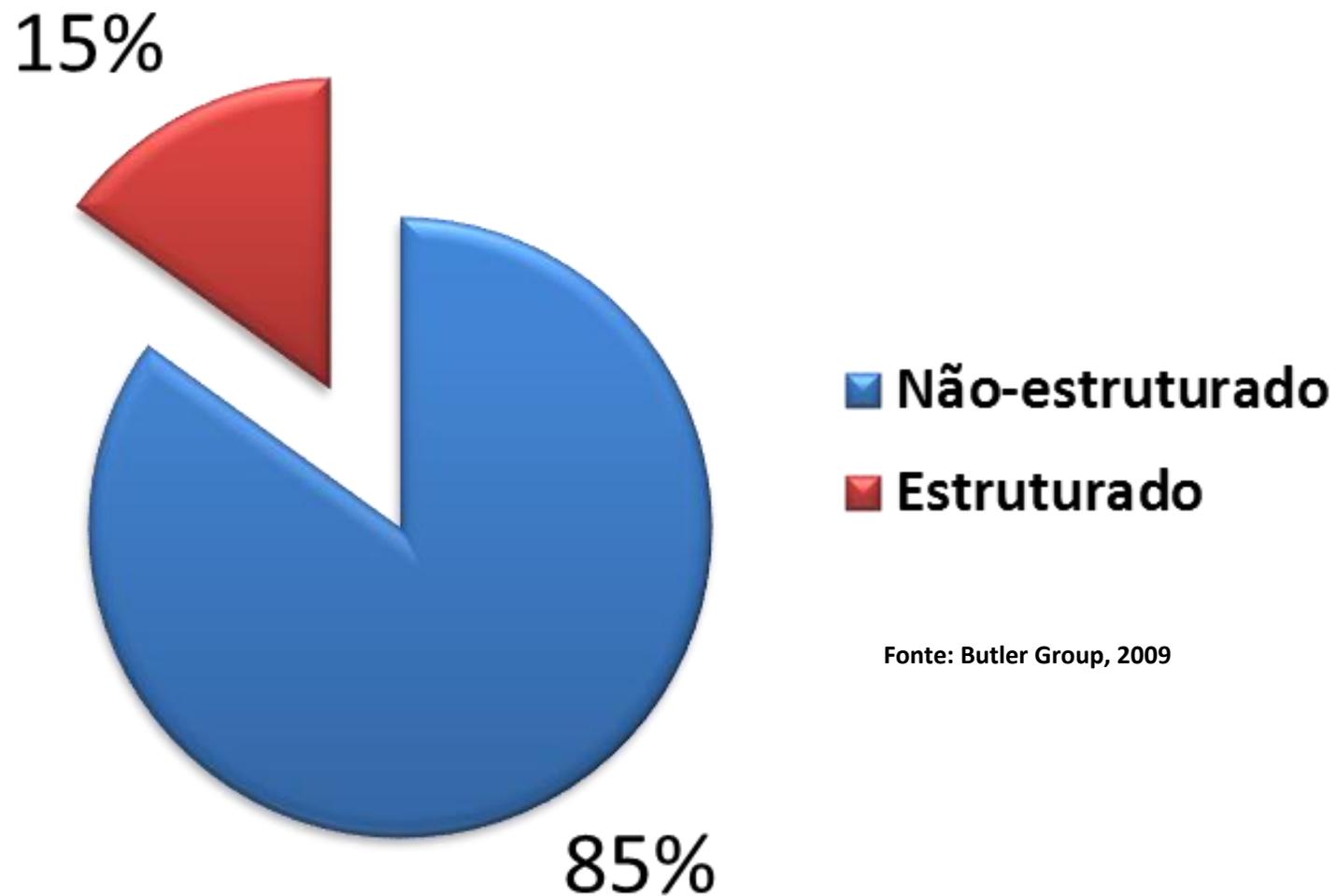
3 V's do Big Data

# PARA ALGUNS SÃO 5 Vs



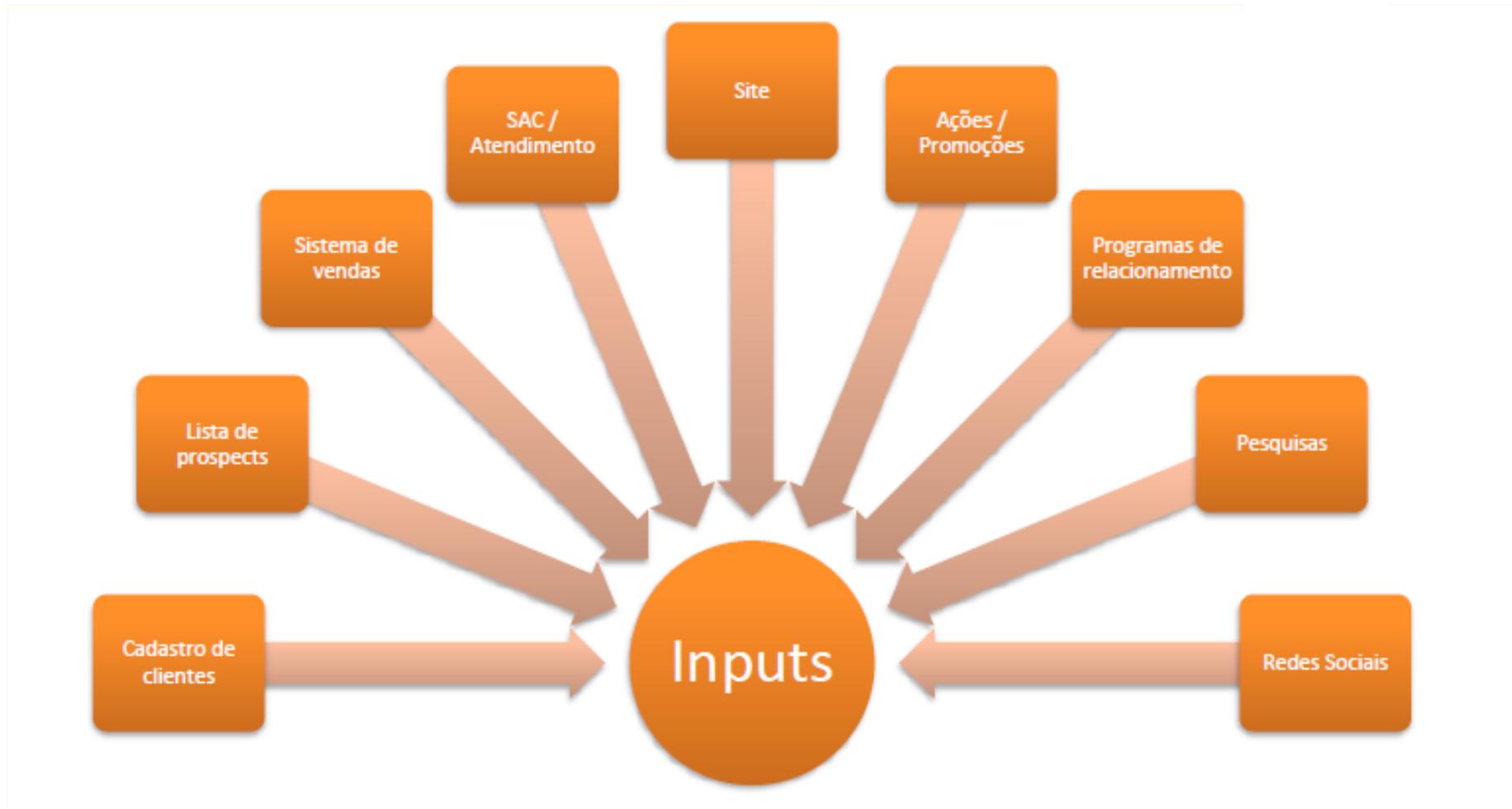


# TIPOS DE INFORMAÇÃO



Fonte: Butler Group, 2009

# FONTES DE INFORMAÇÃO EM EMPRESAS



# DATA SCIENTIST

Estatística/

Mas quem é esse  
profissional no  
mercado hoje?

Conhecimento de Negócio

Comunicação

## 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.

### PROGRAMMING DATABASE

Computer science fundamentals  
Programming language e.g. Python  
Statistical computing package e.g. R  
Databases SQL and NoSQL  
Linear algebra  
Distributed databases and parallel query  
Processing  
MapReduce concepts  
Hadoop and Hive/Pig  
MapReduce reducers  
Experience with xaaS like AWS

### COMMUNICATION VISUALIZATION

Ability to engage with senior  
Management  
Storytelling skills  
Ability to translate data-driven insights into  
Business decisions and actions

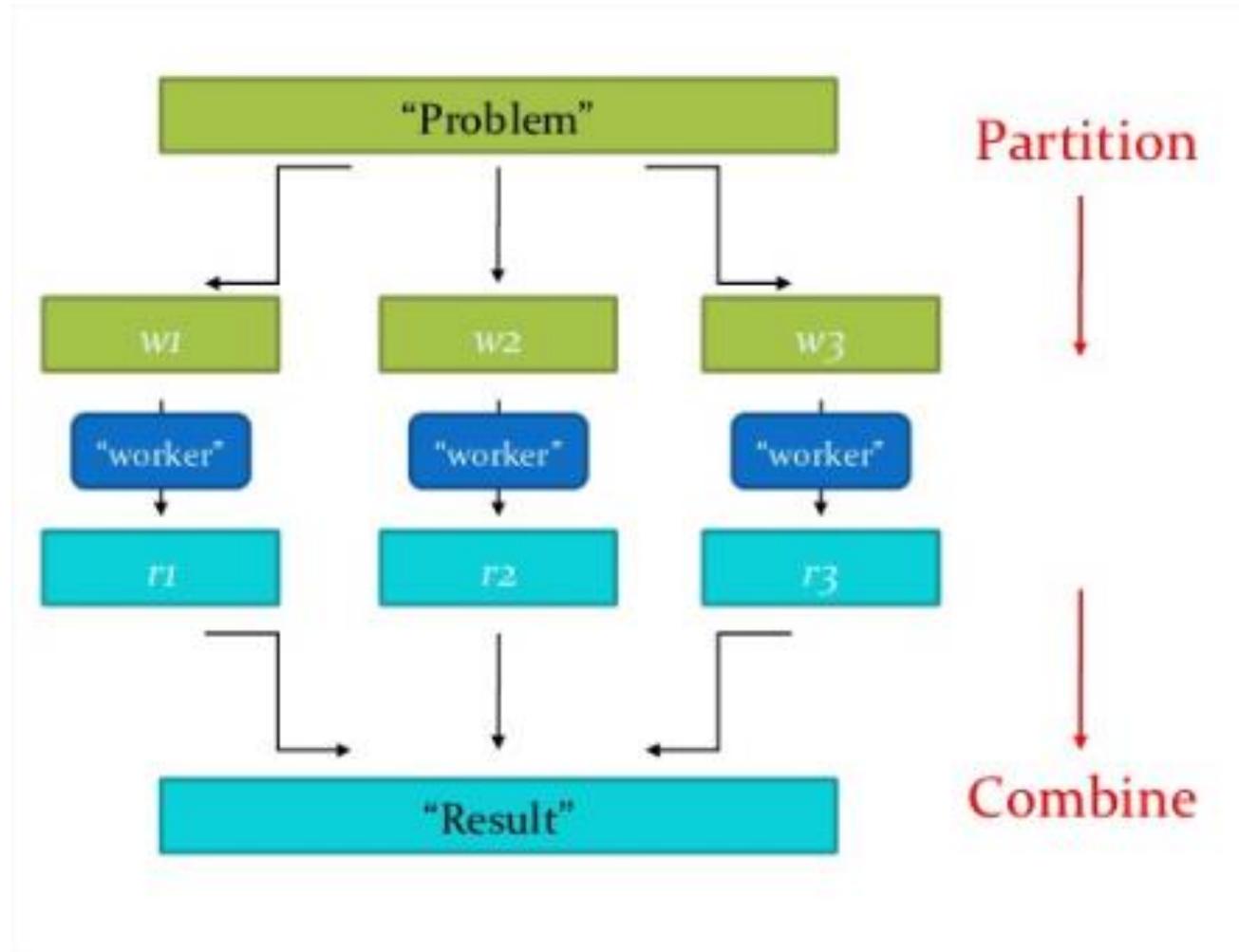
Visual art design

- ☆ Innovative and collaborative
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

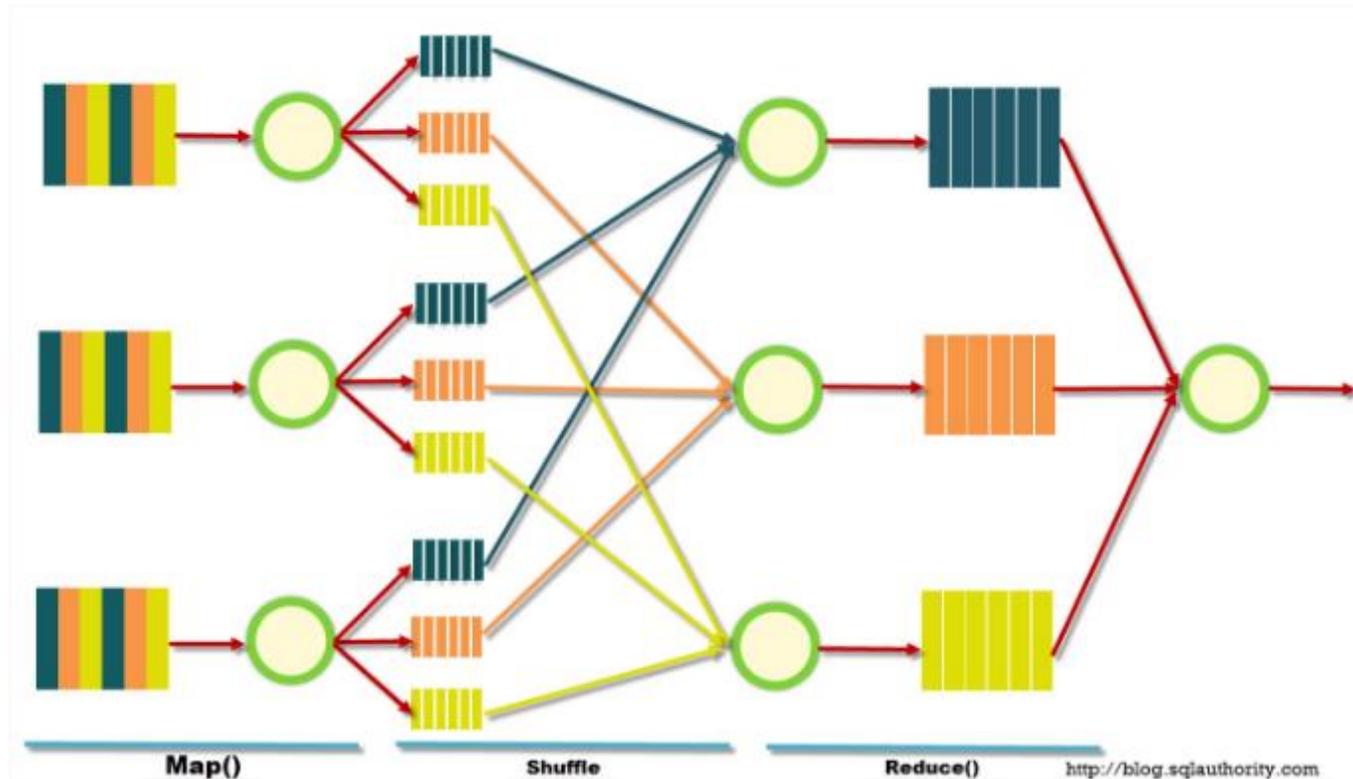


# DIVIDIR E CONQUISTAR

## COMPUTAÇÃO PARALELA



# MAP REDUCE



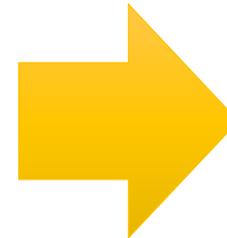
<https://www.youtube.com/watch?v=bcjSe0xCHbE>  
<https://www.youtube.com/watch?v=MfF750YVDxM>

# TENDÊNCIAS

## ❑ Big Data Framework



Armazenamento  
baseado  
em disco



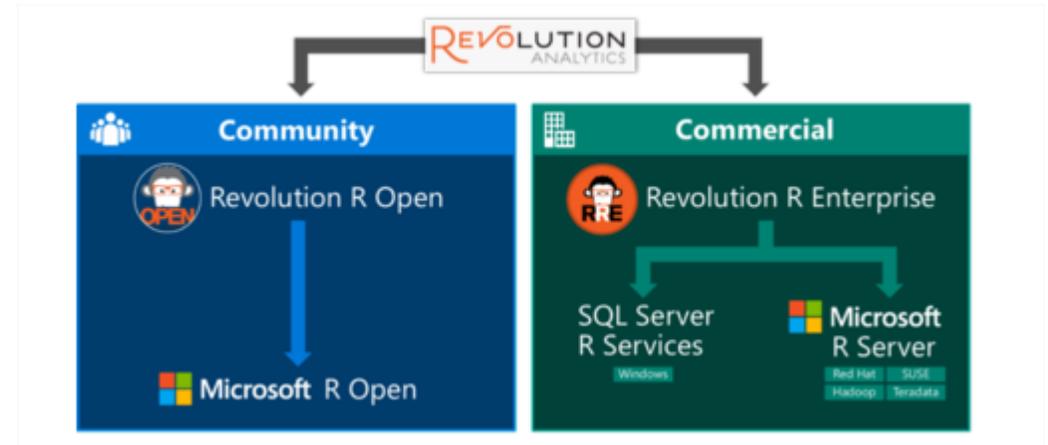
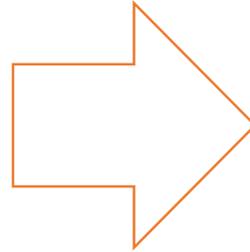
Aplicar os dois  
conjuntamente!



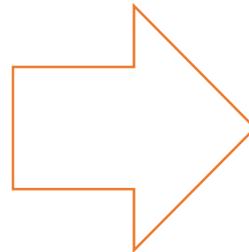
Armazenamento  
baseado  
em memória

# TENDÊNCIAS

Empresas com solução de banco de dados se criando soluções de analytics.



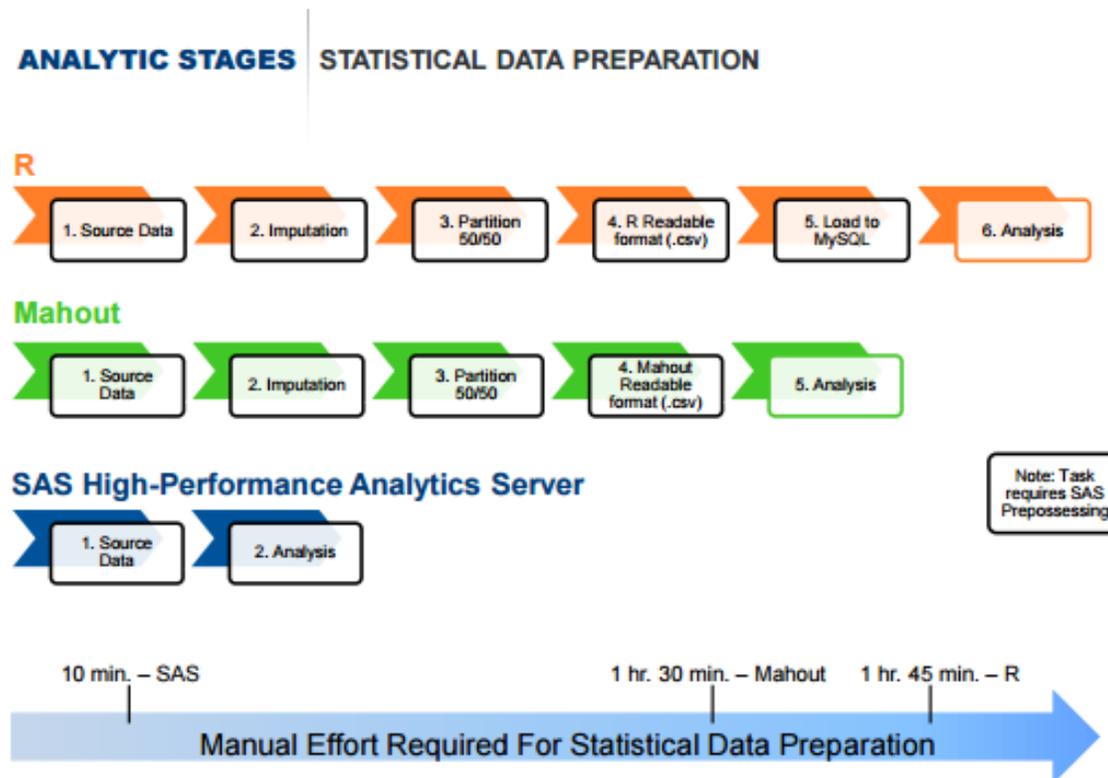
Microsoft Azure



An advertisement for R Technologies from Oracle. The text reads "R Technologies from Oracle" and "Bringing the Power of R to the Enterprise". Below the text is the R logo, which is a red circle containing a white "R" with a small "n" above it.

# TENDÊNCIAS

## Discurso comercial contra soluções Open Source



*White Paper  
da SAS*

MAS ESTÁ TECNOLOGIA  
É SEMPRE BOA? EM TODAS AS  
SITUAÇÕES?

**NÃO!**



# PRÓS E CONTRAS

## □ Prós

- Armazenamento distribuído
  - Agregação de servidores menores ao seu parque tecnológico
- Navegação simples
  - Independente da estrutura de dados (Json, ql, ...)
- Uso de linguagem próxima ao SQL (HIVE, exemplo)
- Permite transformação de dados (tratamento) em linguagens como PIG, JAVA, Python.
- Performance de grandes volumes de dados

## □ Contras

- Performance deficiente em bases de dados pequenas.
- Fonte de dados matricial já existente e mapeada.
- Dificuldade de conseguir mão de obra qualificada no mercado.
- Optando pelo desenvolvimento próprio, não existe suporte.
- Necessidade de adaptação de algoritmos desenvolvidos inicialmente para lidar com informações armazenadas em bancos de dados tradicionais.

# R E BIG DATA

## ☐ Pacotes de Big Data para o R



The screenshot shows the SparkR documentation page. The URL is <https://spark.apache.org/docs/latest/sparkr.html>. The page title is "SparkR (R on Spark)". The navigation menu includes "Overview" and "Programming Guides". The main content is a list of links for SparkR documentation:

- Overview
- SparkR DataFrames
  - Starting Up: SparkContext, SQLContext
  - Starting Up from RStudio
  - Creating DataFrames
    - From local data frames
    - From Data Sources
    - From Hive tables
  - DataFrame Operations
    - Selecting rows, columns
    - Grouping, Aggregation
    - Operating on Columns
  - Running SQL Queries from SparkR
- Machine Learning
  - Gaussian GLM model
  - Binomial GLM model
- R Function Name Conflicts
- Migration Guide
  - Upgrading From SparkR 1.5.x to 1.6

<https://www.youtube.com/watch?v=nrmASvsU-lo>

<https://www.youtube.com/watch?v=wGmg6k3wGUg>

### Hadoop InteractiVE

Hadoop InteractiVE facilitates distributed computing via the MapReduce paradigm through R and Hadoop. An easy to use interface to Hadoop, the Hadoop Distributed File System (HDFS), and Hadoop Streaming is provided.

### bigmemory: Manage Massive Matrices with Shared Memory and Memory-Mapped Files

Create, store, access, and manipulate massive matrices. Matrices are allocated to shared memory and may use memory-mapped files. Packages 'biganalytics', 'bigtabulate', 'synchronicity', and 'bigalgebra' provide advanced functionality.

### biglm: bounded memory linear and generalized linear models

Regression for data too large to fit in memory ty.

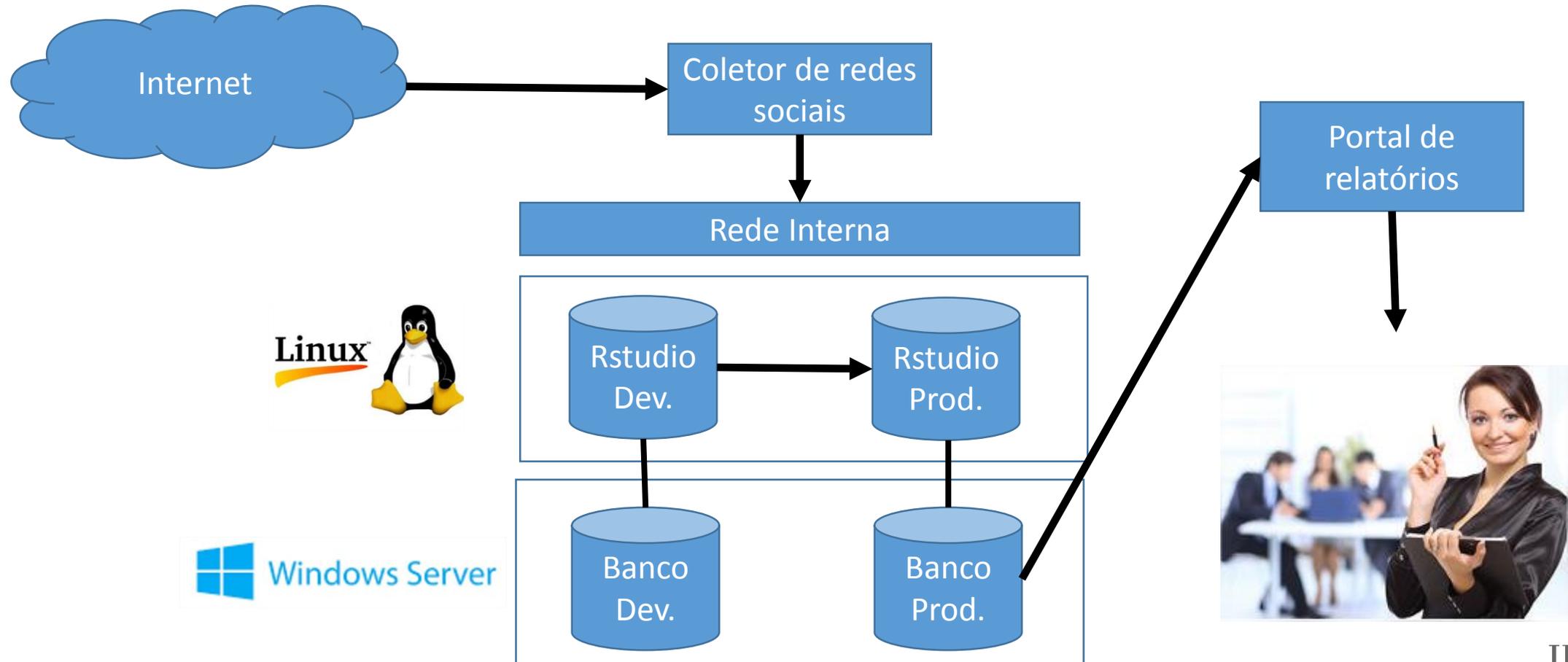
### ff: memory-efficient storage of large data on disk and fast access functions

### snow: Simple Network of Workstations

Support for simple parallel computing in R.

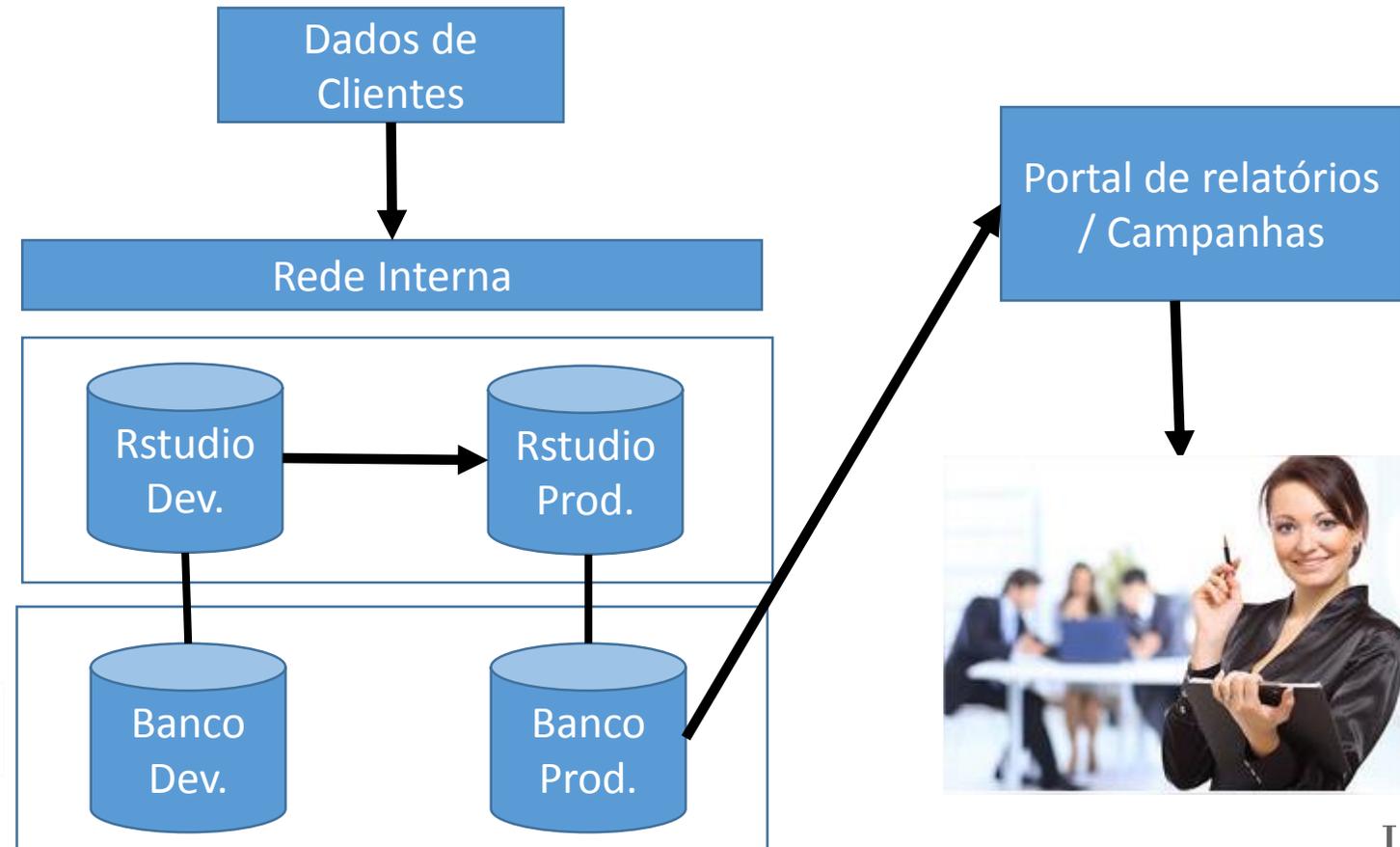
# COMO TRABALHAMOS NO IBOPE DTM?

Projeto  
Politics



# COMO TRABALHAMOS NO IBOPE DTM?

Projetos de  
CRM



# EM QUE USAMOS R

## Serviços (Consultoria)

- Modelos de Retenção
- Modelos de Valor do Cliente (CLV)
- Análise de Cluster
- Modelos de propensão a compra de produtos/serviços
- Modelos de inadimplência
- Estatística Espacial
- Modelos de MMM (***Marketing Mix Modeling***)
- Modelos de previsão (junto ao Tableau)

## Produtos

- Sistemas de Recomendação
- PeopleScope

# PEOPLESCOPE

❑ Algoritmo de Cluster desenvolvido em R

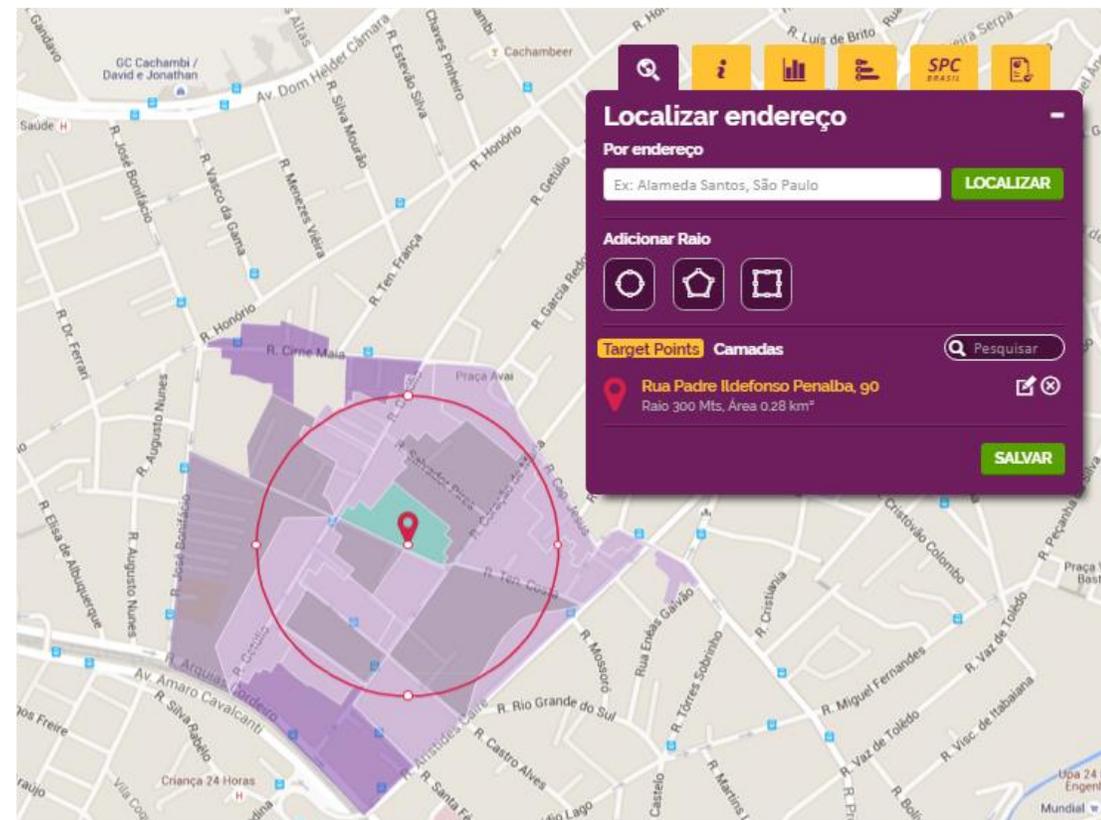
Ferramenta de negócio que possibilita a análise de *segmentos populacionais geolocalizados* identificados com base em informações sociodemográficas e pesquisas comportamentais/atitudinais.

**Target Group Index**

**IBOPE BUS**



**PNAD** Contínua  
Pesquisa Nacional por Amostra de Domicílios Contínua



# SISTEMA DE RECOMENDAÇÃO

□ Algoritmos desenvolvidos em R



# POLITICS

## ❑ Algoritmo de classificação desenvolvido em R

### ***Etapa 1***

Desenvolvimento de metodologia para polarização manual por doutores em **Língua Portuguesa**.

### ***Etapa 2***

Polarização manual de amostra significativa de postagens pelos especialistas.

### ***Etapa 3***

Elaboração de modelo estatístico para polarização automática de postagens do Twitter sobre política pelos especialistas em **Big Data Analytics** do IBOPE DTM.

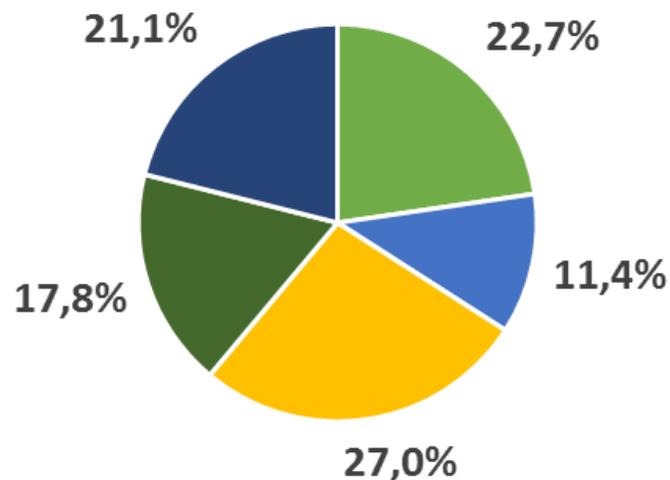
### ***Etapa 4***

**Processo Contínuo:** Polarização manual de novas amostras, recalibragem do algoritmo desenvolvido com novas informações.

# POLITICS

## ☐ Análise de tópicos em R

Representatividade dos Tópicos



■ Tópico 1 ■ Tópico 2 ■ Tópico 3 ■ Tópico 4 ■ Tópico 5

Onda 9: Tópicos				
1	2	3	4	5
aecioneves	40	dilmabr	marina	marina
dilmabr	silvamarina	chegaderouba...	silva	turno
45aacioconfir...	domingo	presidenta	luciana	presidente
corrupção	votar	vai	pra	neves
pt	conheça	novo	levy	aécio
sobre	razões	marina	neves	aecioneves
aécio	httpcoaz...	pastormalafaia	pergunta	ser
petrobras	neste	pra	candidato	silva
diz	vou	pt	pastor	segundo
correios	brasilmarina40	inflação	aécio	pode
minas	marina40	13	everaldo	debate
neves	dia	controle	corrupção	pois
educação	fazer	dilma	noite	pesquisa
fhc	dias	presidente	falar	globo
frase	econômica	povo	vai	qualquer

**Tópico 1:** Contém predominantemente postagens de apoio ao candidato Aécio Neves, denunciando escândalos envolvendo o partido da candidata petista.

**Tópico 2:** Estiveram mais presentes as razões para se optar pela candidata Marina Silva

**Tópico 3:** Dilma aparece como figura central, sendo o controle da inflação a argumentação de destaque em de defesa, aparecendo também fortes negativas a sua candidatura.

**Tópico 4:** Aparecem postagens relacionadas a outros candidatos, tais como Luciana, Levy e Pastor Everaldo.

**Tópico 5:** Relaciona principalmente postagens especulativas sobre o resultado das eleições, mencionando também o debate da Rede Globo.

# VAGA PARA SP

Analista Júnior de Inteligência de Marketing

**Formação Concluída**

✓ Estatística, Física, Matemática ou Engenharia

**Conhecimento Técnico**

✓ R, SQL Server e Pacote Office

**Local em São Paulo**

Região da Paulista (Alameda Santos, 2.101)

Enviar currículo para [rodrigo.ribeiro@ibopedtm.com](mailto:rodrigo.ribeiro@ibopedtm.com)

Obrigado!

**IBOPE dtm**

[www.ibopedtm.com](http://www.ibopedtm.com)

**SP** 11 3069-3730

**RJ** 21 3439-1777

**RS** 51 3382-3300