



Estudo da Ferramenta Hadoop

Marcos Antonio de Souza,
Thaynara Mábile Marques Ribeiro



Motivação



GLOBAL INTERNET POPULATION GROWTH 2012-2016
(IN BILLIONS)

SOURCES: SNAPCHAT, NETFLIX, GOOGLE, INSTAGRAM, TINDER, THE WEATHER COMPANY, DROPBOX, GITHUB, GIPHY, YOUTUBE, BUZZFEED, AMAZON, CTV, MARY MECKERS 2016 INTERNET TRENDS REPORT, USA TODAY, GLOBAL WEB INDEX

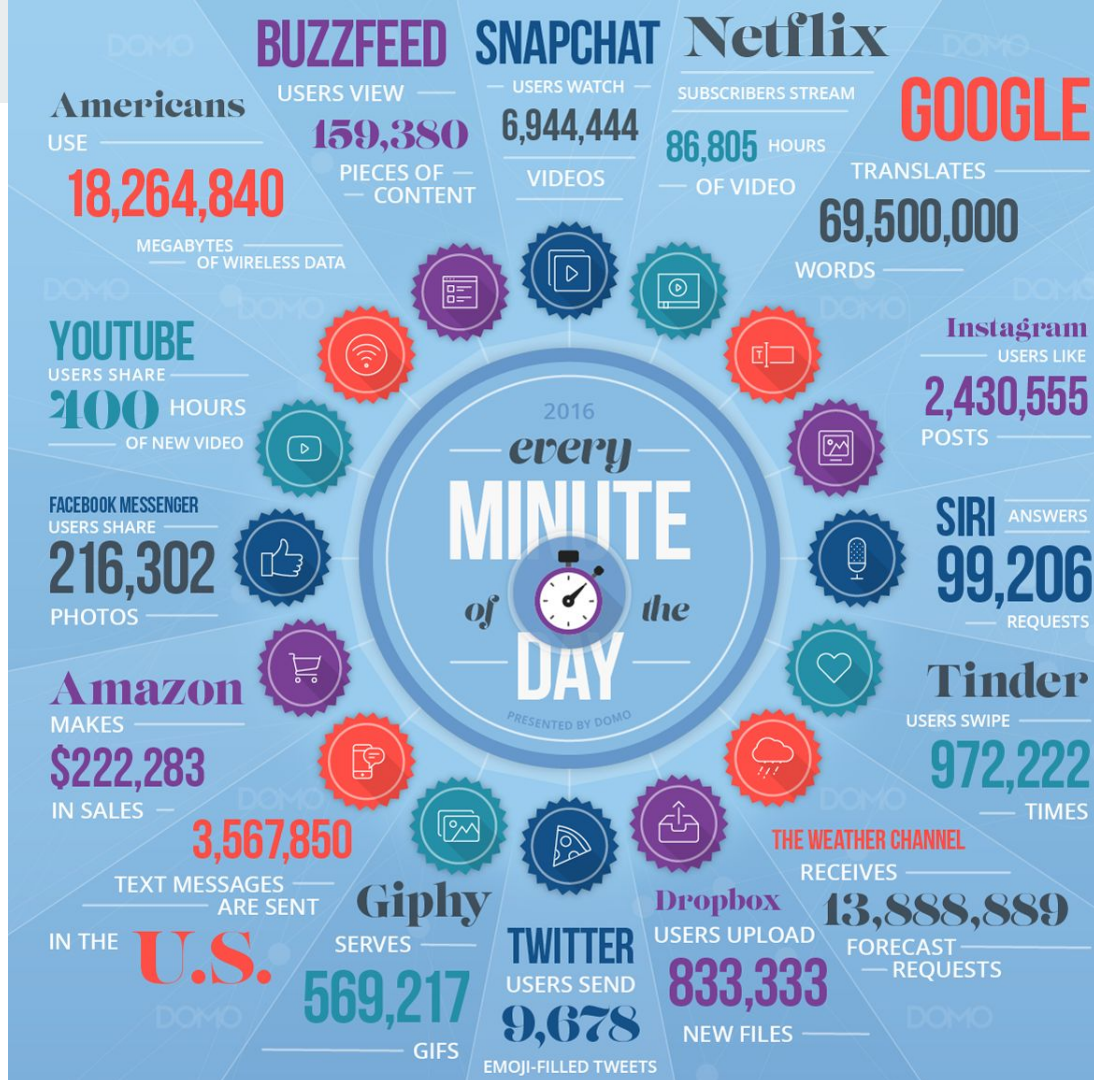
Data has become the new enterprise currency. The ability to collect, analyze, and leverage it effectively will distinguish the best from the rest. Domo helps you stay ahead by bringing your data and people together in the cloud, where everyone in your organization can easily access the information they need to make faster, better-informed decisions and optimize business performance.

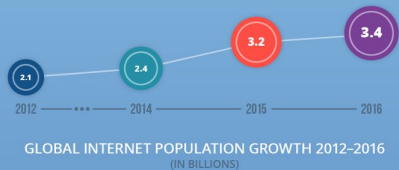
Learn more at www.domo.com



DATA NEVER SLEEPS 4.0

How much data is generated every minute? In the fourth annual edition of Data Never Sleeps, newcomers like Giphy and Facebook Messenger illustrate the rise of our multimedia messaging obsession, while veterans like YouTube and Snapchat highlight our insatiable appetite for video. Just how many GIFs, videos, and emoji-filled Tweets flood the Internet every minute? See for yourself below.





Data has become the new enterprise currency. The ability to collect, analyze, and leverage it effectively will distinguish the best from the rest. Domo helps you stay ahead by bringing your data and people together in the cloud, where everyone in your organization can easily access the information they need to make faster, better-informed decisions and optimize business performance.

Learn more at www.domo.com

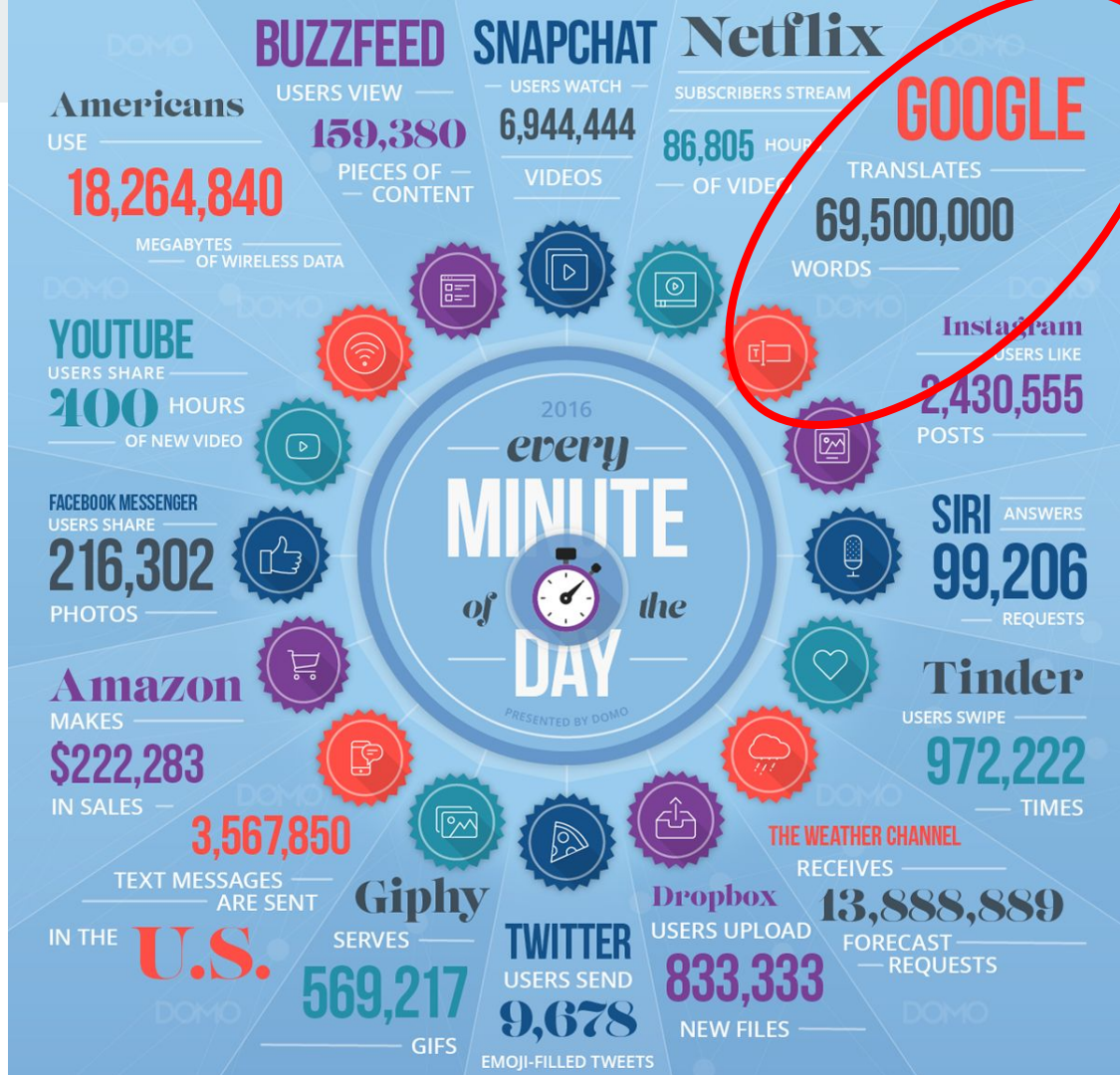


SOURCES: SNAPCHAT, NETFLIX, GOOGLE, INSTAGRAM, TINDER, THE WEATHER CHANNEL, DROPBOX, GITHUB, GIPHY, YOUTUBE, BUZZFEED, AMAZON, CTA, MARY MECKERS 2016 INTERNET TRENDS REPORT, USA TODAY, GLOBAL WEB INDEX



DATA NEVER SLEEPS 4.0

How much data is generated every minute? In the fourth annual edition of Data Never Sleeps, newcomers like Giphy and Facebook Messenger illustrate the rise of our multimedia messaging obsession, while veterans like YouTube and Snapchat highlight our insatiable appetite for video. Just how many GIFs, videos, and emoji-filled Tweets flood the Internet every minute? See for yourself below.



The world internet population has grown 75% from 2012 and now represents 3.7 billion people.



GLOBAL INTERNET POPULATION GROWTH 2012-2017
(IN BILLIONS)

SOURCES: EXPANDED RAILINGS.COM, WEARESOCIAL.COM, WIKIPEDIA, FORBES, ADWEEK.COM, FORTUNE.COM, BLOOMBERG.COM, ONEREACTION.COM, IBM, BUZZFEED, INTERNET LIVE STATS, INTERNET WORLD STATS, BBC



With each click, swipe, share, and like, businesses are using data to make decisions about the future. Domo gives everyone in your business real-time access to data from virtually any data source in a single platform for smarter decision-making at any moment.

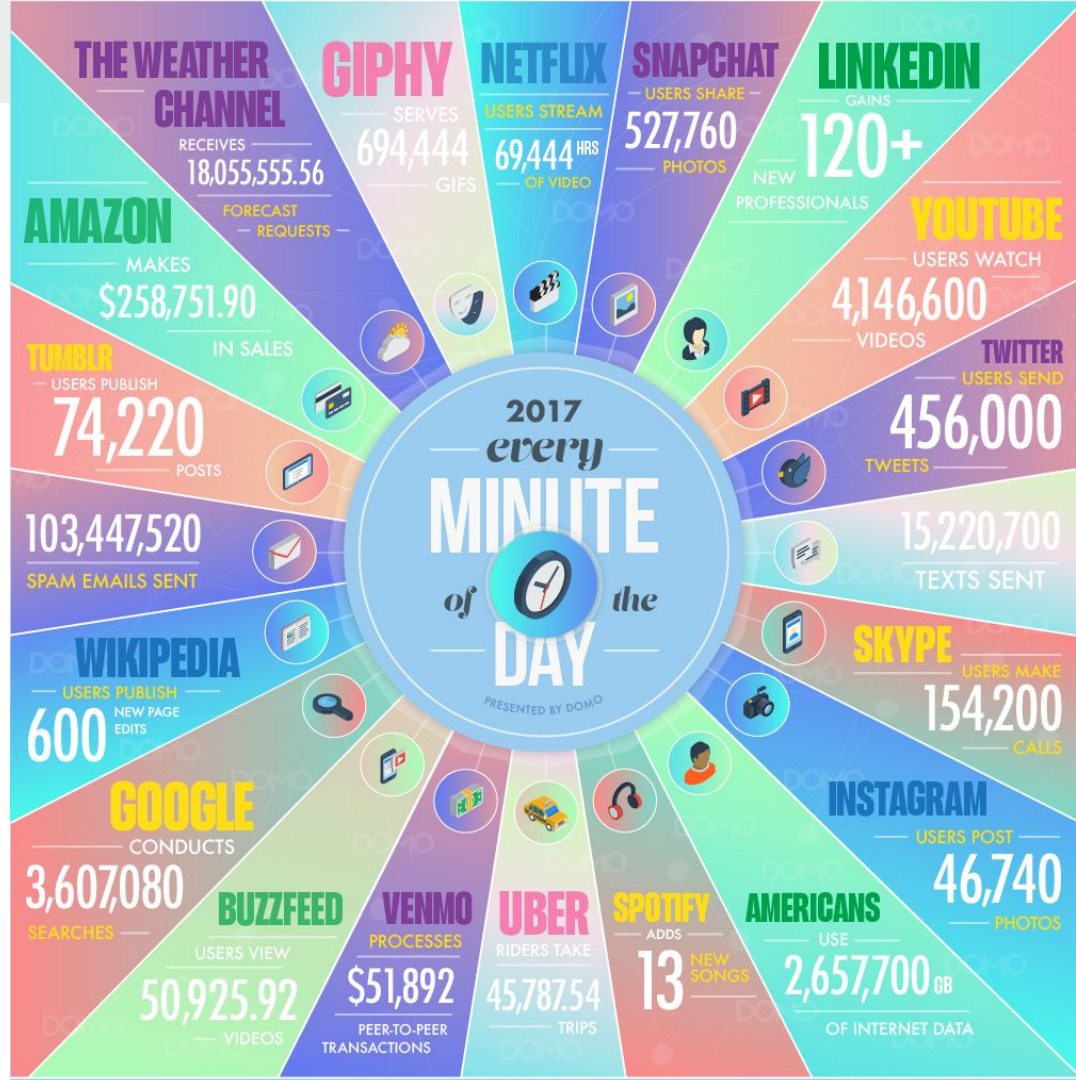
Learn more at domo.com



DATA NEVER SLEEPS 5.0

How much data is generated *every minute*?

90% of all data today was created in the last two years—that's 2.5 quintillion bytes of data per day. In our 5th edition of Data Never Sleeps, we bring you the latest stats on just how much data is being created in the digital sphere—and the numbers are staggering.



The world internet population has grown 75% from 2012 and now represents 3.7 billion people.



GLOBAL INTERNET POPULATION GROWTH 2012-2017
(IN BILLIONS)

SOURCES: EXPANDED RAILINGS.COM, WEARESOCIAL.COM, WIKIPEDIA, FORBES, ADWEEK.COM, FORTUNE.COM, BLOOMBERG.COM, ONEREACH.COM, IBM, BUZZFEED, INTERNET LIVE STATS, INTERNET WORLD STATS, BBC



With each click, swipe, share, and like, businesses are using data to make decisions about the future. Domo gives everyone in your business real-time access to data from virtually any data source in a single platform for smarter decision-making at any moment.

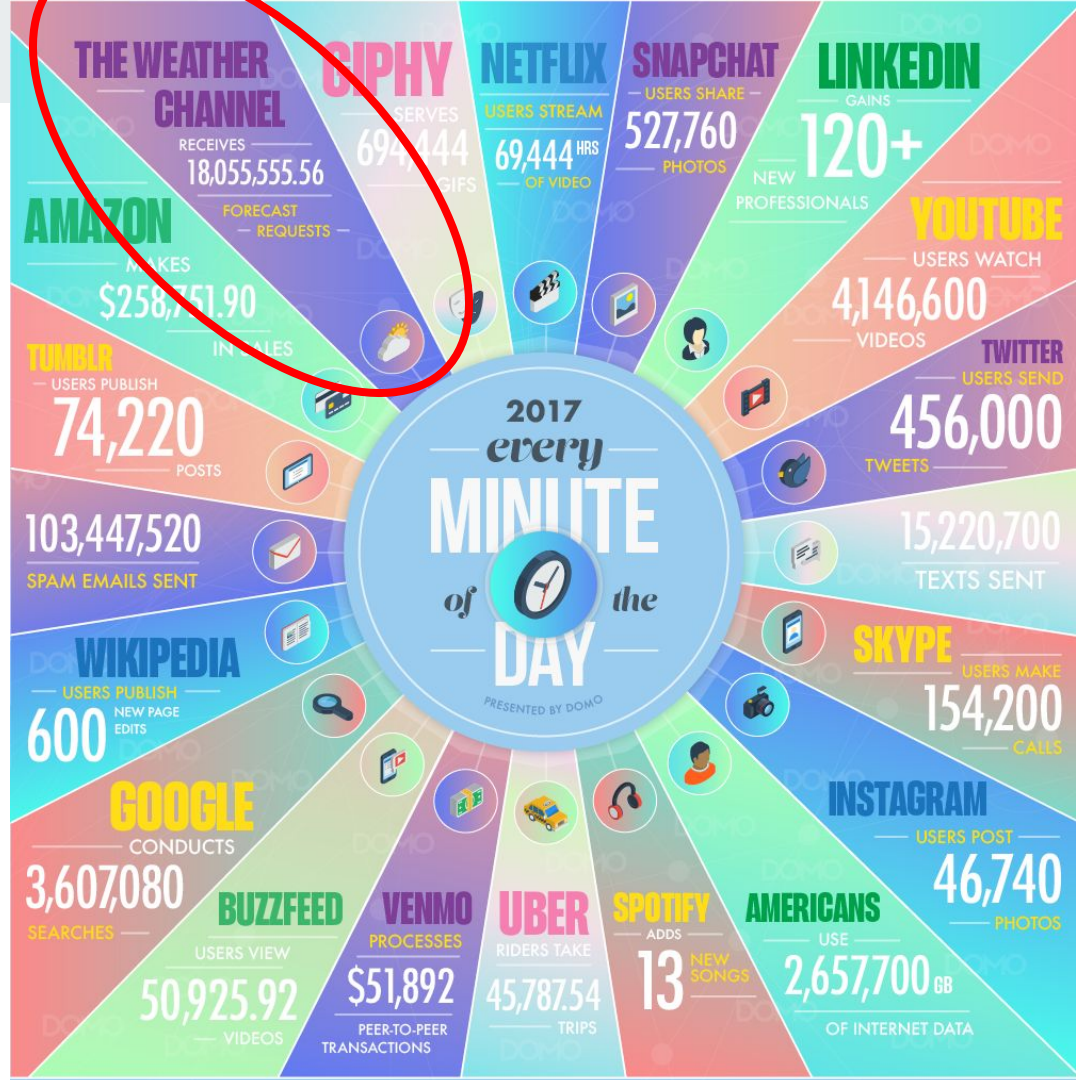
Learn more at domo.com



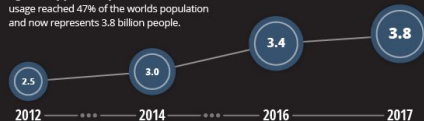
DATA NEVER SLEEPS 5.0

How much data is generated *every minute*?

90% of all data today was created in the last two years—that's 2.5 quintillion bytes of data per day. In our 5th edition of Data Never Sleeps, we bring you the latest stats on just how much data is being created in the digital sphere—and the numbers are staggering.



The world's internet population is growing significantly year-over-year. In 2017, internet usage reached 47% of the world's population and now represents 3.8 billion people.



GLOBAL INTERNET POPULATION GROWTH 2012-2017
(IN BILLIONS)

SOURCES: STATISTA, LINKEDIN, INTERNET LIVE STATS, EXPANDED RAMBLINGS, SLASH FILM, RIAA, BUSINESS OF APPS, INTERNATIONAL TELECOMMUNICATIONS UNION, INTERNATIONAL DATA CORPORATION

The ability to make data-driven decisions is crucial to any business. With each click, swipe, share, and like, a world of valuable information is created. Domo puts the power to make those decisions right into the palm of your hand by connecting your data and your people at any moment, on any device, so they can make the kind of decisions that make an impact.

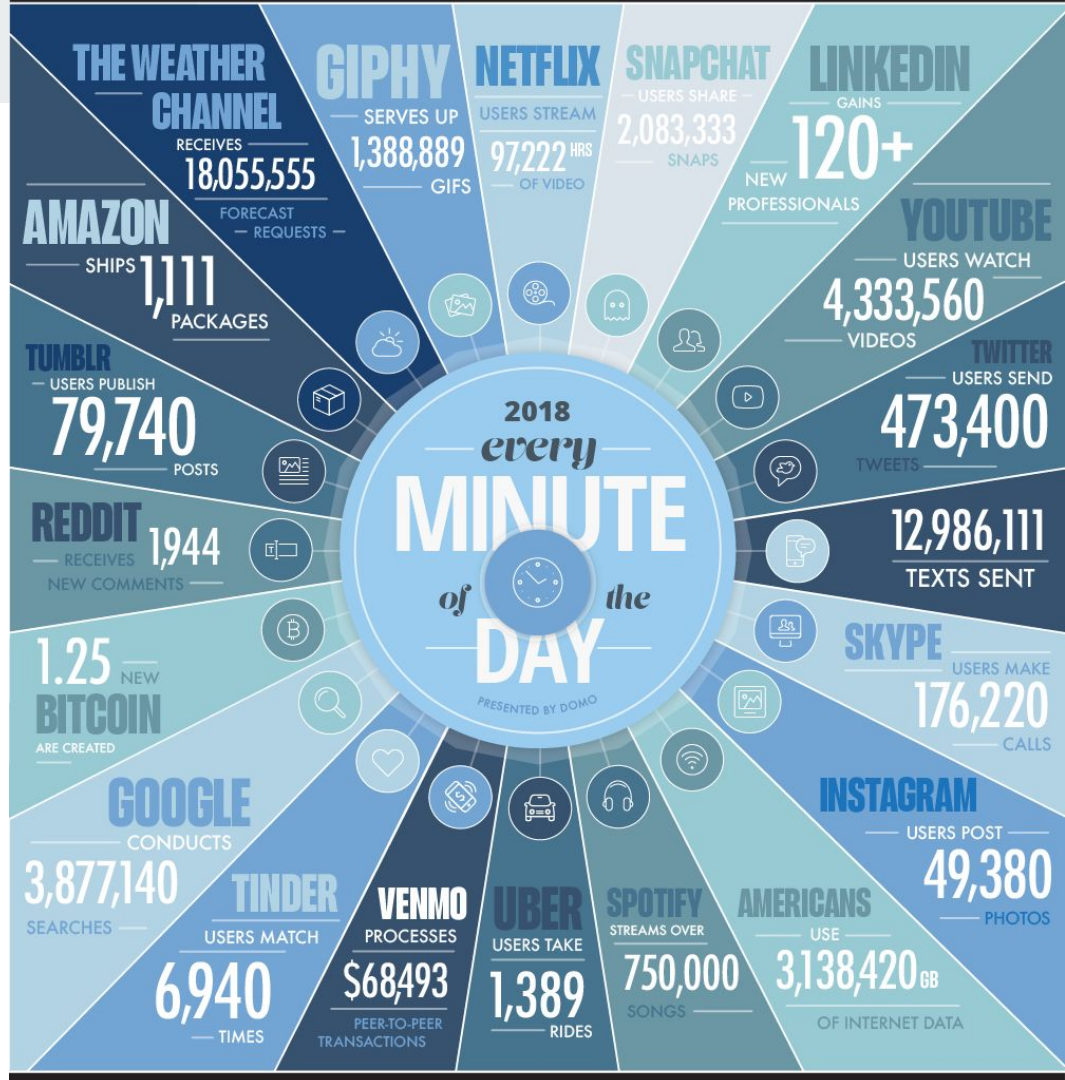
Learn more at domo.com



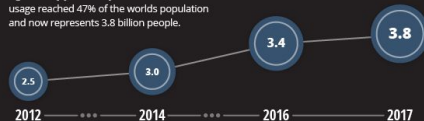
DATA NEVER SLEEPS 6.0

How much data is generated *every minute*?

There's no way around it: big data just keeps getting bigger. The numbers are staggering, but they're not slowing down. By 2020, it's estimated that for every person on earth, 1.7 MB of data will be created every second. In our 6th edition of Data Never Sleeps, we once again take a look at how much data is being created all around us every single minute of the day—and we have a feeling things are just getting started.



The world's internet population is growing significantly year-over-year. In 2017, internet usage reached 47% of the world's population and now represents 3.8 billion people.



GLOBAL INTERNET POPULATION GROWTH 2012-2017
(IN BILLIONS)

SOURCES: STATISTA, LINKEDIN, INTERNET LIVE STATS, EXPANDED RAMBLINGS, SLASH FILM, RIAA, BUSINESS OF APPS, INTERNATIONAL TELECOMMUNICATIONS UNION, INTERNATIONAL DATA CORPORATION

The ability to make data-driven decisions is crucial to any business. With each click, swipe, share, and like, a world of valuable information is created. Domo puts the power to make those decisions right into the palm of your hand by connecting your data and your people at any moment, on any device, so they can make the kind of decisions that make an impact.

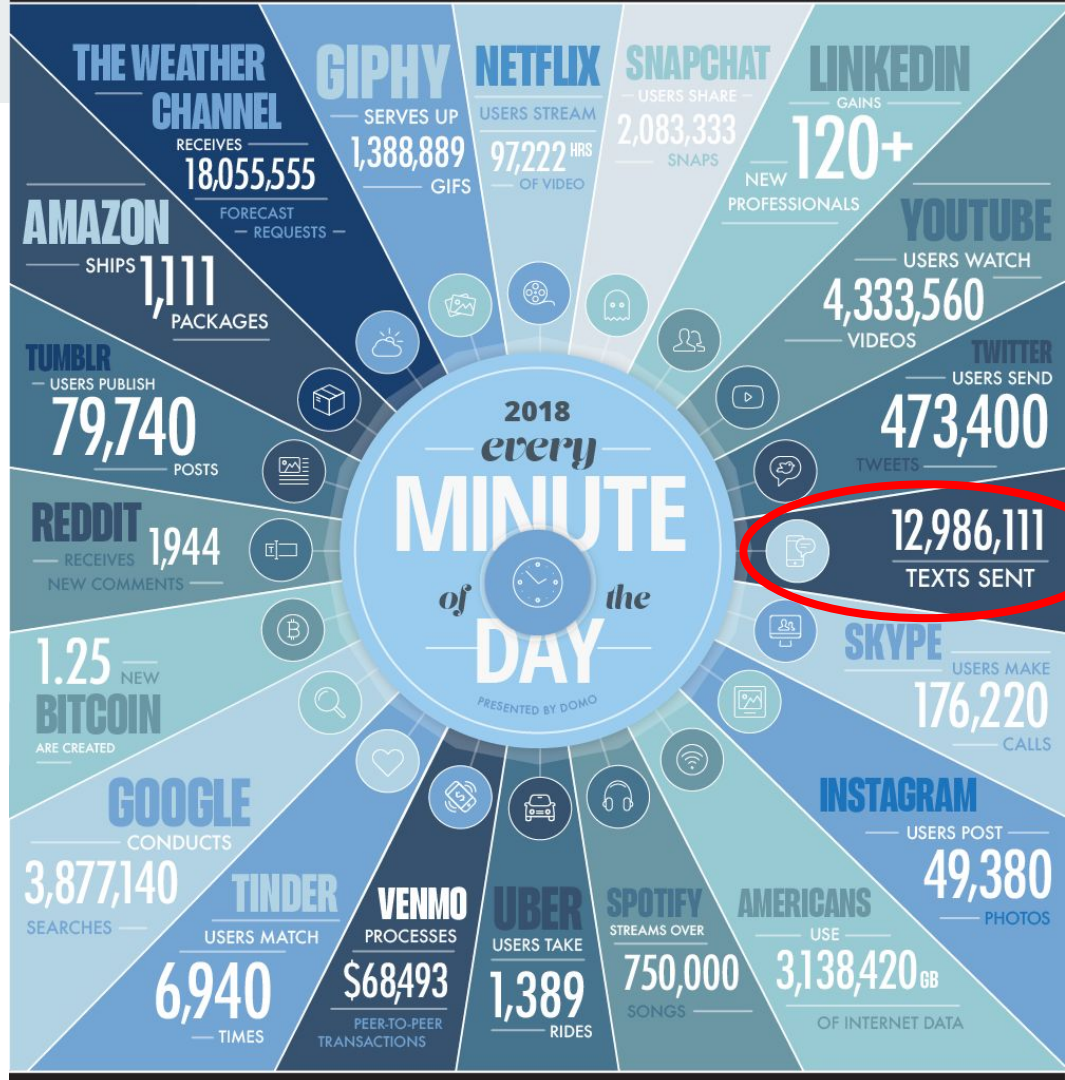
Learn more at domo.com



DATA NEVER SLEEPS 6.0

How much data is generated *every minute*?

There's no way around it: big data just keeps getting bigger. The numbers are staggering, but they're not slowing down. By 2020, it's estimated that for every person on earth, 1.7 MB of data will be created every second. In our 6th edition of Data Never Sleeps, we once again take a look at how much data is being created all around us every single minute of the day—and we have a feeling things are just getting started.



The world's internet population is growing significantly year-over-year. As of January 2018, the internet reaches 36.1% of the world's population and now represents 4.39 billion people — a 9% increase from January 2017.



GLOBAL INTERNET POPULATION GROWTH 2012-2018
(IN BILLIONS)

SOURCES: STATISTA, INTERNET LIVE STATS, EXPANDED RAMBLINGS, NATIONAL ASSOCIATION OF CITY TRANSPORTATION OFFICIALS, WIRED



The ability to make data-driven decisions is crucial to any business. With each click, swipe, share, and like, a world of valuable information is created. Domo puts the power to make those decisions right into the palm of your hand by connecting your data and your people at any moment, on any device, so they can make the kind of decisions that make an impact.

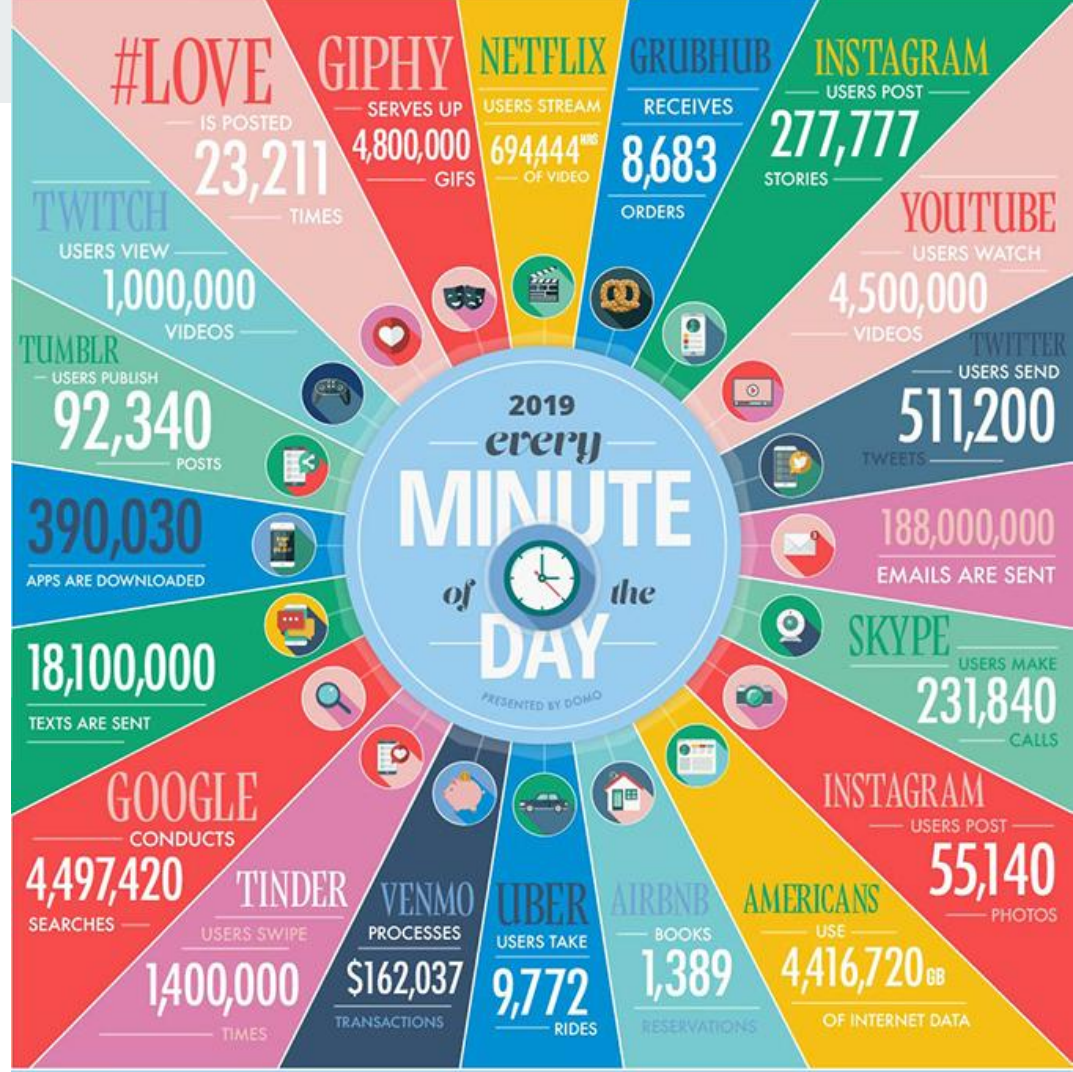
Learn more at domo.com



DATA NEVER SLEEPS 7.0

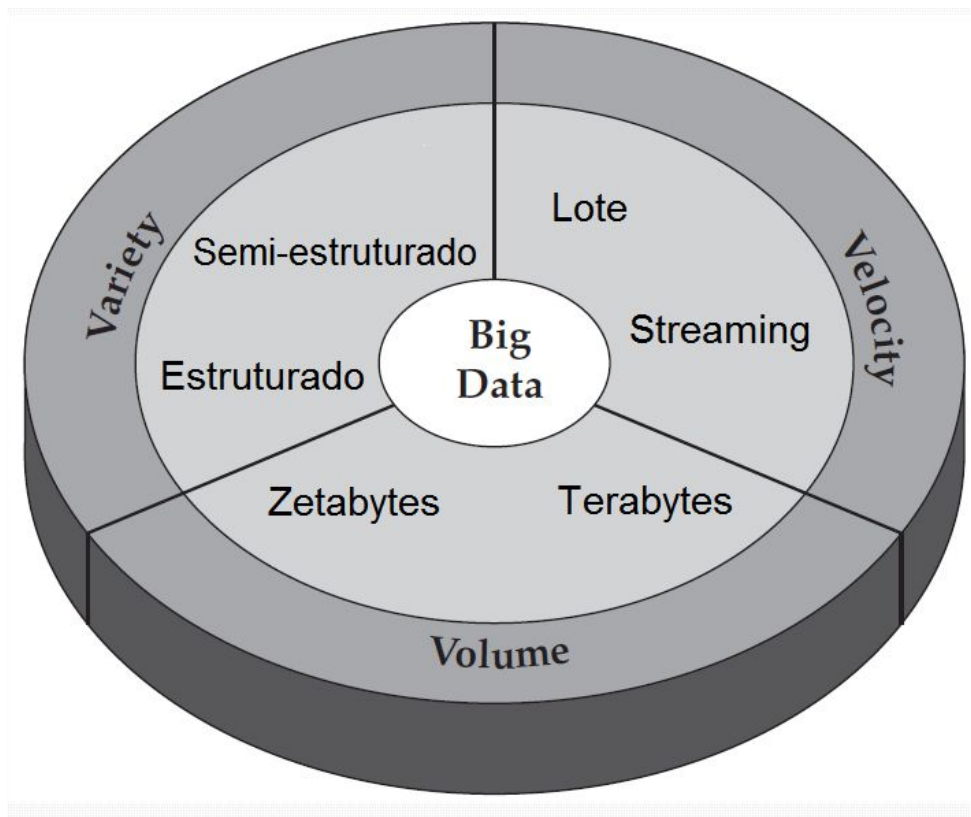
How much data is generated *every minute*?

There's no way around it: big data just keeps getting bigger. The numbers are staggering, and they're not slowing down. By 2020, there will be 40x more bytes of data than there are stars in the observable universe. In our 7th edition of Data Never Sleeps, we bring you the latest stats on how much data is being created in every digital minute — and the numbers are staggering.





Contextualização






Histórico



1999

Apache Software
Foundation (ASF)
formed as a non-profit

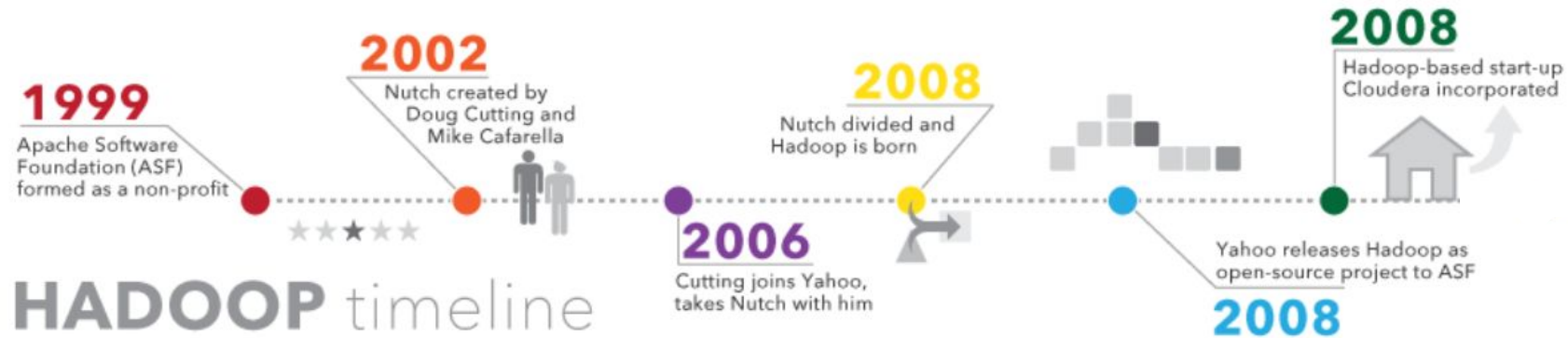


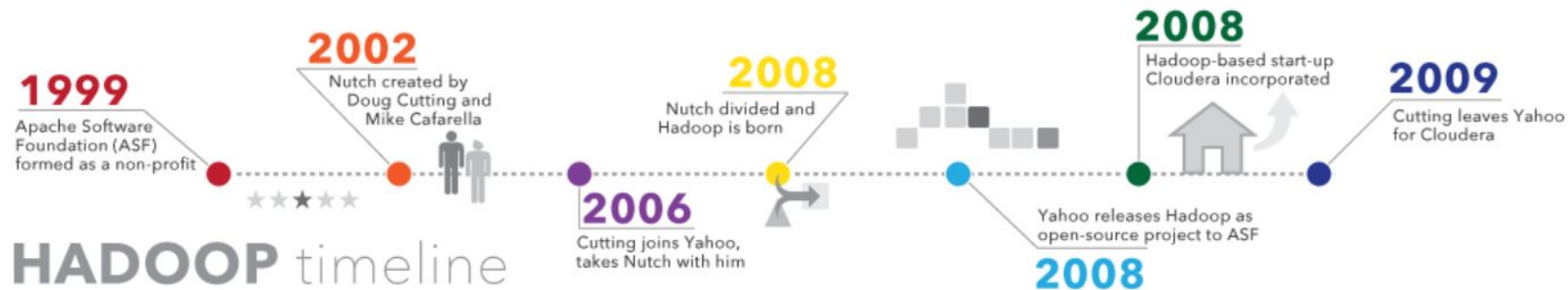
HADOOP timeline

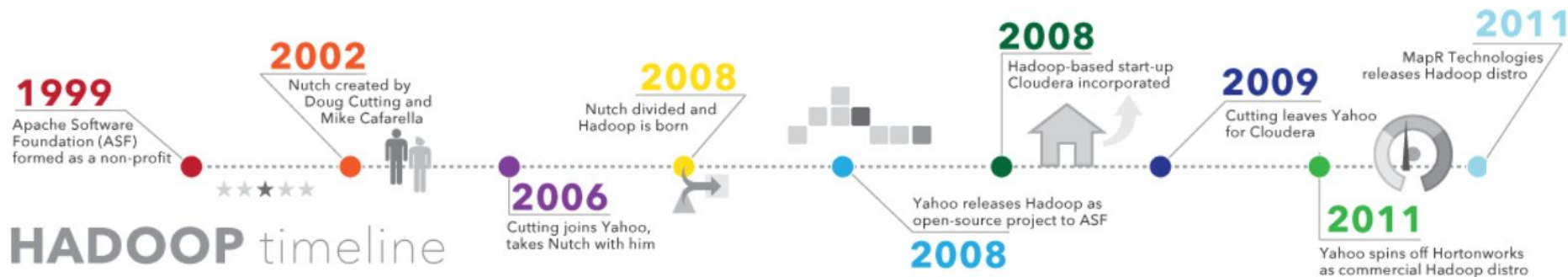


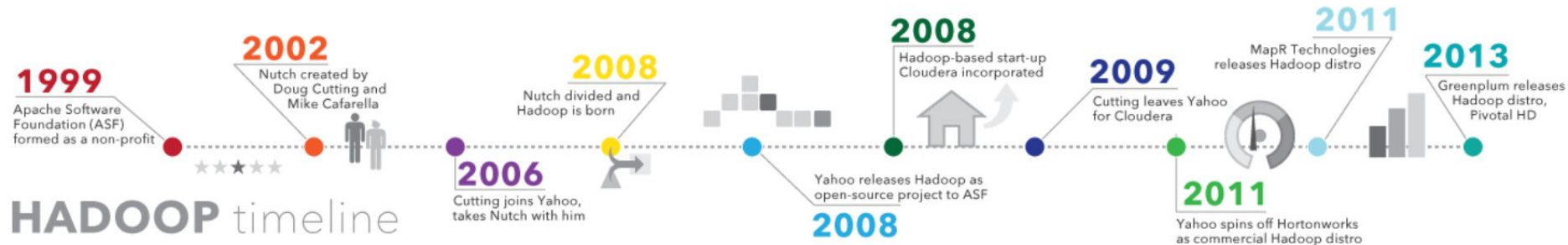
HADOOP timeline













Arquitetura

Aplicações Hadoop

Map Reduce

Distributed Processing Framework

HDFS

Hadoop Distributed File System

Aplicações Hadoop

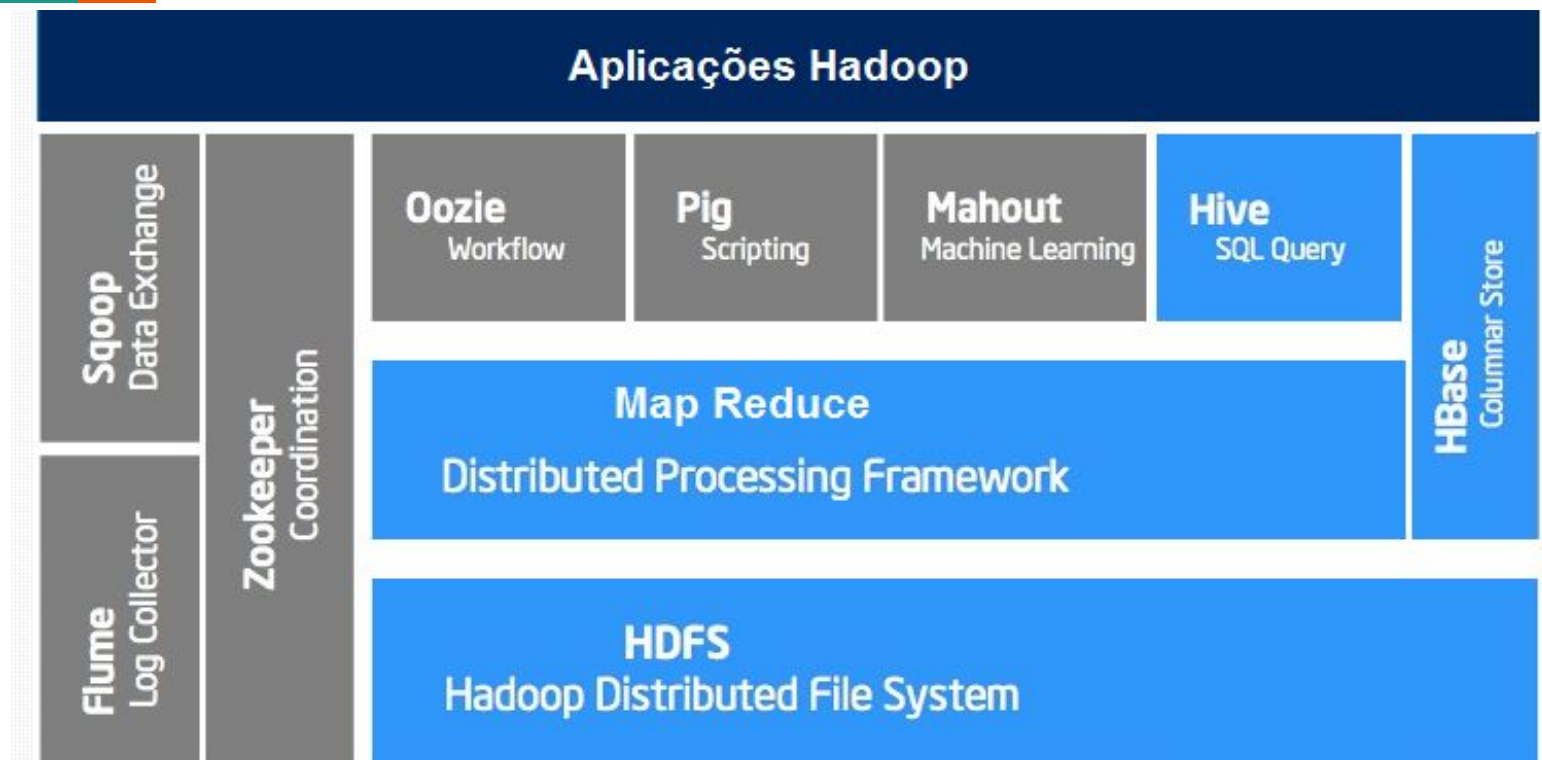
Hive
SQL Query

Map Reduce
Distributed Processing Framework

HDFS
Hadoop Distributed File System

HBase
Columnar Store

Componentes



Fonte: <https://www.devmedia.com.br/hadoop-fundamentos-e-instalacao/29466>

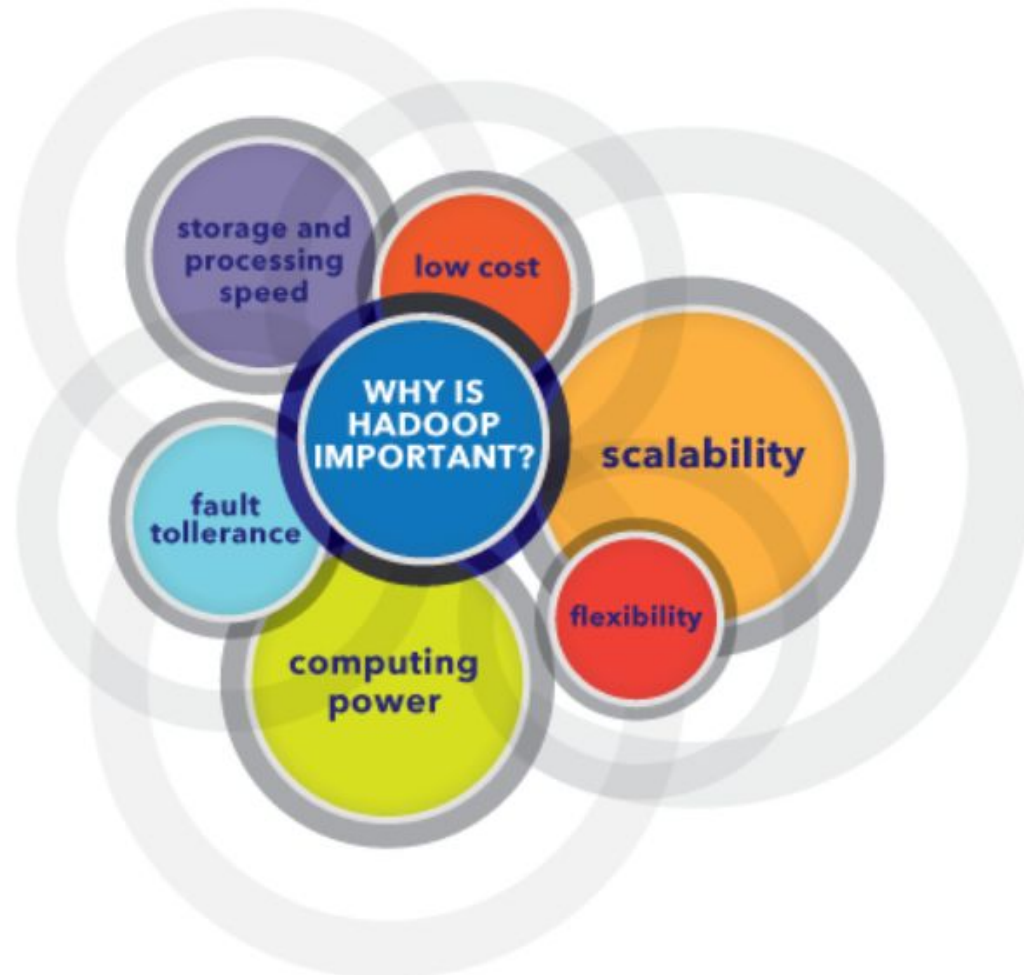
Instalação



- Modo Local ou Independente
- Modo Pseudo Distribuído
- Modo Totalmente Distribuído



Por que Hadoop?



Fonte: https://www.sas.com/pt_br/insights/big-data/hadoop.html



Instalação



Instalação

Pré-Requisitos:

- GNU/Linux ou Windows
- Java 8 - JDK8 (Java 11 suporte em progresso)
- ssh (Hadoop usa SSH para acessar outros nós)
- Hadoop (.tar.gz baixado)



Instalação

- Extração do `hadoop.tar.gz`
- Variáveis de Ambiente (Java e Hadoop)
- Criar as pastas (`store/namenode` e `store/datanode`)
- Alterar `core-site.xml`, `hdfs-site.xml`, `yarn-site.xml`, `mapred-site.xml` e o `hadoop-env.sh`
- Formatar o hdfs do namenode



Aplicação Prática



Aplicação Prática - Entrada

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque sed interdum lorem. Donec gravida nisi et nisl ullamcorper bibendum. Nulla dictum urna ex, at rhoncus nulla malesuada fermentum. Donec gravida vel ante quis ultricies. Phasellus blandit elementum rutrum. Vestibulum ut condimentum massa. Nunc a eros et nibh imperdiet finibus eget a est.

Cras dignissim ipsum non mollis fringilla. Suspendisse potenti. Integer fermentum ornare felis, a auctor orci finibus malesuada. Quisque nec feugiat tellus, quis malesuada erat. Fusce orci sapien, elementum sit amet tellus non, eleifend ornare dui. Quisque mattis lacus vel fermentum porttitor. ...



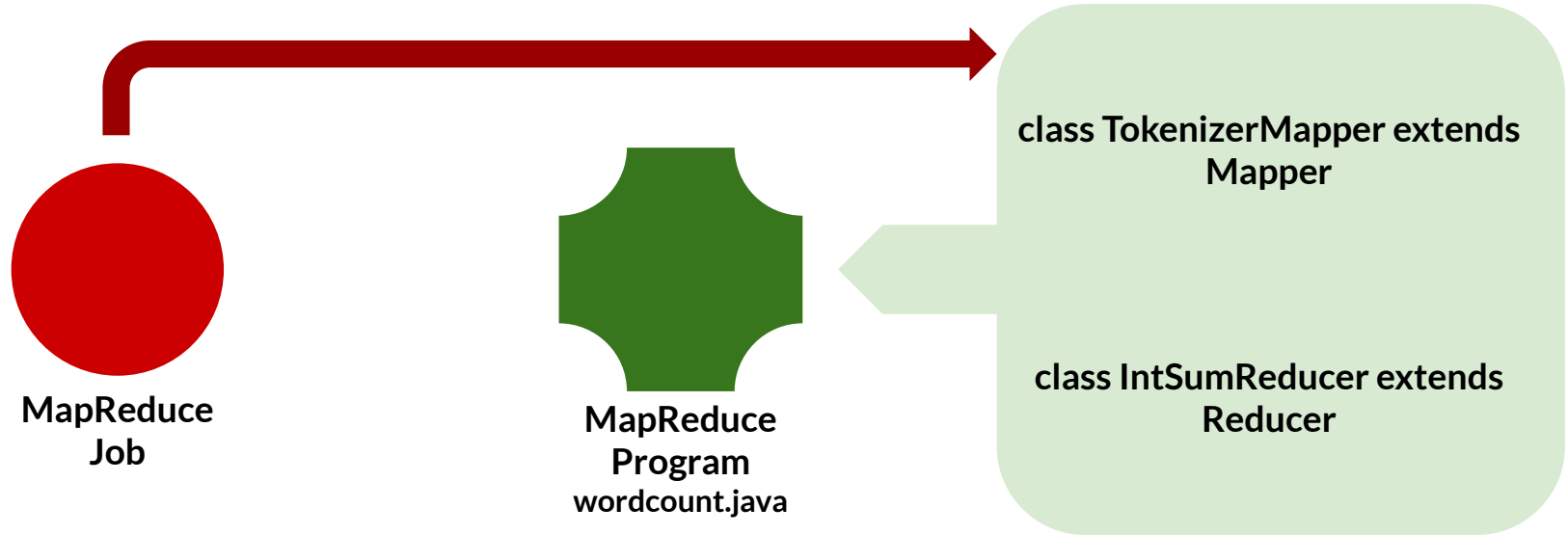
Aplicação Prática - Saída

Aenean 2
Aliquam 2
Class 1
Cras 2
Donec 4
Fusce 3
In 3
Integer 4
Lorem 1

Maecenas 1
Nam 2
Nulla 3
Nullam 2
Nunc 8
Pellentesque
3
Phasellus 2
Proin 1

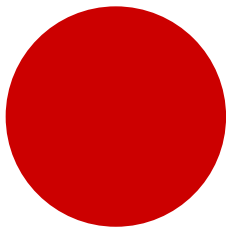
Quisque 3
Sed 4
Suspendisse 4
Ut 3
Vestibulum 6
Vivamus 1
a 6
ac 5
ac, 1

Aplicação Prática - Abstração





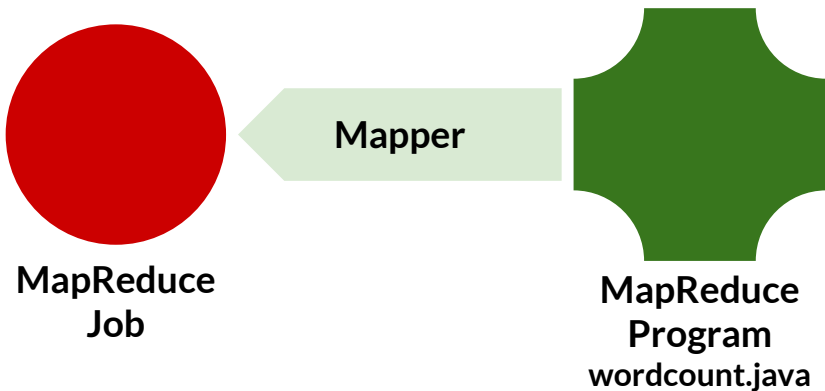
Aplicação Prática - Abstração



MapReduce
Job

1. Fase: Map
2. Fase: Reduce

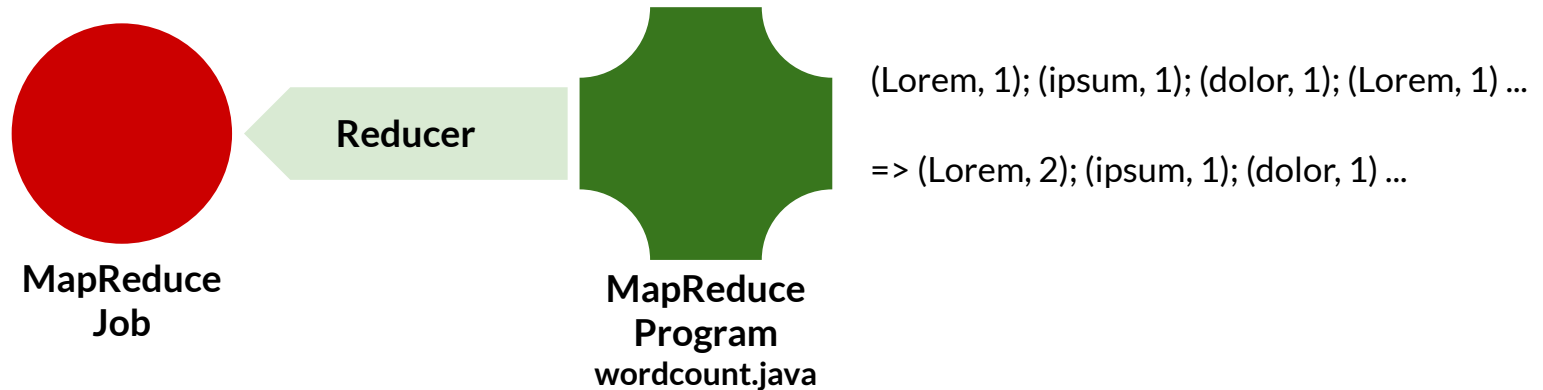
Aplicação Prática - Fase 1



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque sed interdum lorem. Donec gravida nisi et nisl...

=> (Lorem, 1); (ipsum, 1); (dolor, 1) ...

Aplicação Prática - Fase 2





Aplicação Prática

AO VIVO

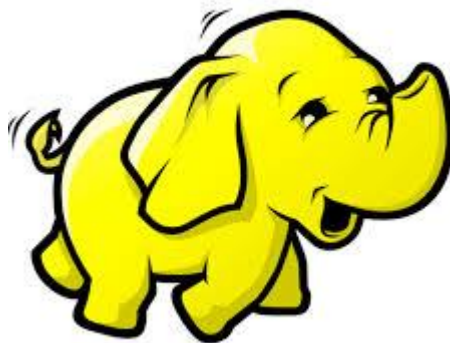


Conclusão

ebay

Linked in

facebook



YAHOO!

amazon



The
New York
Times



< > ↺ || www.amazon.com.br

AliExpress Booking.com Danu Americanas Facebook Galeria do Web Slice Sites Sugeridos Importado do IE Observar e absorver Estados do Brasil jo...

Inspirado por suas compras

Página 2 de 10 | [Iniciar novamente](#)



Inteligência emocional

Daniel Goleman

★★★★★ 179

Capa comum

R\$44,13



O homem mais rico da Babilônia

George S. Clason

★★★★★ 619

Capa comum

R\$20,33



O poder do subconsciente

Joseph Murphy

★★★★★ 269

Capa comum

R\$30,65



Trabalhe 4 horas por semana

Timothy Ferriss

★★★★★ 356

Capa comum

R\$42,48



A Culpa É das Estrelas

John Green

★★★★★ 391

Capa comum

R\$28,12

Seu histórico de navegação [Exiba ou edite seu histórico de buscas](#)

Página 1 de 3

https://www.amazon.com.br/Trabalhe-4-Horas-por-Semana/dp/8542211561/ref=pd_rhf_gw_s_qp_0_?_encoding=UTF8&pd_rd_i=8542211561&pd_rd_r=e4a1845c-408c-4c84-97f9-da2ebfbbe498&pd_rd_w=8ldWz...

17:08
04/12/2019



Obrigado