



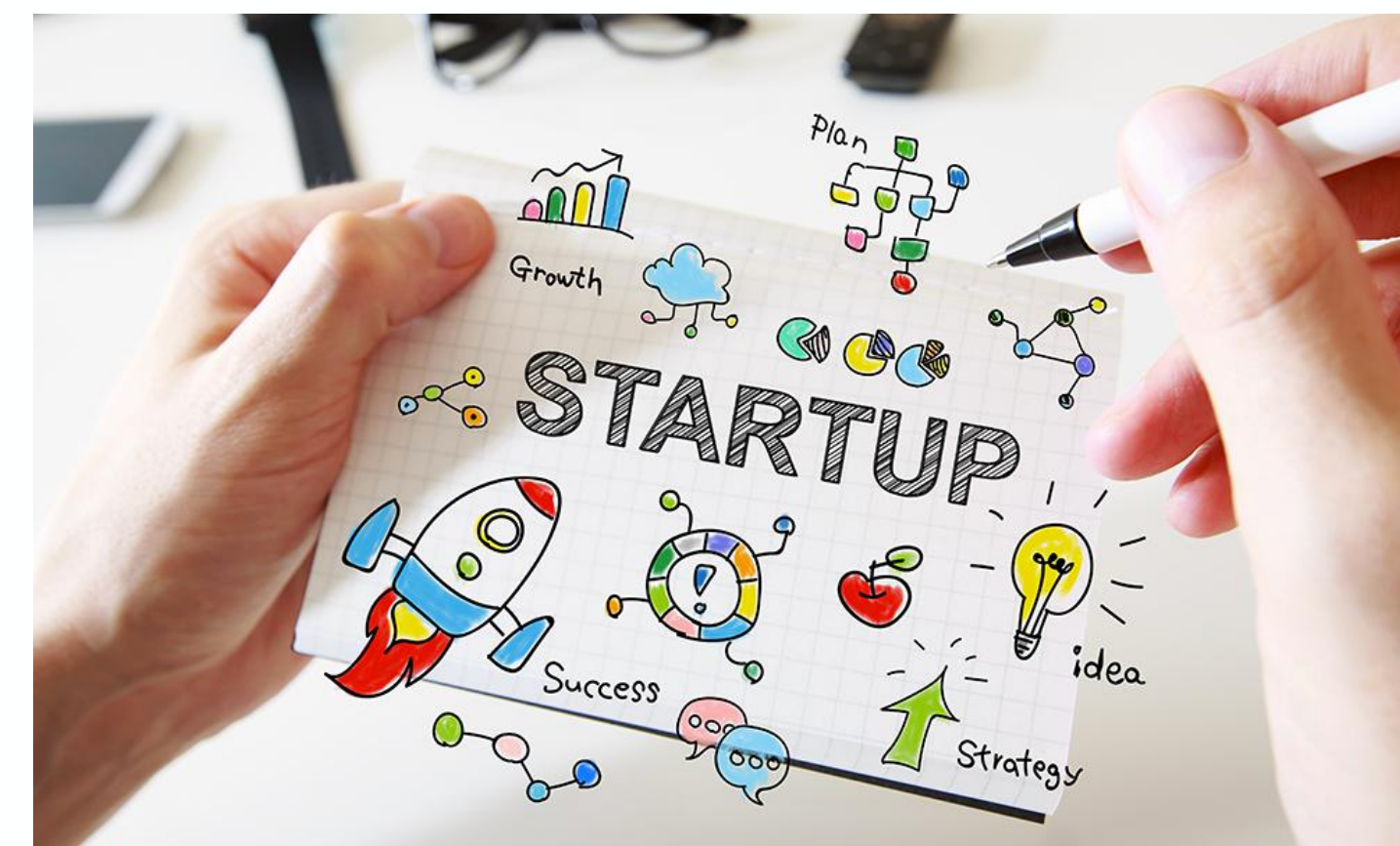
Big Data

Pós Disruptiva

Quem é essa pessoa que vos fala?



Quem é essa pessoa que vos fala?



AULA 1

Definições



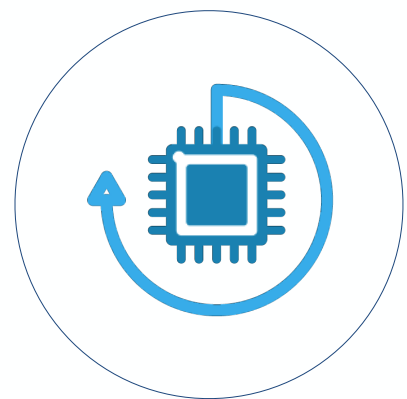


Modulo 1

Definições

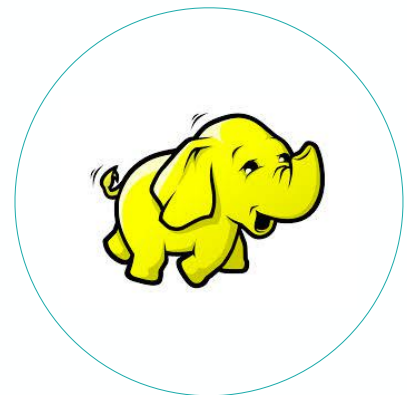
O que **vamos** aprender ?

Nesta matéria



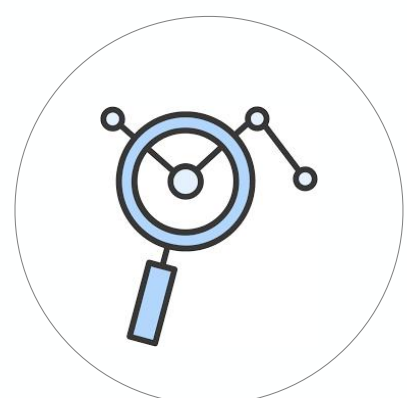
Configurar Hadoop Family

Nesta matéria você aprenderá os principais caminhos para ter uma solução Hadoop



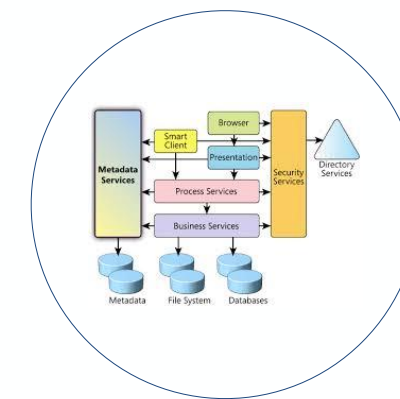
Hadoop

Vamos aprender o princípio do MapReduce, história do Hadoop e o porquê este sistema é tão importante.



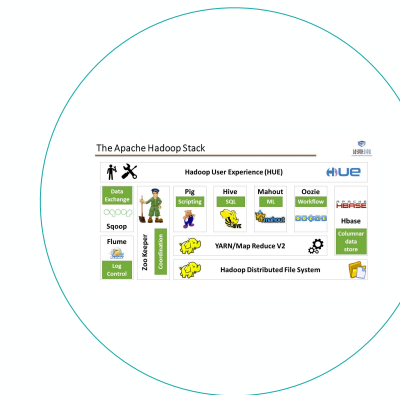
Análise de dados

Esta não é uma matéria de Análise de Dados, mas podemos nos divertir com eles, não ?



Arquitetura completa Hadoop

Vamos entender como é montado uma estrutura Hadoop de ponta e por onde começar.



Tecnologias agregadas a BigData

Vamos conhecer as principais tecnologias, para que elas servem, e como usa-las de forma básica.



! Piece of Cake

Não será moleza, muita informação, muito conteúdo e pouco tempo.



Modulo 1

O que não é e o que é BigData



VS



VS



DATA SCIENCE

BIG DATA

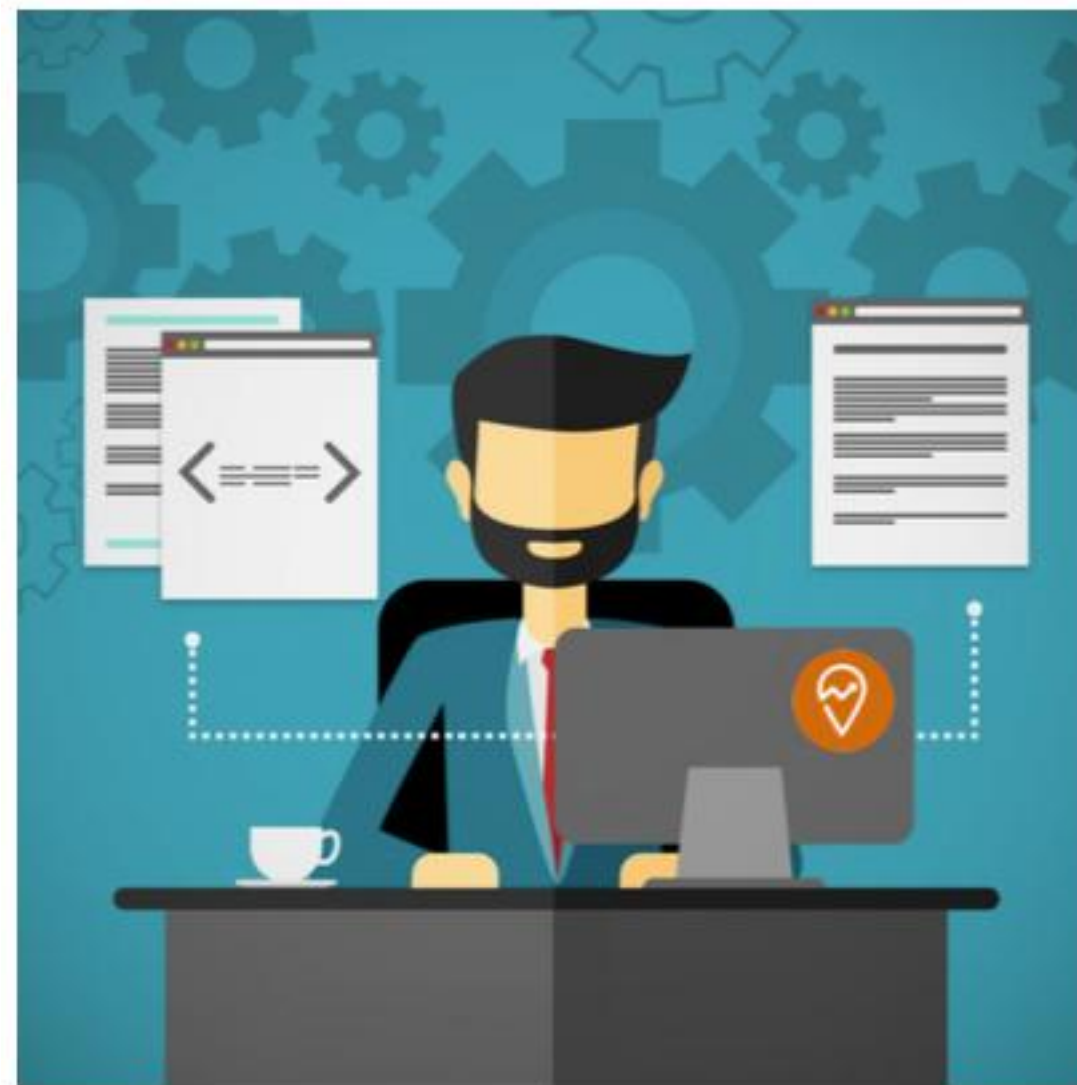
DATA ANALYTICS

Então quem é quem?

Data Science



Data Analysis

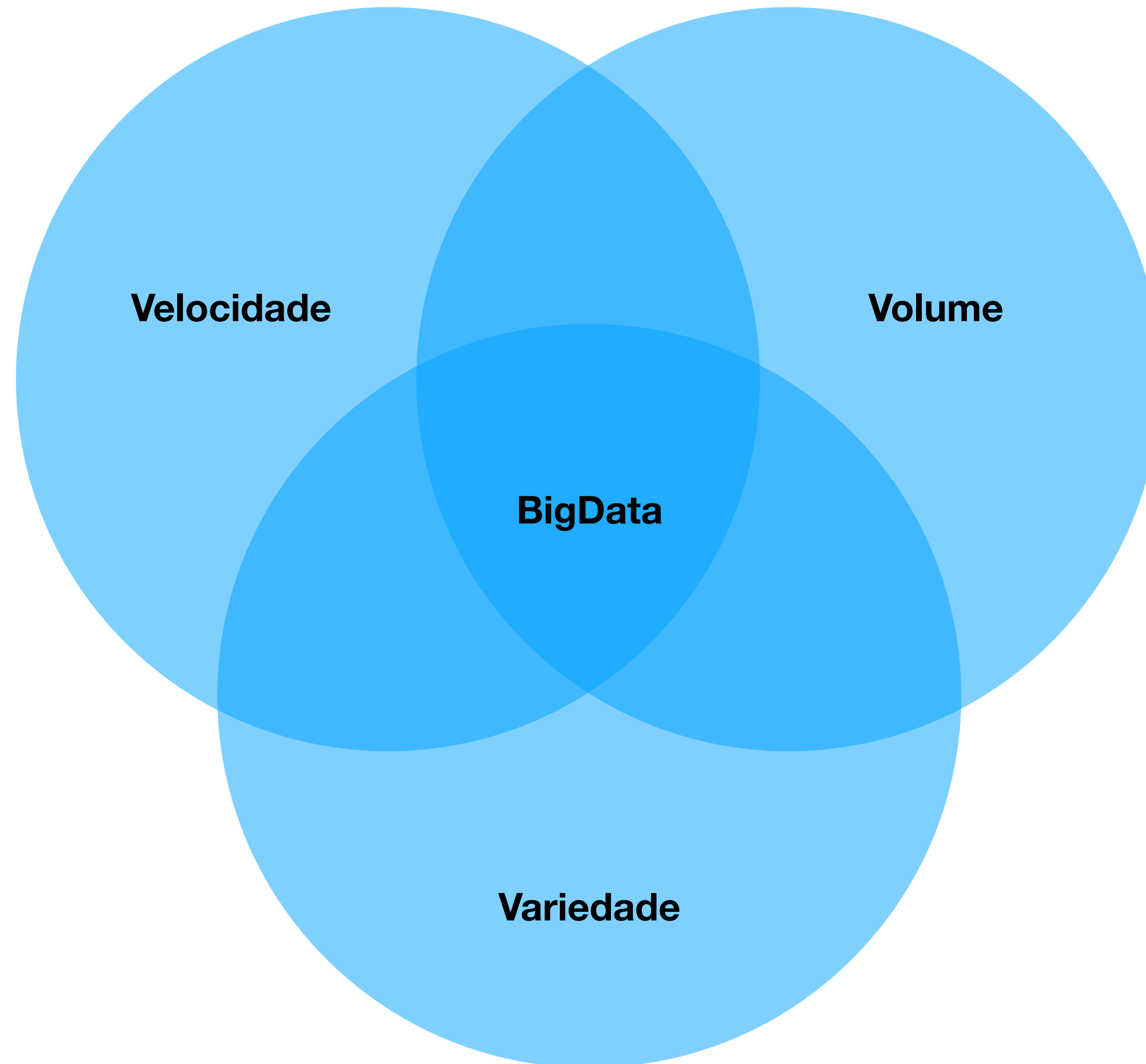


BigData





Bigdata é um problema, não uma solução!



Bigdata é um problema, não uma solução!



Eu tenho uma solução BigData

Precisamos usar BigData



Eu tenho uma solução para BigData

Temos um problema de BigData



Modulo 2

A origem das tecnologias atuais

Jeffrey Dean

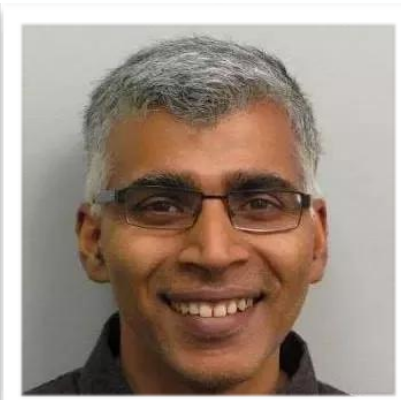


Sanjay Ghemawat

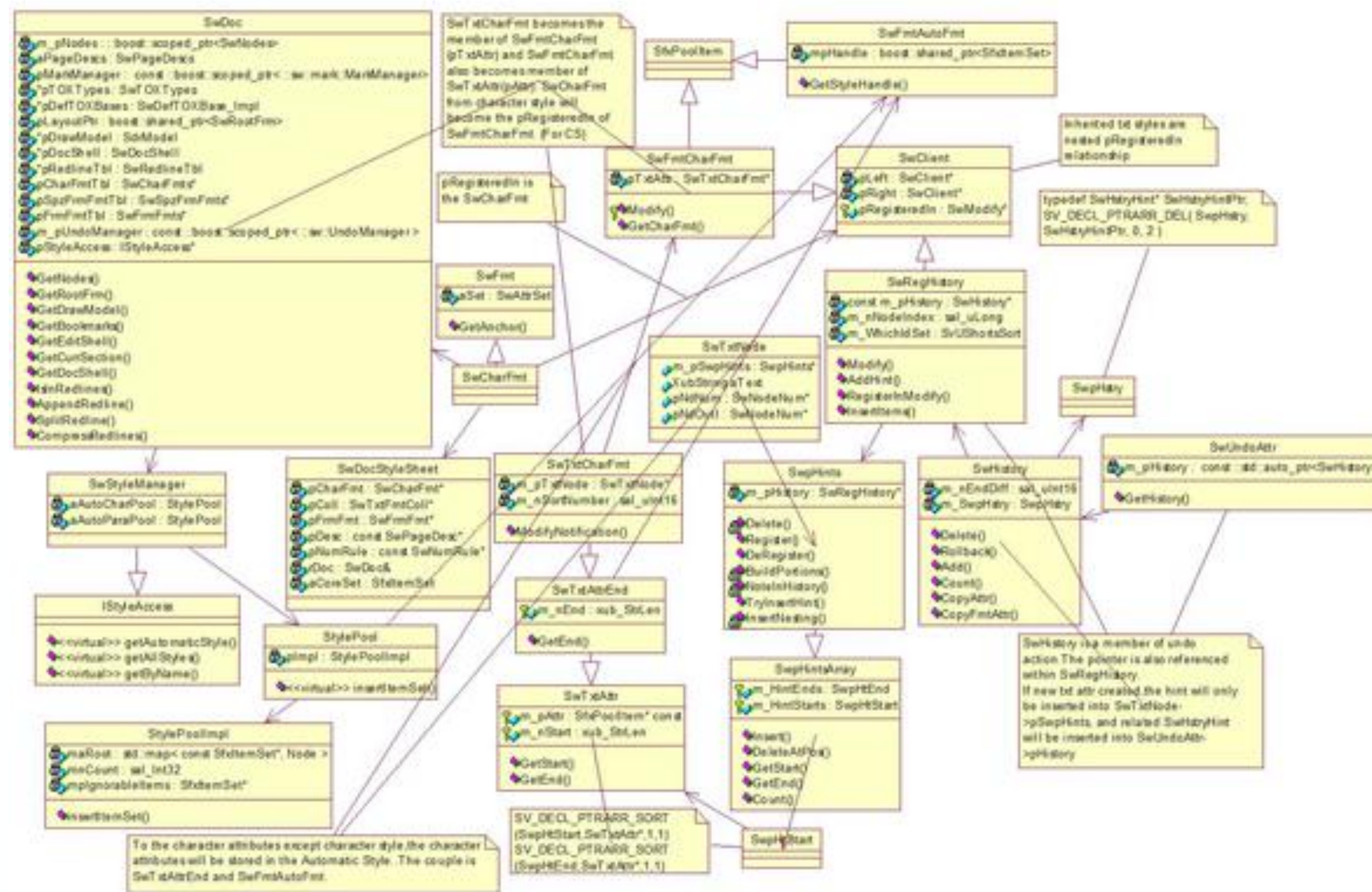
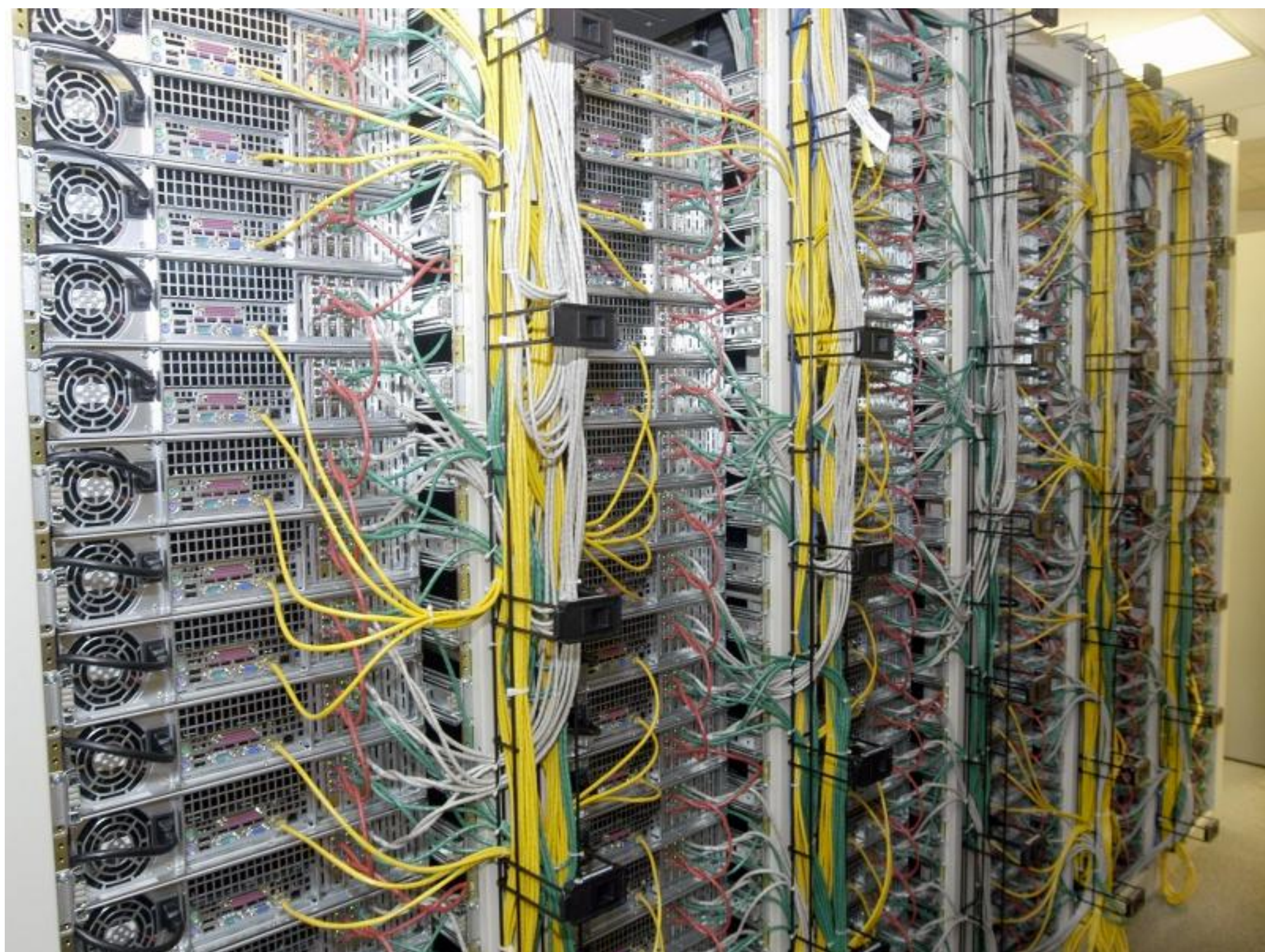


Doug Cutting





A perspective view of a road made of numbers receding into the distance. The road is composed of two parallel lines of numbers that converge towards a vanishing point in the center of the image. The numbers are arranged in a grid-like pattern, with each row of numbers forming a step in the road's surface. The numbers are black and set against a white background. The perspective is created by the way the lines of numbers recede, giving a sense of depth and distance. The numbers themselves are a mix of integers and decimal values, some with negative signs, suggesting a mathematical or scientific theme. The overall effect is a surreal and intriguing visual metaphor for a journey through data or a path of calculation.







KEEP IT
SIMPLE

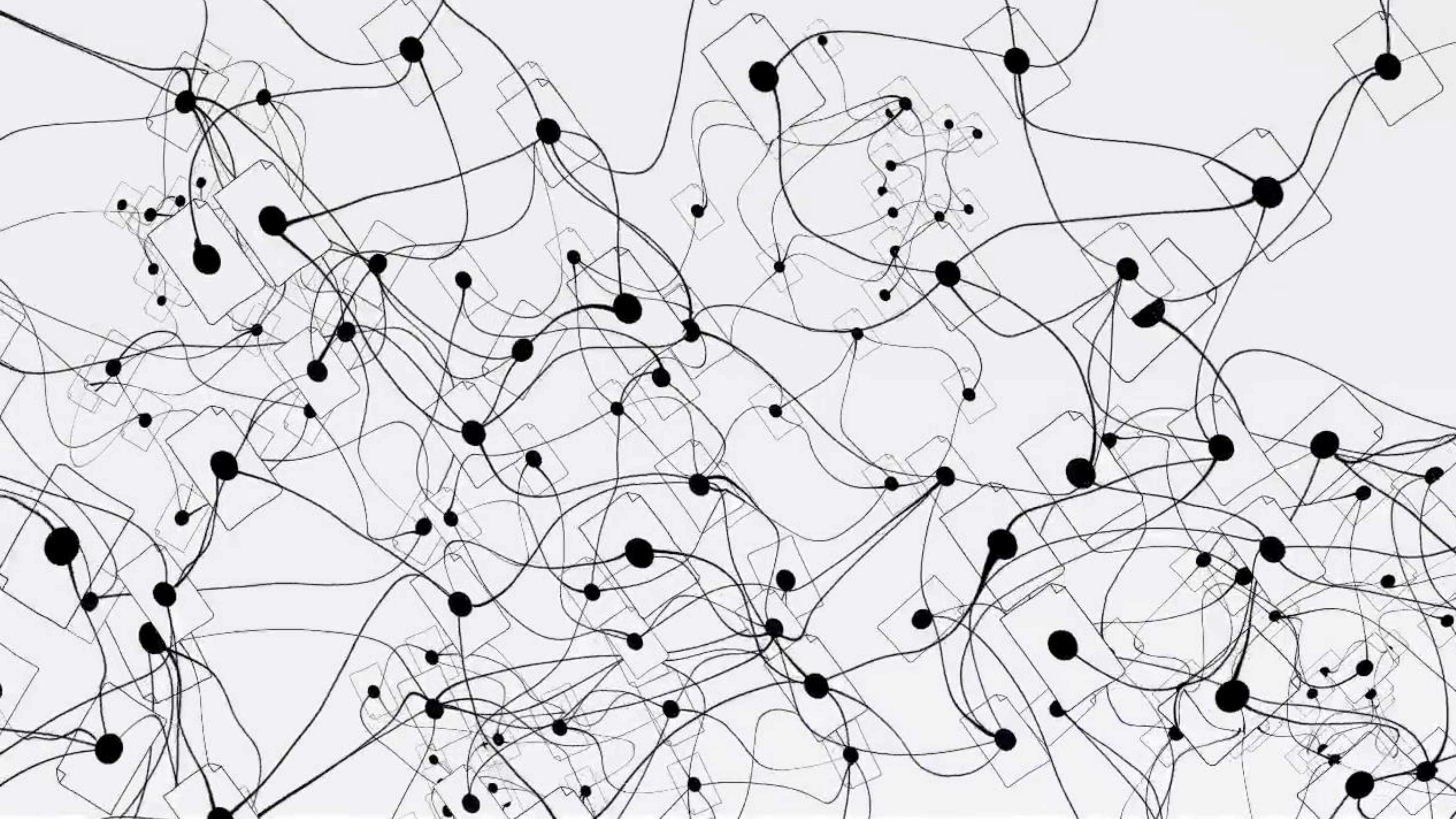
Ainda não consegui
entender muito bem
essa necessidade



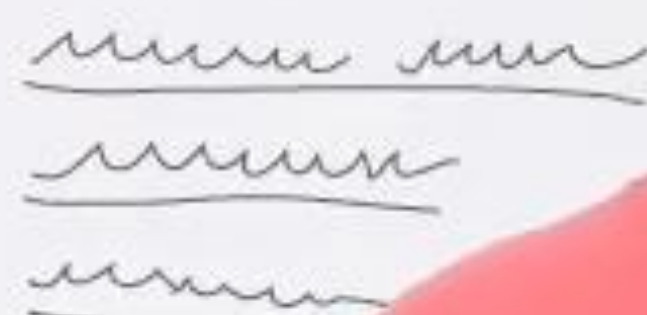
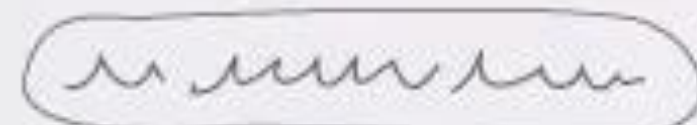
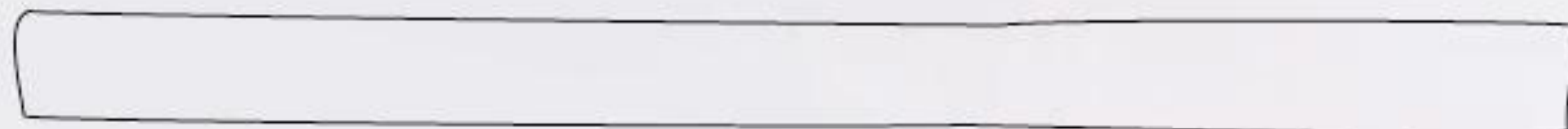
Pesquisar no Google ou digitar URL

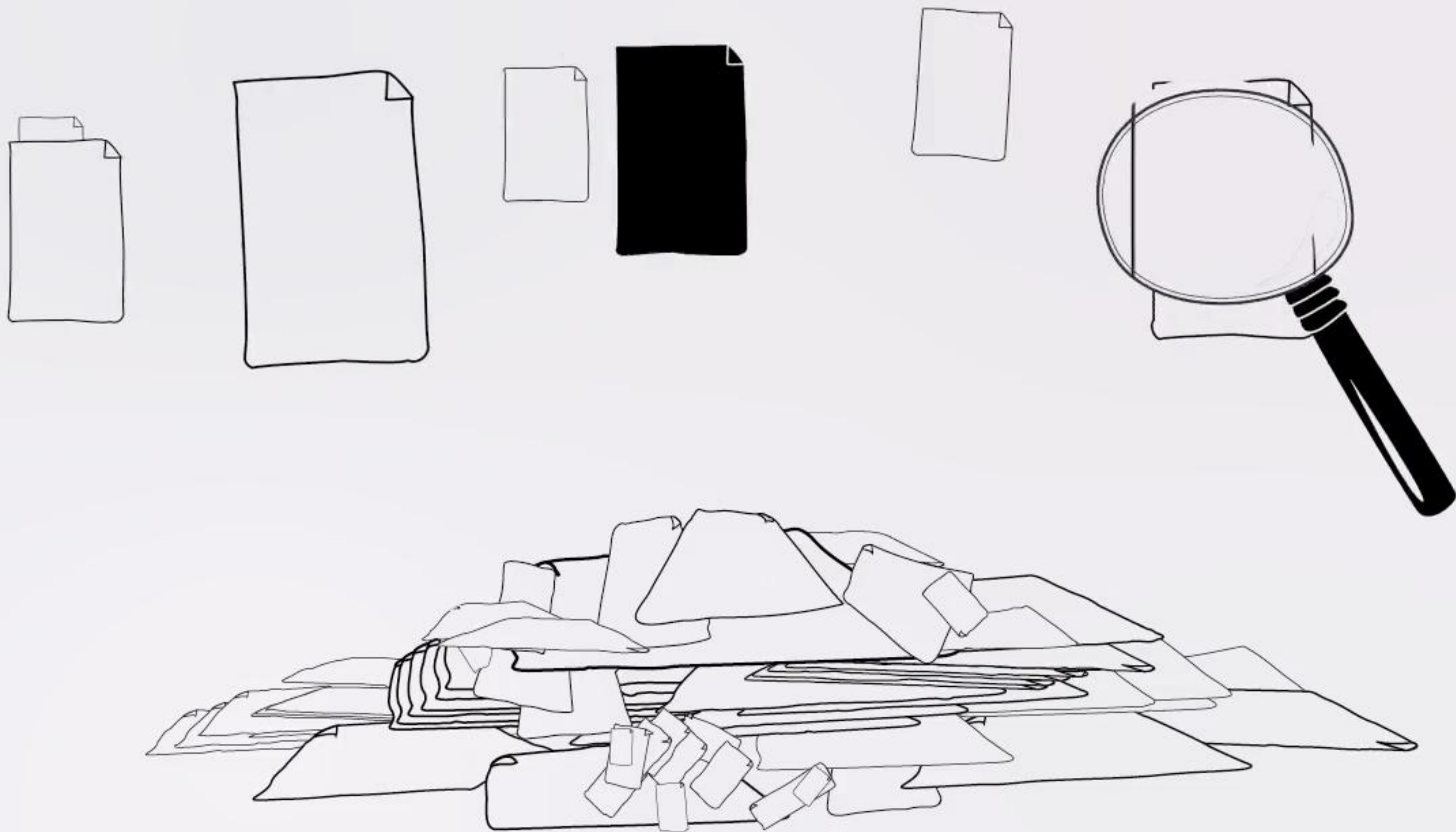


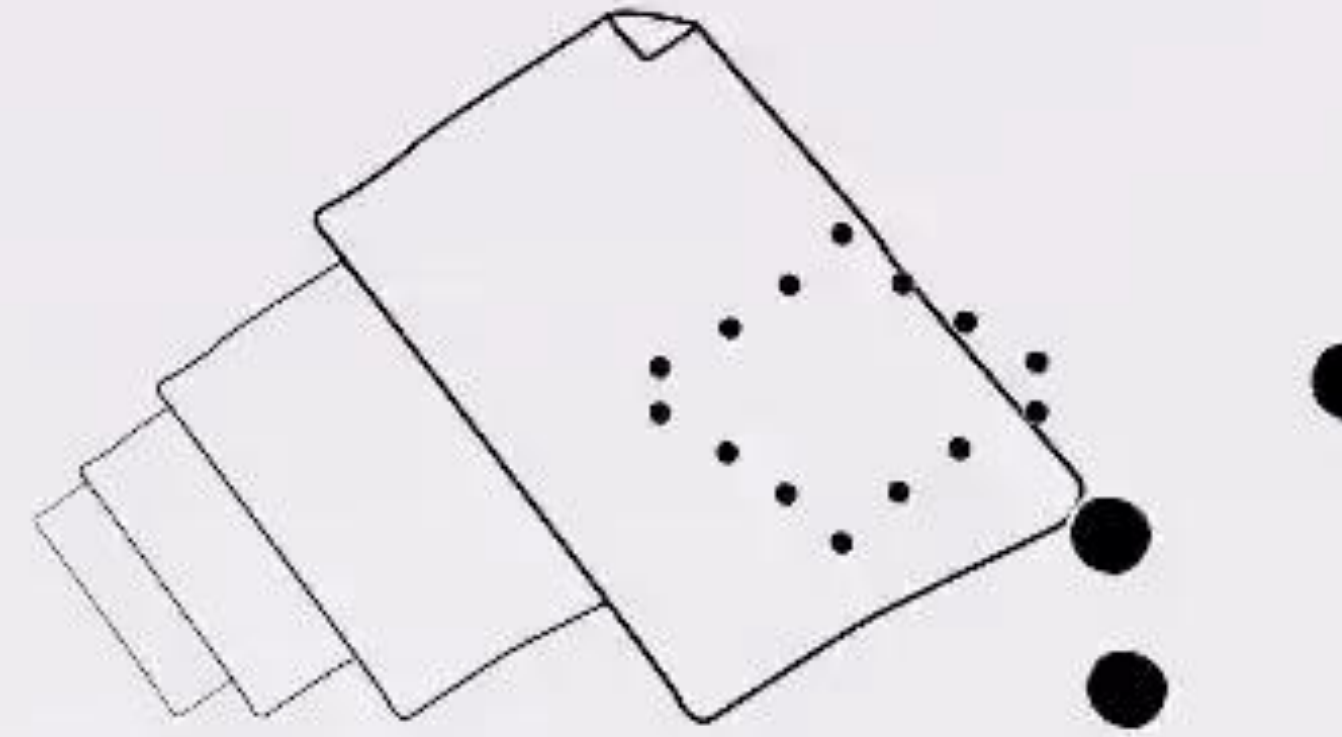




Google

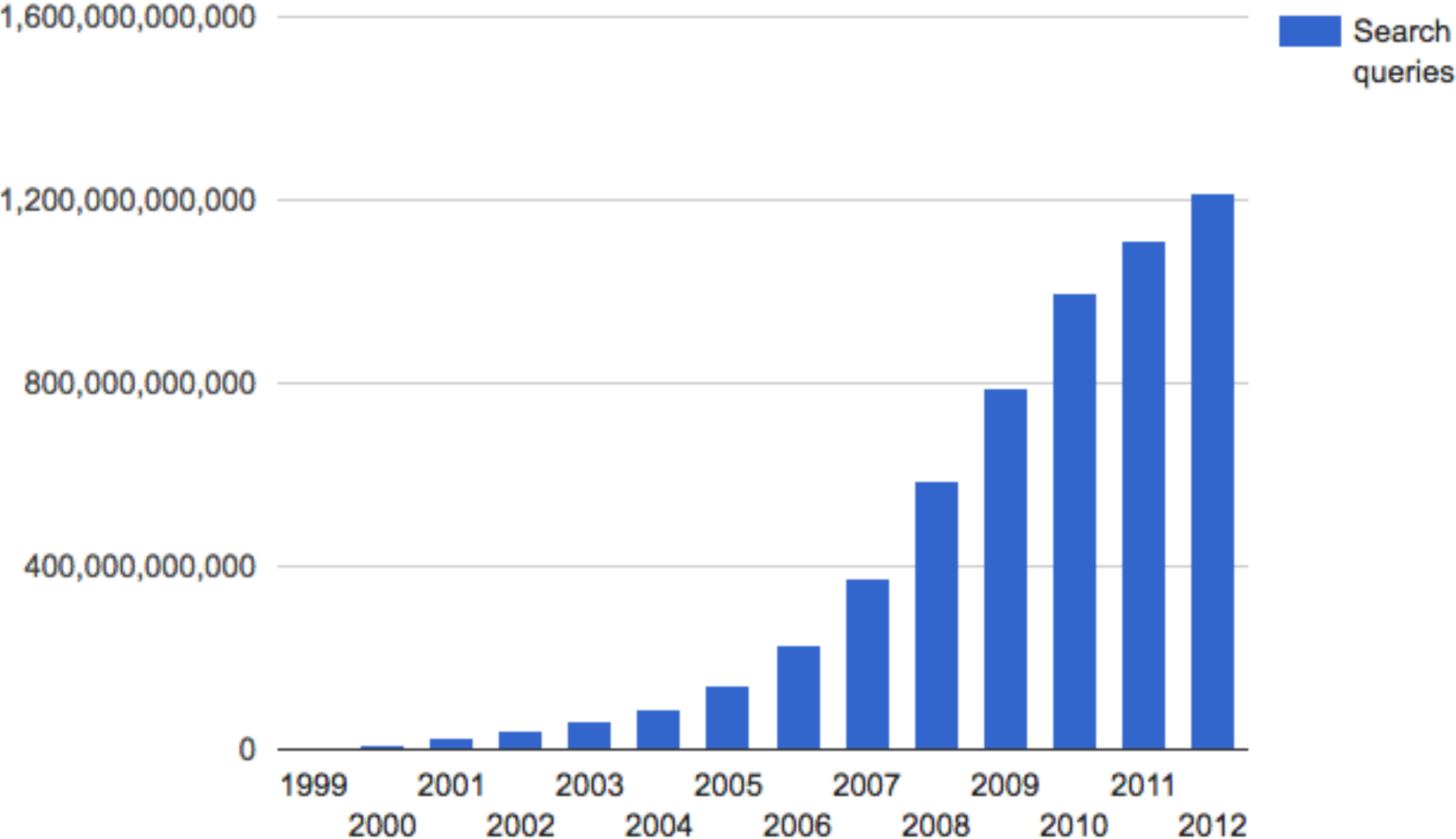




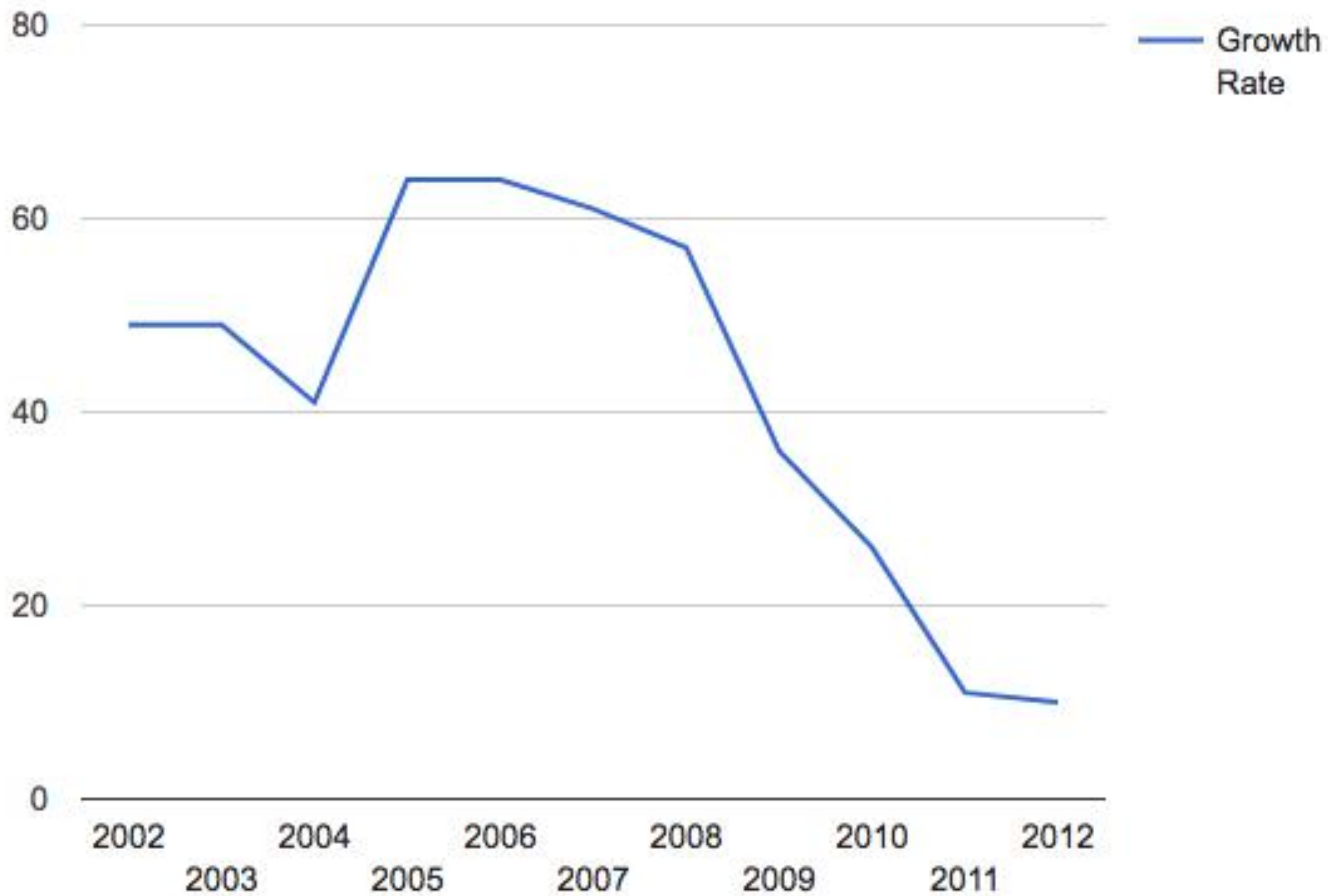


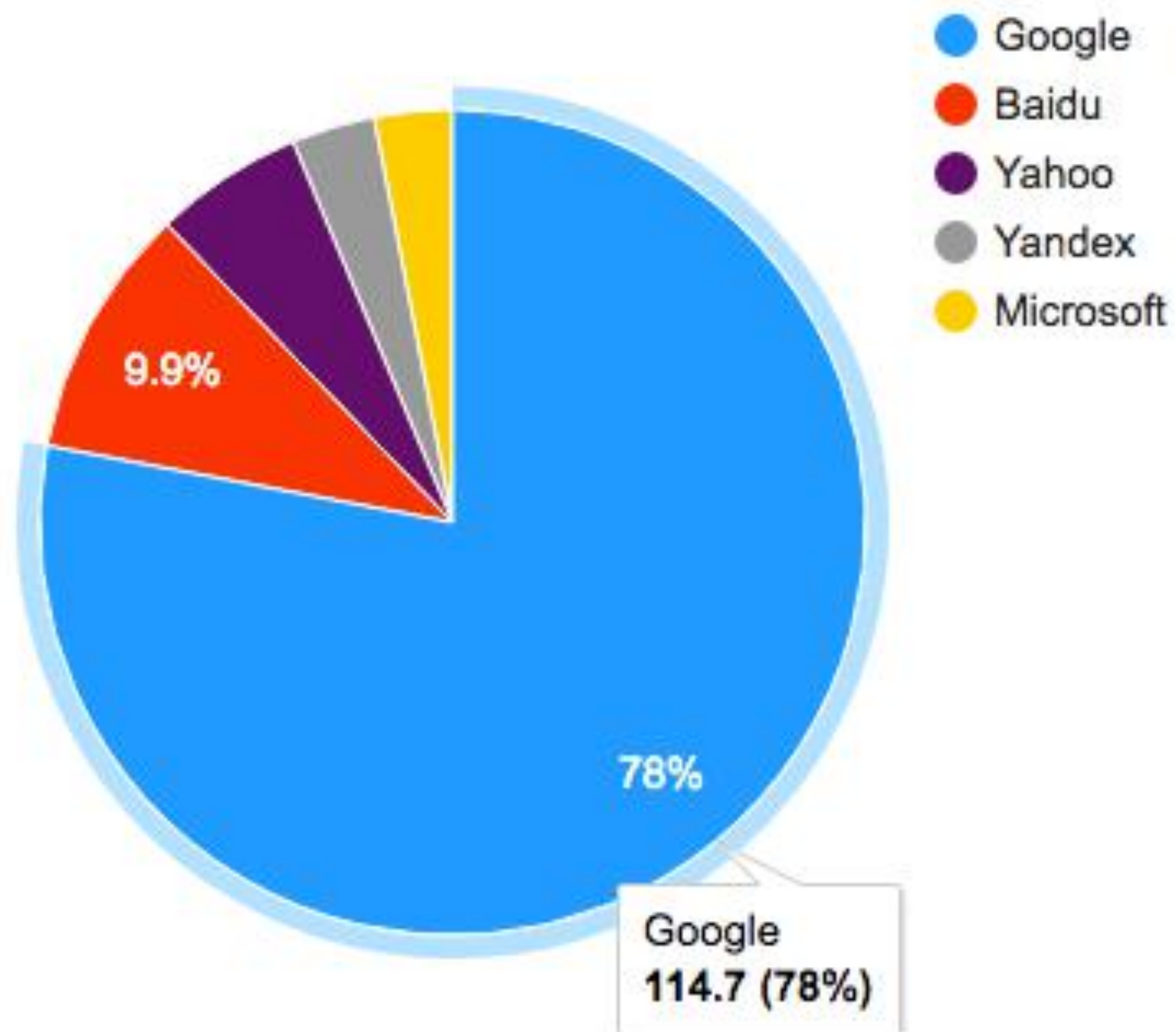
Ainda não consegui
entender muito bem
essa necessidade

Google Searches per Year



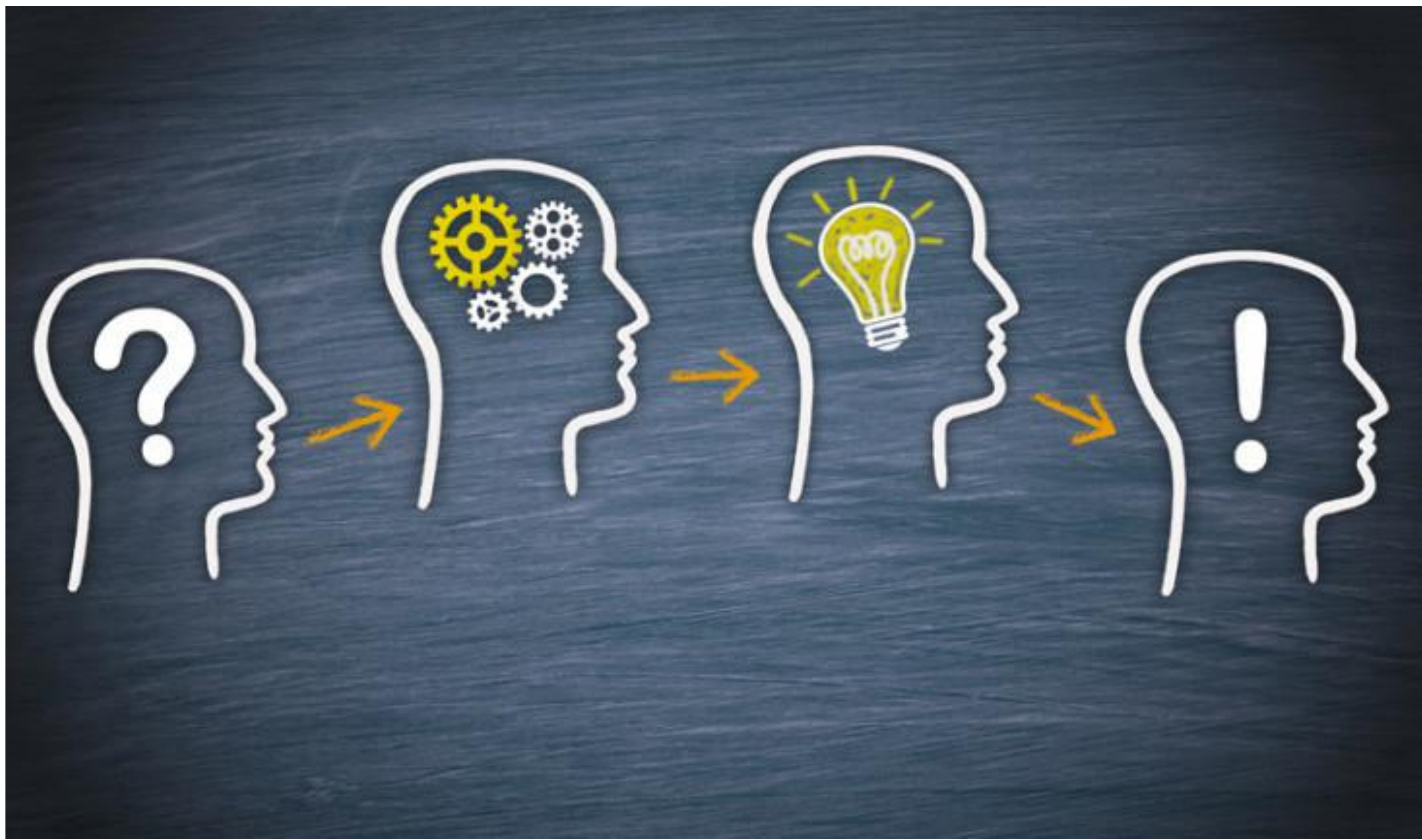
Google Search Volume Growth Rate (%)





Google 65.2% 114.7 bilhões mês
Baidu 8.2% 14.5 bilhões mês
Yahoo 4.9% 8.6 bilhões mês
Yandex 2.8% 4.8 bilhões mês
Microsoft 2.5% 4.5 bilhões mês

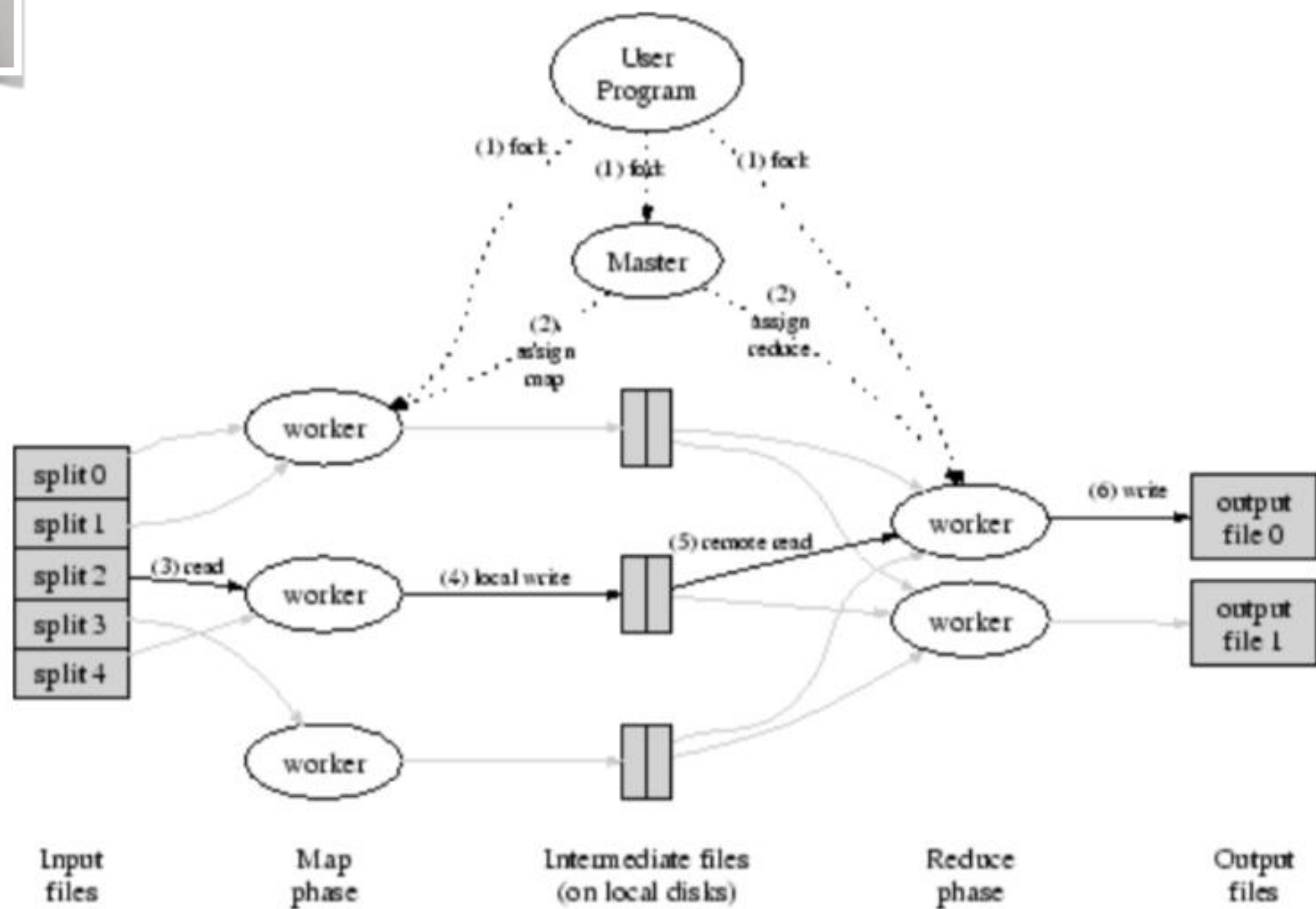
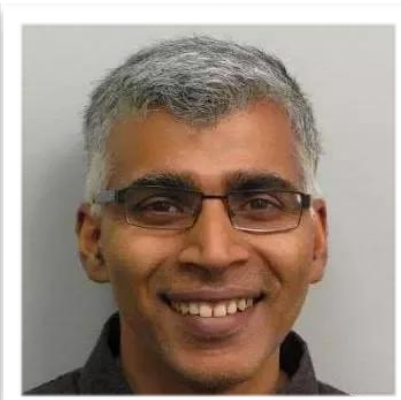
<http://www.internetlivestats.com/>





Modulo 3

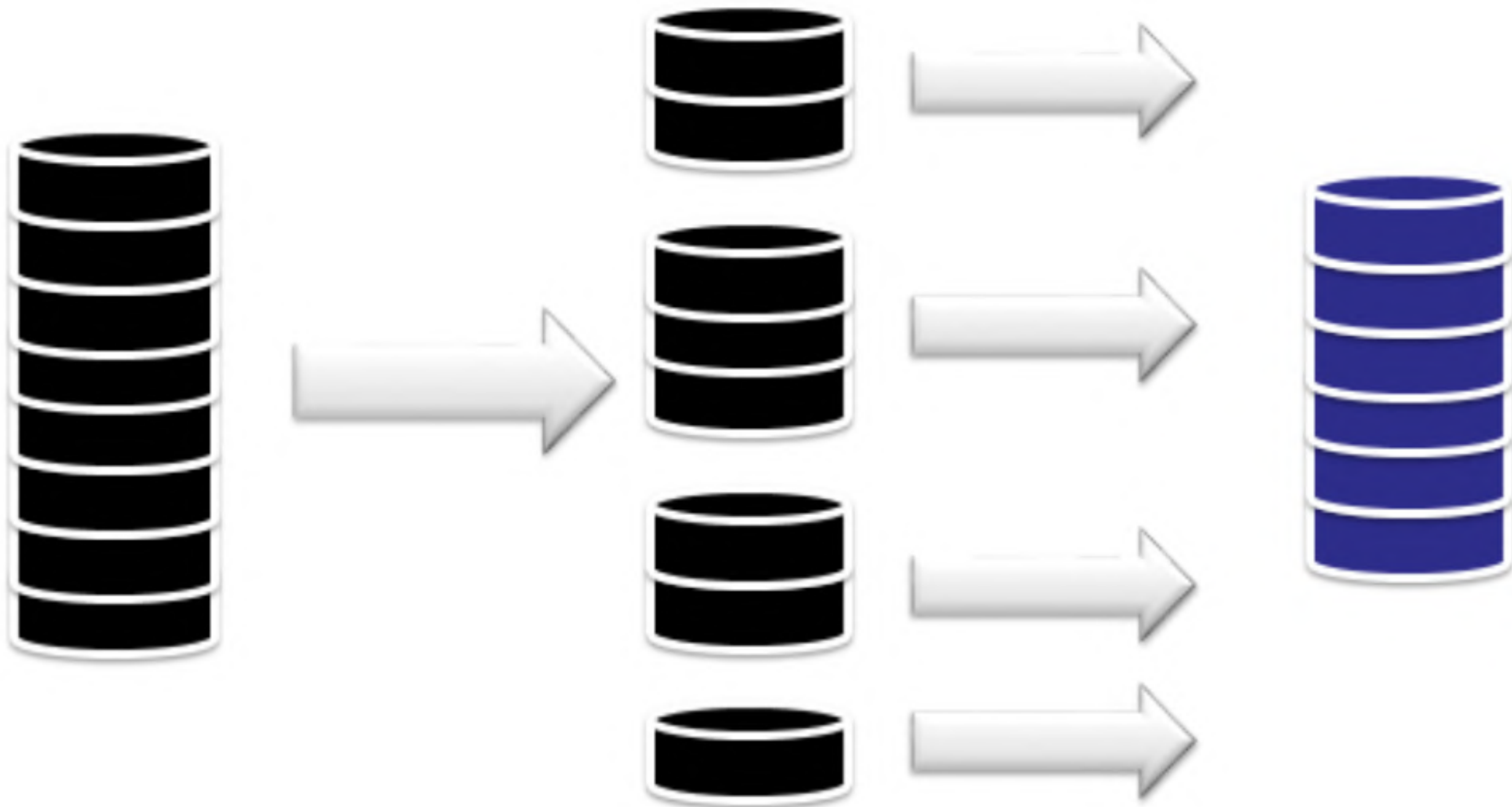
MapReduce

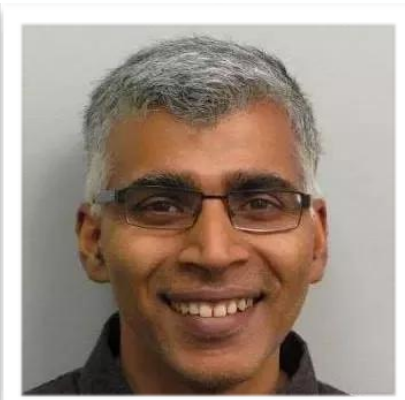




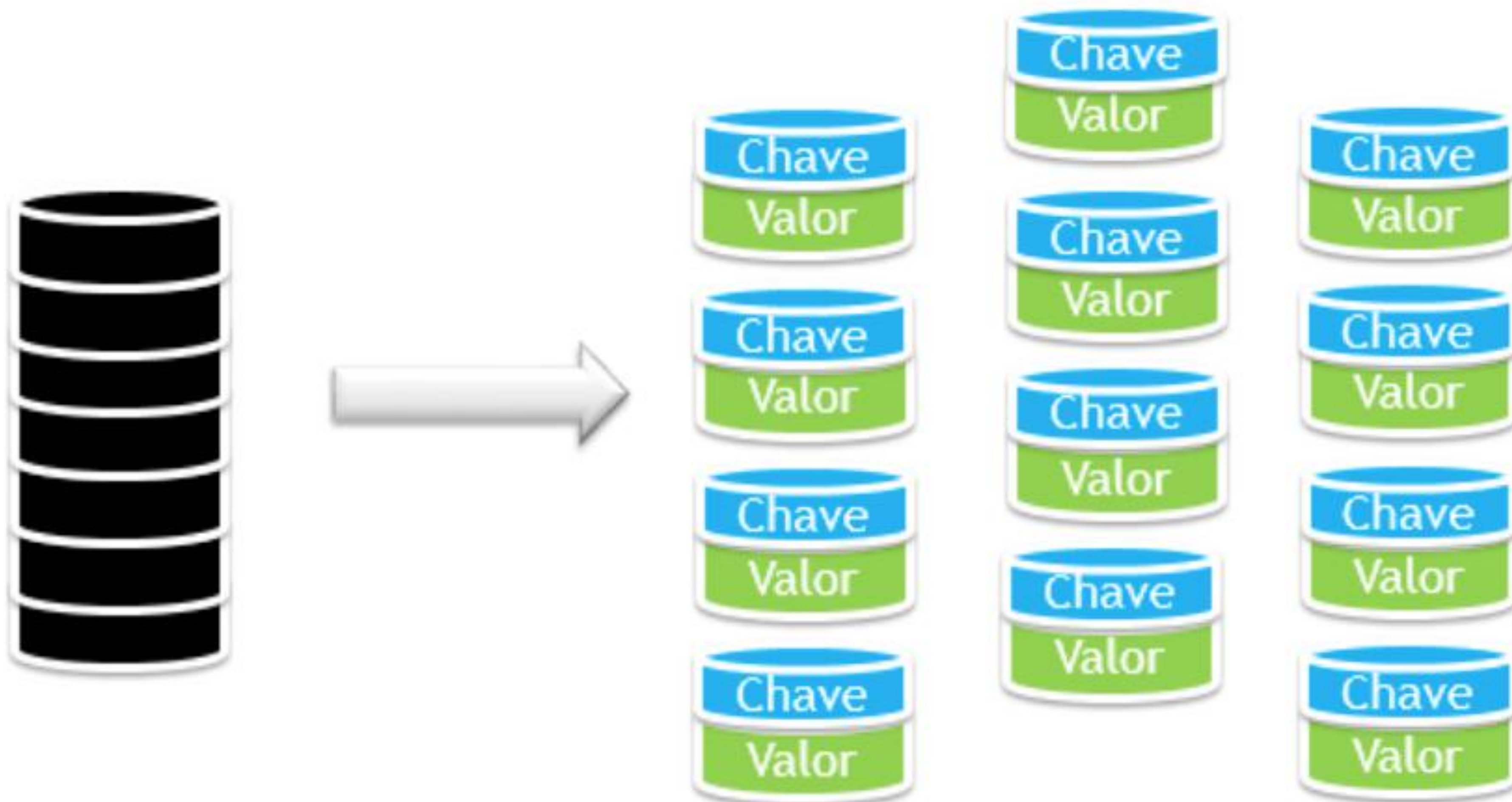
Map

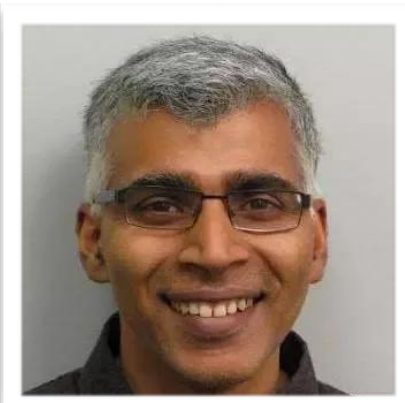
Reduce



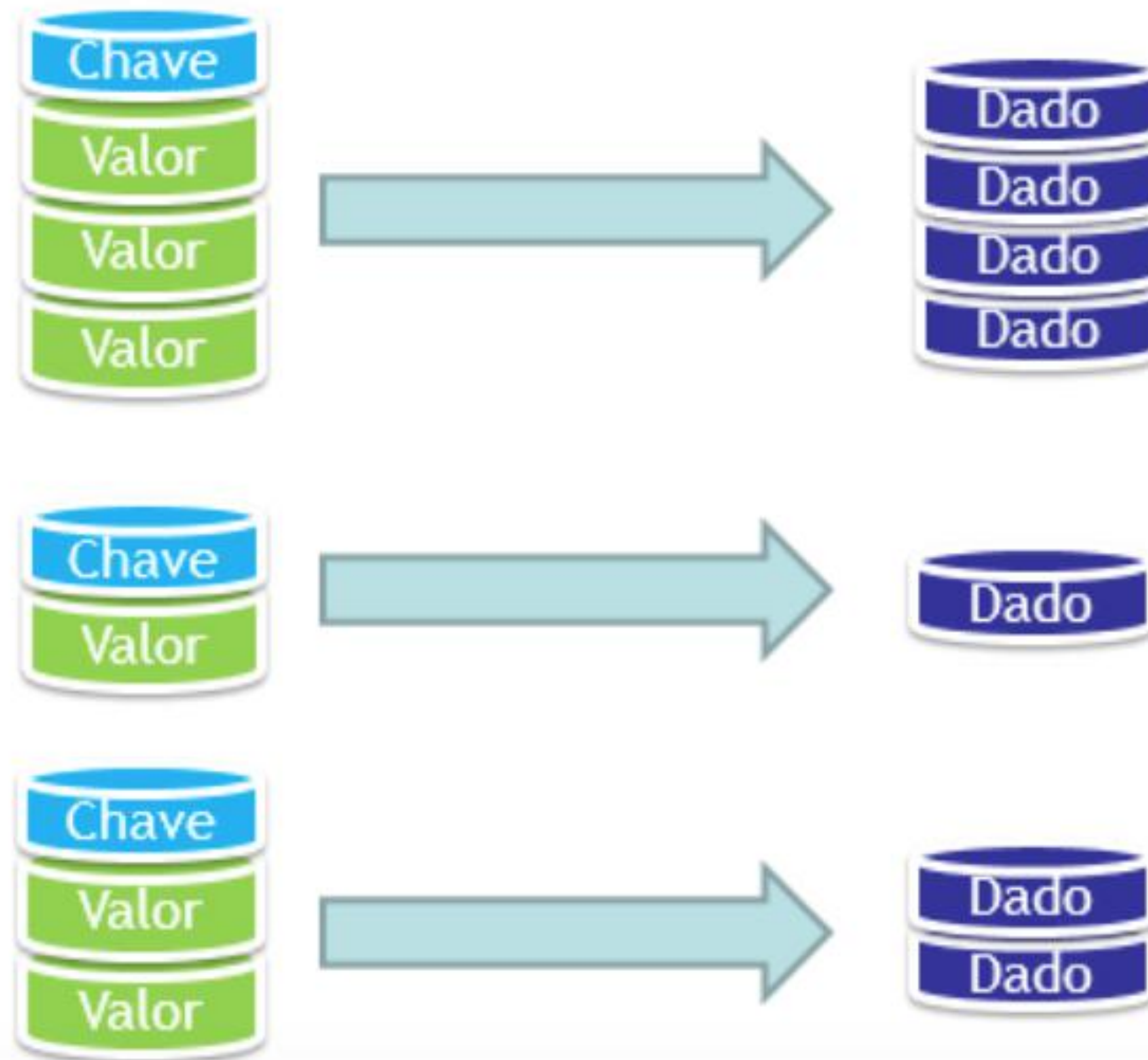


Map





Reduce



LIFE



Y U NO GIVE ME A BREAK

Documento 1

Este documento é um exemplo de como as coisas funcionam usando map reduce e indexam

Documento 2

Lorem Ipsum é simplesmente uma simulação de texto da indústria tipográfica e de impressos, e vem sendo utilizado desde o século XVI.

Documento 3

Este é um exemplo de texto que pode ser utilizado como entrada no MapReduce. Com isso é possível ter uma ideia de como é o processo de mapeamento

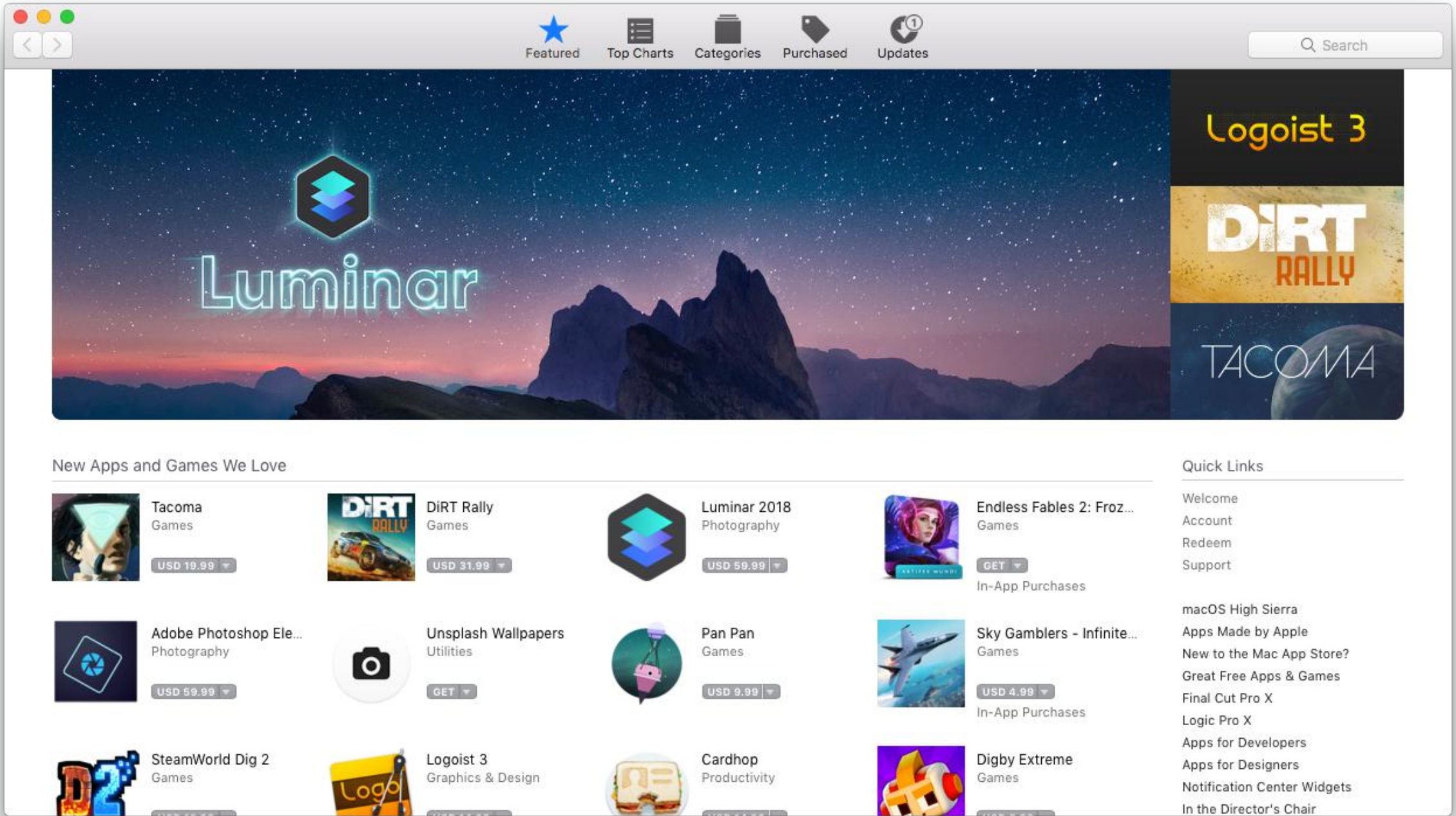
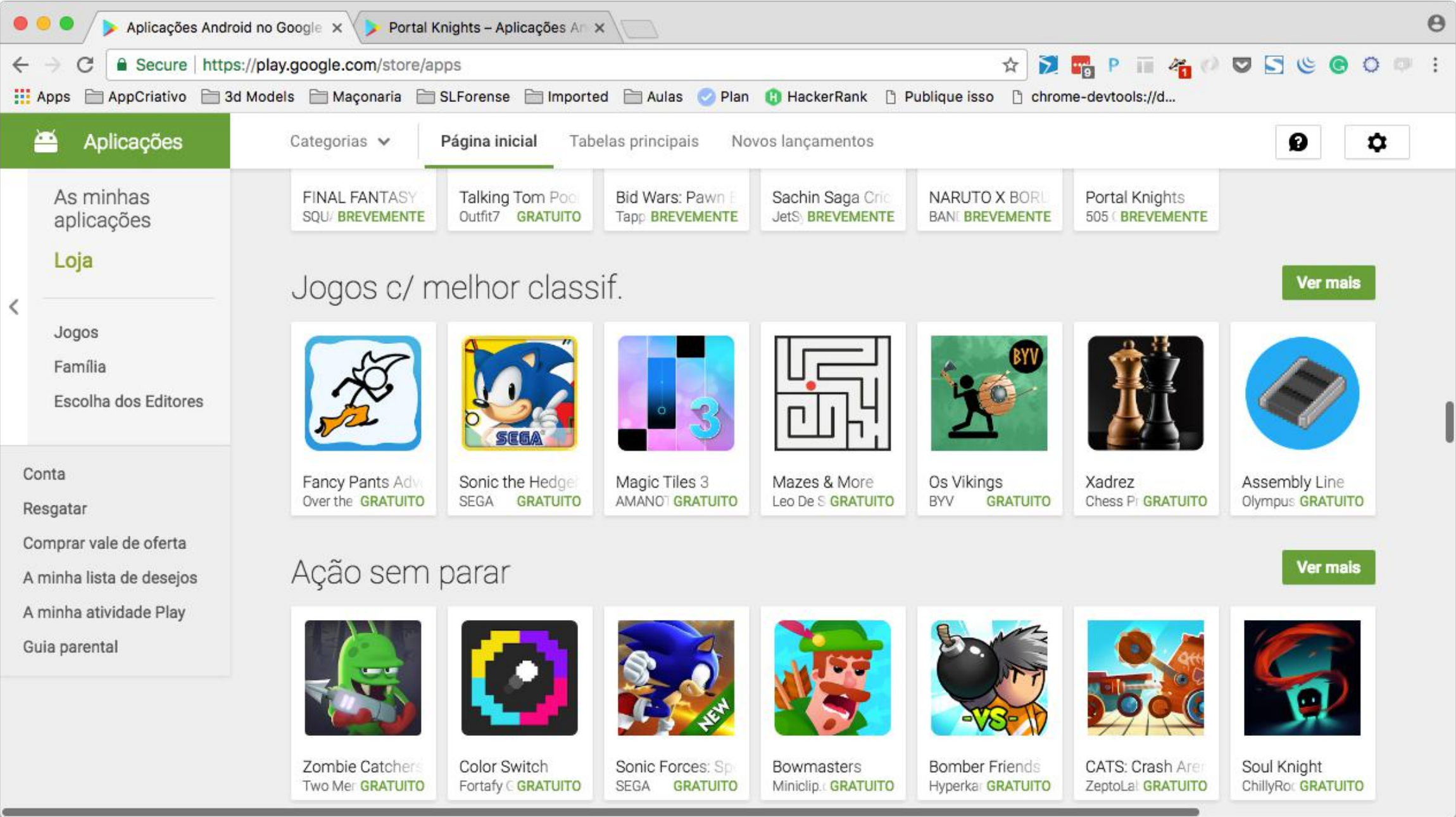


THE CHALLENGE



Modulo 3

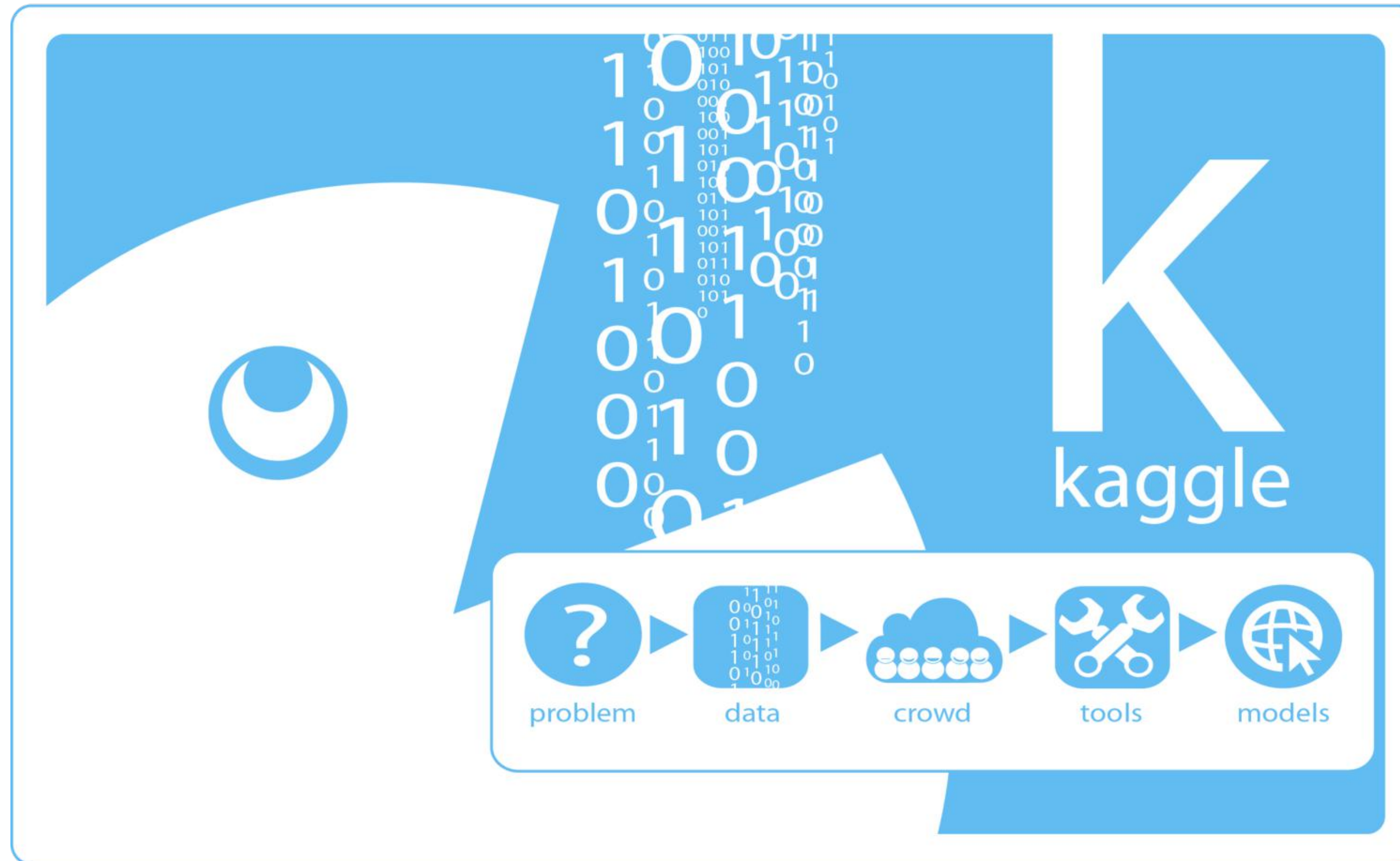
Exemplos





THE CHALLENGE

Dataset Públicos



Dataset Públicos

Crimes in Chicago | Kaggle

Secure | https://www.kaggle.com/currie32/crimes-in-chicago/data

Apps AppCriativo 3d Models Maçonaria SLForense Imported Aulas Plan HackerRank Publique isso chrome-devtools://d...

Overview **Data** Kernels Discussion Activity

Download (489 MB) New Kernel

Download All

Preview (first 100 rows) Column Metadata

	ID	Case Number	Date	Block	IUCR	Primary Type	Description
1108	10508702	HZ250447	05/03/2016 10:35:00 PM	082XX S MARYLAND AVE	041A	BATTERY	AGGRAVATED: HANDGUN
1130	10508703	HZ250489	05/03/2016 10:30:00 PM	027XX S STATE ST	0460	BATTERY	SIMPLE
1801	10508704	HZ250514	05/03/2016 09:30:00 PM	002XX E 46TH ST	0460	BATTERY	SIMPLE
1868	10508709	HZ250523	05/03/2016 04:00:00 PM	014XX W DEVON AVE	0460	BATTERY	SIMPLE
1891	10508982	HZ250667	05/03/2016 10:30:00 PM	069XX S ASHLAND AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE
1935	10508710	HZ250469	05/03/2016 09:44:00 PM	074XX S SOUTH SHORE DR	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN
2150	10508715	HZ250541	05/03/2016 11:11:00 PM	006XX N WABASH AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE
2193	10508717	HZ250415	05/03/2016 05:30:00 PM	011XX W JACKSON BLVD	0890	THEFT	FROM BUILDING

Como resolver este problema ?





THE CHALLENGE

Isso que fizemos é uma solução para BigData?

Crimes in Chicago | Kaggle

Secure

https://www.kaggle.com/currie32/crimes-in-chicago/data

chrome-devtools://d...

Apps

AppCriativo

3d Models

Maçonaria

SLForense

Imported

Aulas

Plan

HackerRank

Publique isso

Overview

Data

Kernels

Discussion

Activity

Download (489 MB)

New Kernel

Download All

Preview (first 100 rows)

Column Metadata

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	
	1108	10508702	HZ250447	05/03/2016 10:35:00 PM	082XX S MARYLAND AVE	041A	BATTERY	AGGRAVATED; HANDGUN
	1130	10508703	HZ250489	05/03/2016 10:30:00 PM	027XX S STATE ST	0460	BATTERY	SIMPLE
	1801	10508704	HZ250514	05/03/2016 09:30:00 PM	002XX E 46TH ST	0460	BATTERY	SIMPLE
	1868	10508709	HZ250523	05/03/2016 04:00:00 PM	014XX W DEVON AVE	0460	BATTERY	SIMPLE
	1891	10508982	HZ250667	05/03/2016 10:30:00 PM	069XX S ASHLAND AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE
	1935	10508710	HZ250469	05/03/2016 09:44:00 PM	074XX S SOUTH SHORE DR	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN
	2150	10508715	HZ250541	05/03/2016 11:11:00 PM	006XX N WABASH AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE
	2193	10508717	HZ250415	05/03/2016 05:30:00 PM	011XX W JACKSON BLVD	0890	THEFT	FROM BUILDING

Weather Data

posted in [Crimes in Chicago](#) 6 months ago

Hey everyone

A lot of people in Chicago think the weather is related to people killing each other so I've collected weather data that works nicely for this data set. Will update this post later with how I extracted the data and a writeup on my analysis. So far I couldn't find much of a link between homicides and weather.

Dataset description: weather data for O'hare airport from 2005-2014

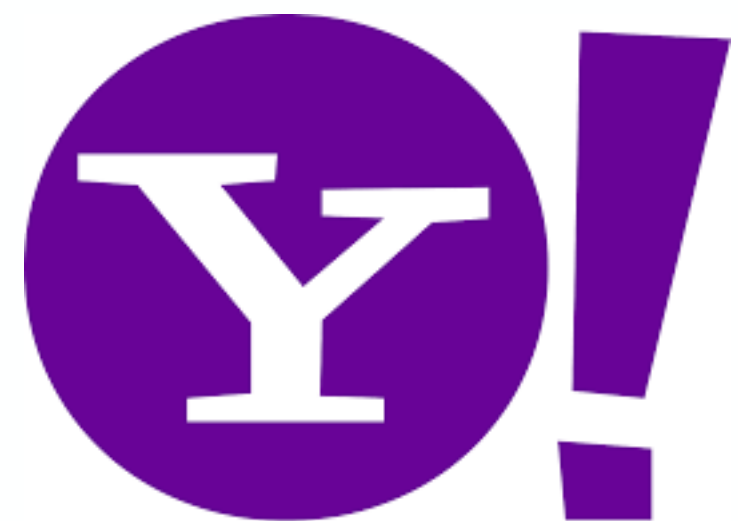
 [weather.csv](#) (137.01 KB)





Modulo 4

Hadoop



Doug Cutting

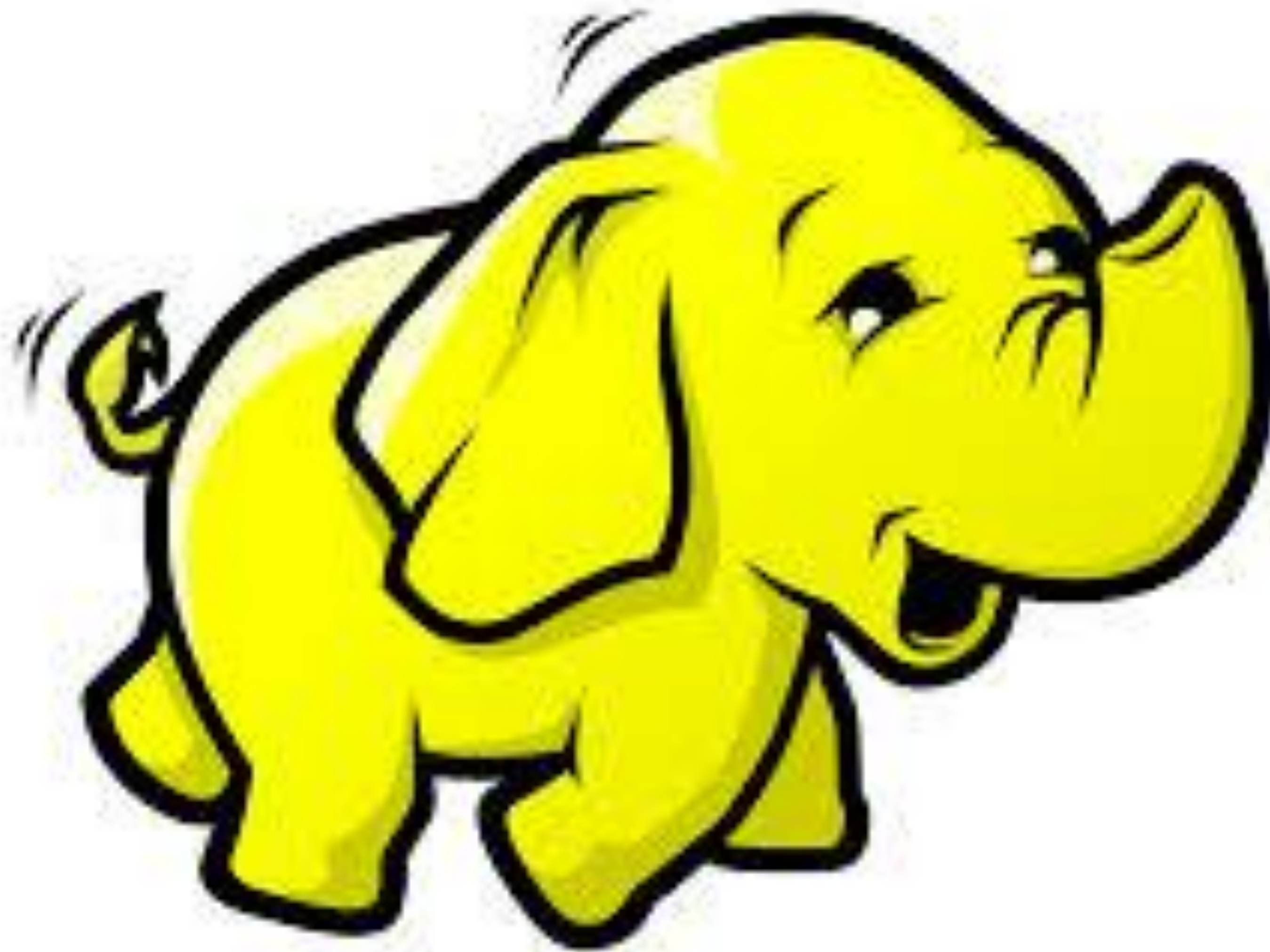


Jeffrey Dean



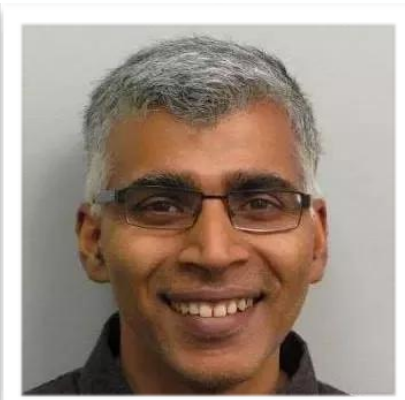
Sanjay Ghemawat





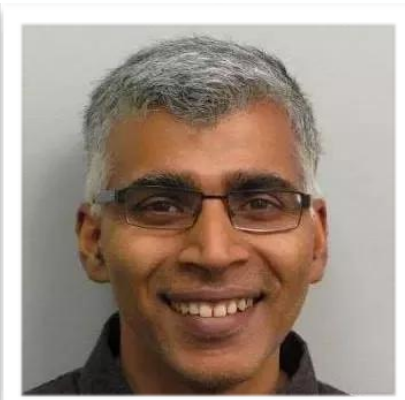


KEEP IT
SIMPLE

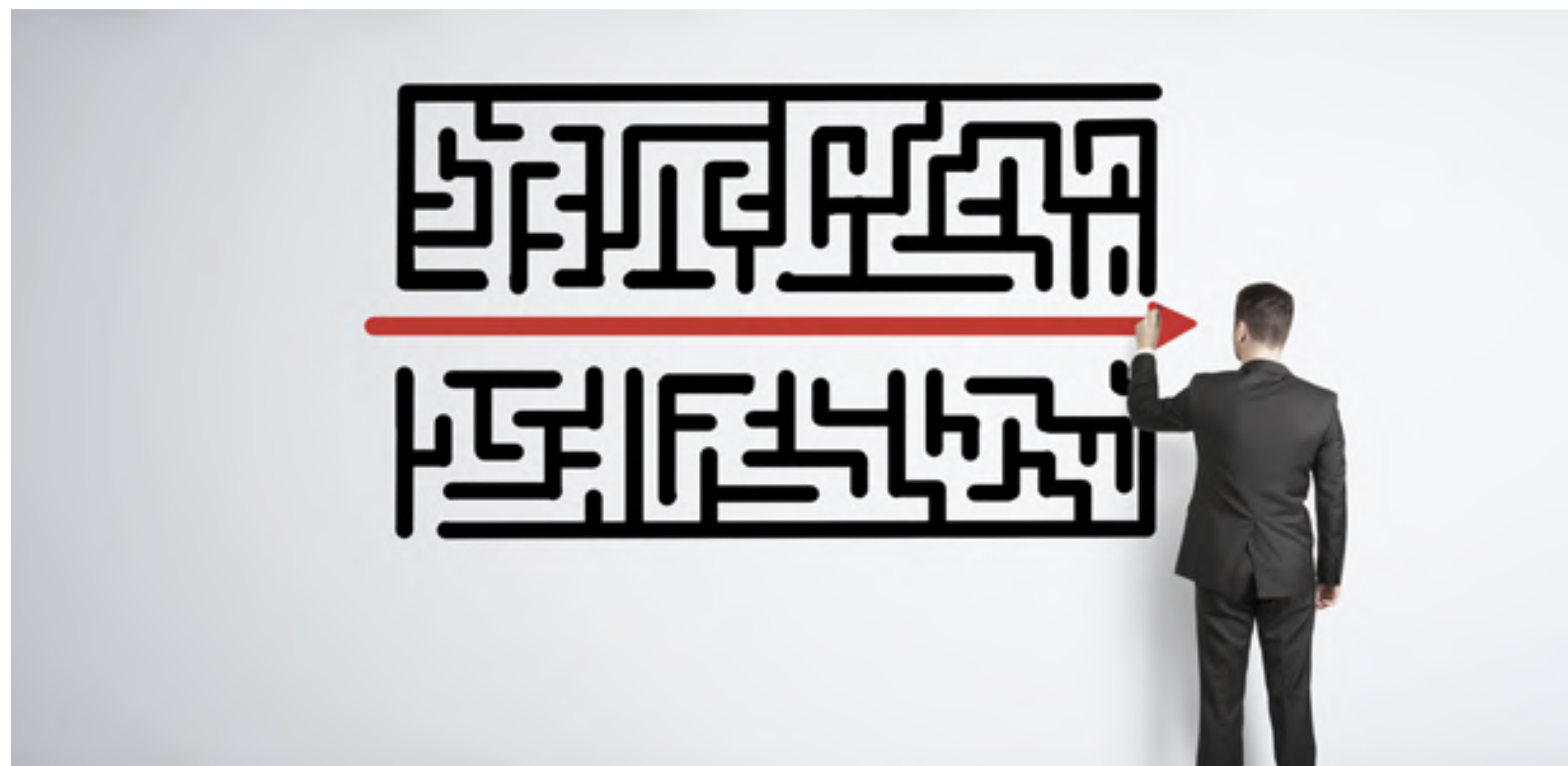


Commodities Hardware





Foco na solução e não no framework





KEEP IT
SIMPLE

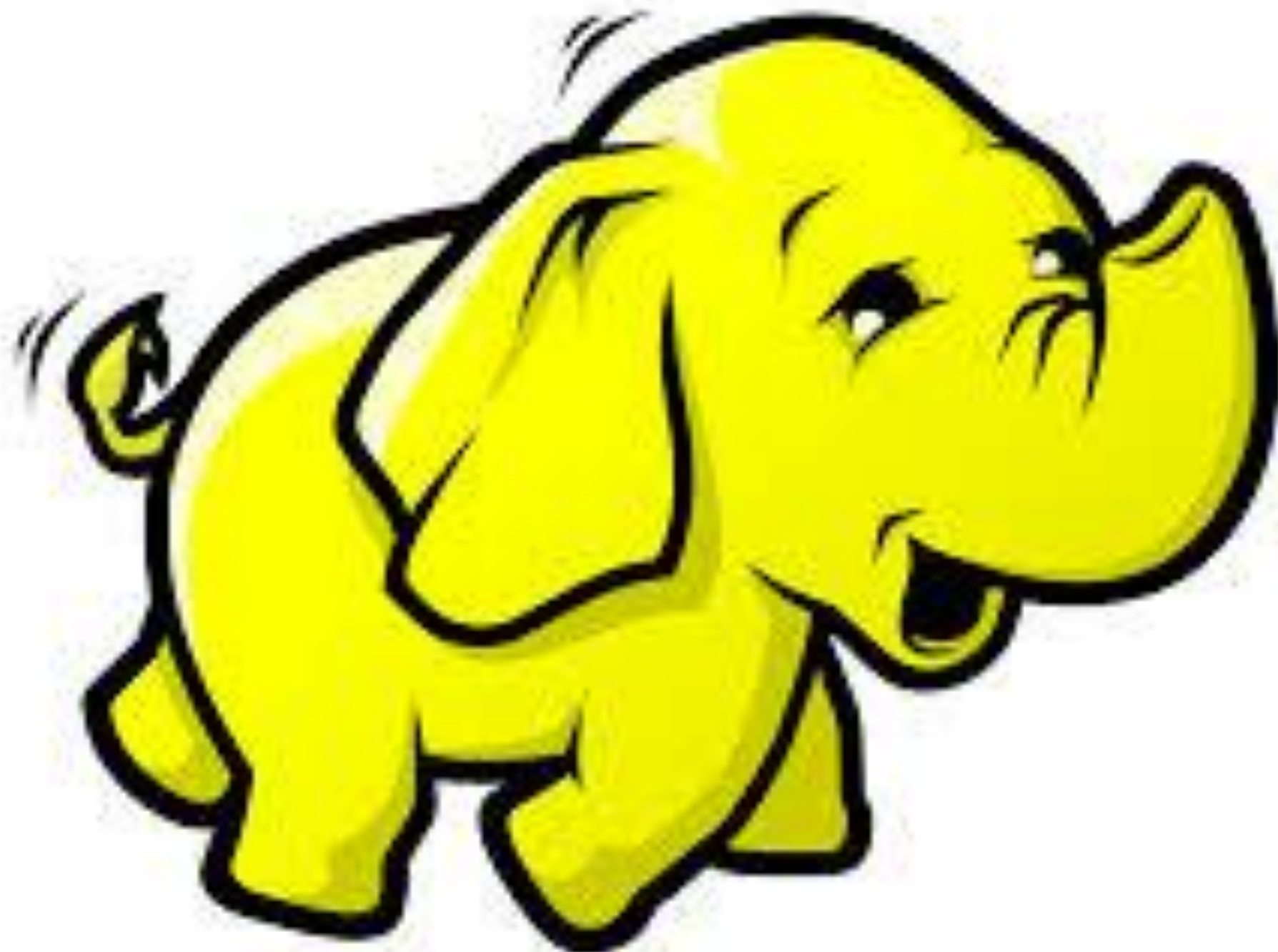


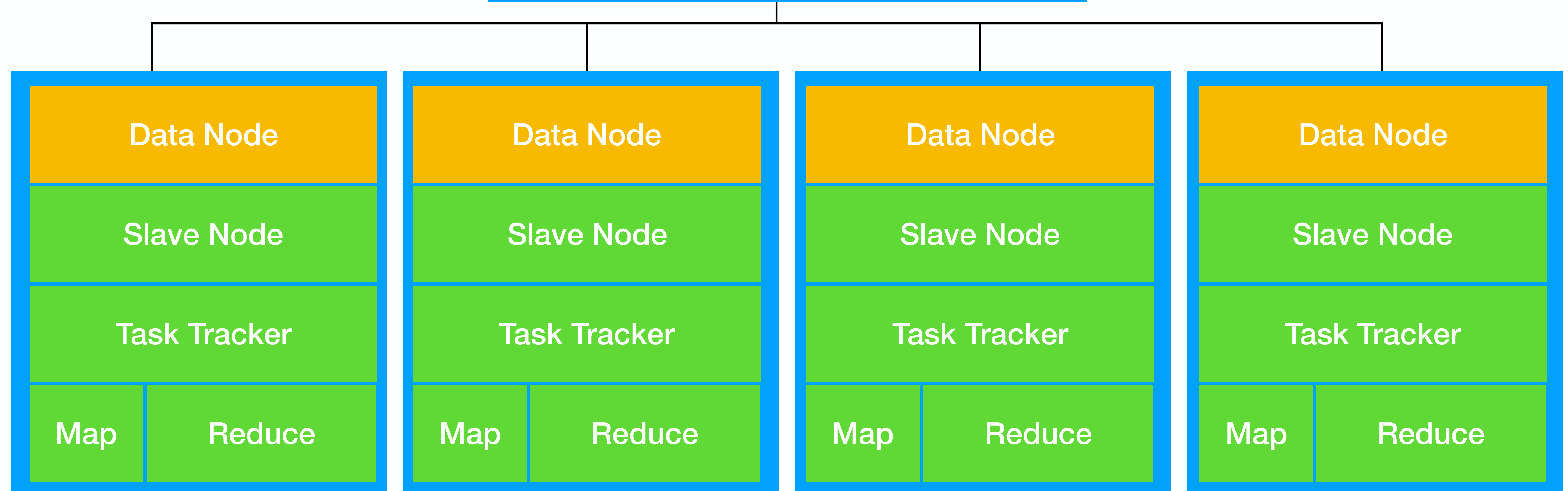
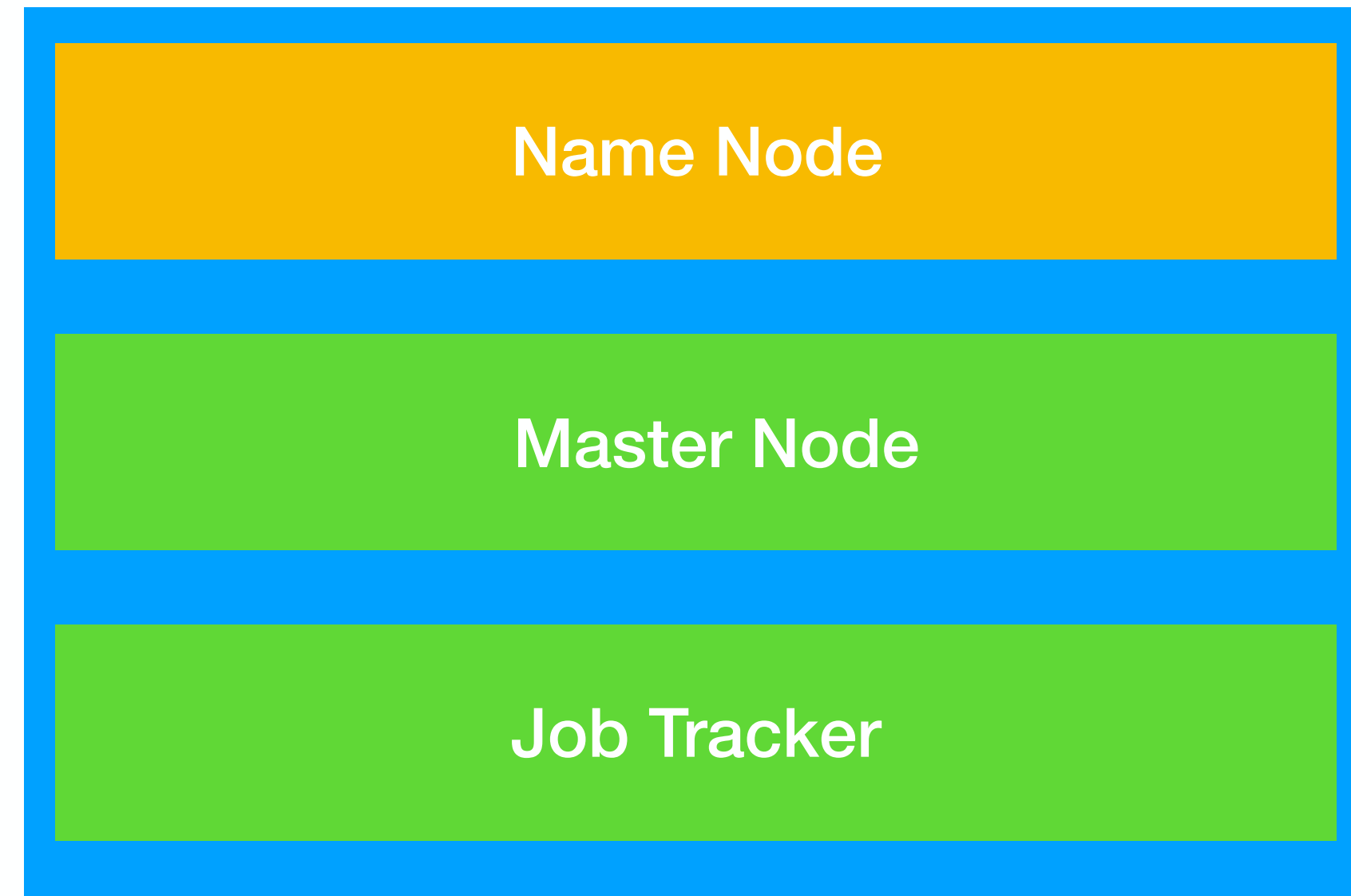
Hadoop
Common

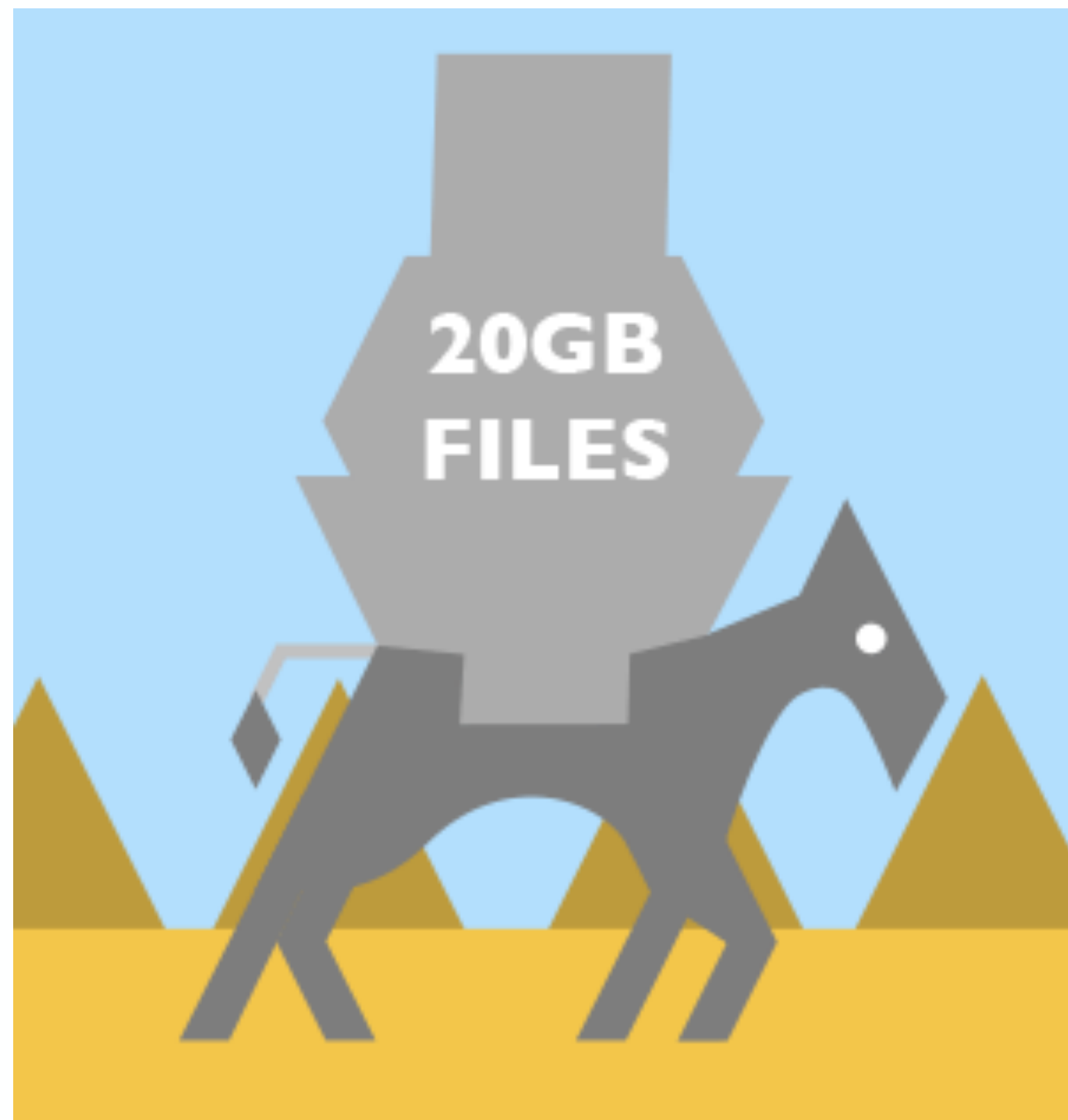
Hadoop
Yarn

Hadoop MapReduce

HDFS







OPERATION
FAULT SYSTEM PERFORMANCE
CRITICAL RELIABILITY SOFTWARE **TOLERANCE** FAILURE HARDWARE DEGRADATION DIVERSITY
GRACEFUL AVAILABILITY **CONTINUE** REDUNDANCY
REPLICATION **RAID** COMPONENT



Modulo 5

Hadoop Family



Hadoop User Experience (HUE)



Data
Exchange



Sqoop



Zoo Keeper

Coordination

Pig
Scripting



Hive
SQL



Mahout
ML



Oozie
Workflow



APACHE
HBASE

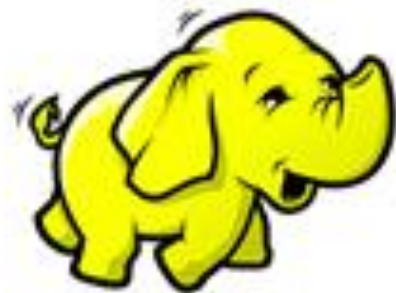
Hbase

Columnar
data
store

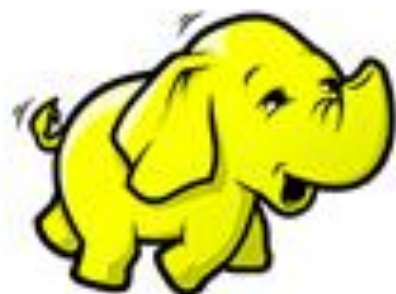
Flume



Log
Control



YARN/Map Reduce V2



Hadoop Distributed File System





Applications

Environments

YARN



DataFrames / SQL / Datasets APIs



Spark SQL

Spark Streaming

MLlib

GraphX

RDD API

Spark Core

Data Sources



APACHE
HBASE

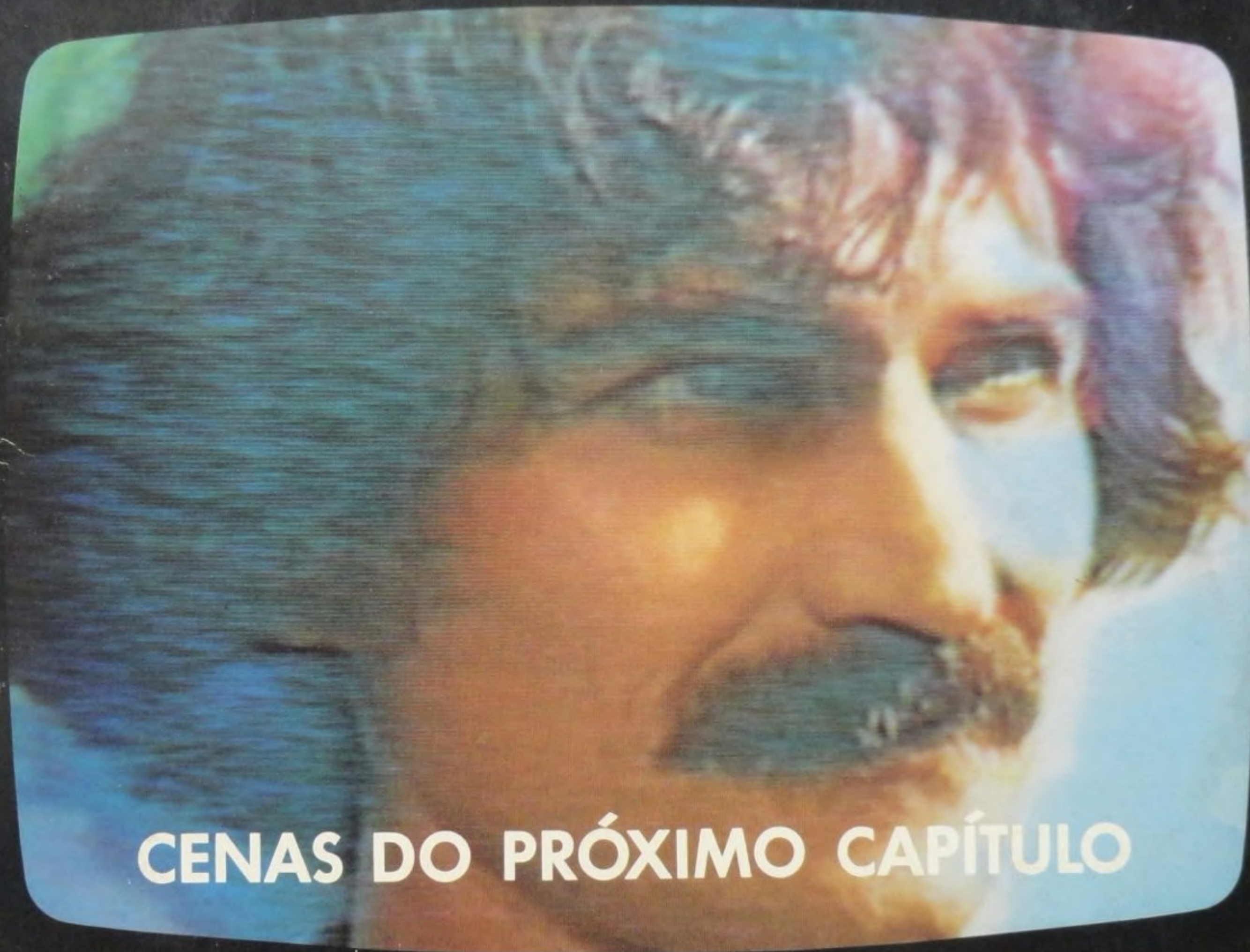




Modulo 6

Do dado à análise

B E L C H I O R



CENAS DO PRÓXIMO CAPÍTULO





THE CHALLENGE



Por hoje é tudo

Até a próxima aula ;)