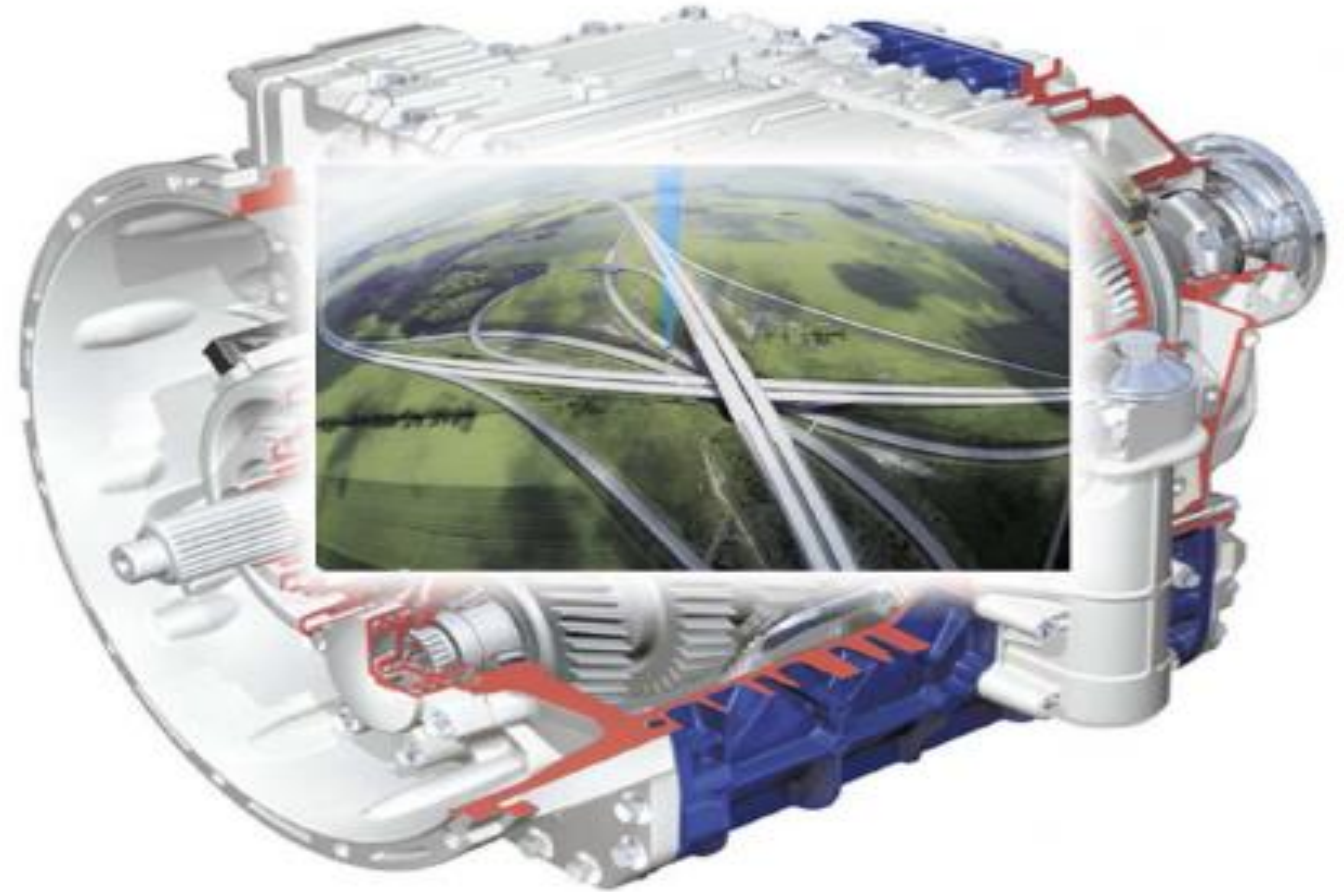


# Big Data

Intro

# Examples - Volvo

- Collect inf. of road profile
- Next time optimizing fuel consuming



# Examples - Google car

- Collect all kind of information
- Self driving

[https://techcrunch.com/2015/05/15/google-self-driving-cars-mountain-view/?ncid=rss&utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%3A+Techcrunch+%28TechCrunch%29&utm\\_content=Google+International](https://techcrunch.com/2015/05/15/google-self-driving-cars-mountain-view/?ncid=rss&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Techcrunch+%28TechCrunch%29&utm_content=Google+International)



# Examples - Formula 1 racing team 'Red Bull'

- Several TB data in just one race
- 40-50 engineers analysing real time -> modify configuration



**RB10**  
RENAULT ENERGY F1-2014

# Examples - Social medias

- Approx.  
562.000.000  
tweets / day
- 4.5 billion likes  
generated daily as  
of May 2013



# What characterize Big Data

- **Volume** - The quantity of generated and stored data.
- **Variety** - The type and nature of the data. – structured / unstructured
- **Velocity** - The speed at which the data is generated and processed.
- **Variability** - Inconsistency of the data set can hamper processes to handle and manage it.
- **Veracity** - The quality of captured data can vary greatly, affecting accurate analysis

# Big Data implementations

- Apache – Hadoop – most known

- Apache - <https://hadoop.apache.org/>

- Sandbox with bundled tools

- **Hortonworks** – <http://hortonworks.com/products/sandbox/> -- We use this

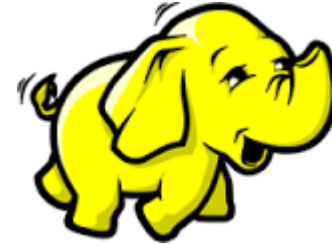
- Cloudera – <http://www.cloudera.com/downloads.html>

- Some others

- Oracle

- IBM – Watson – Beats Jeopardy world masters

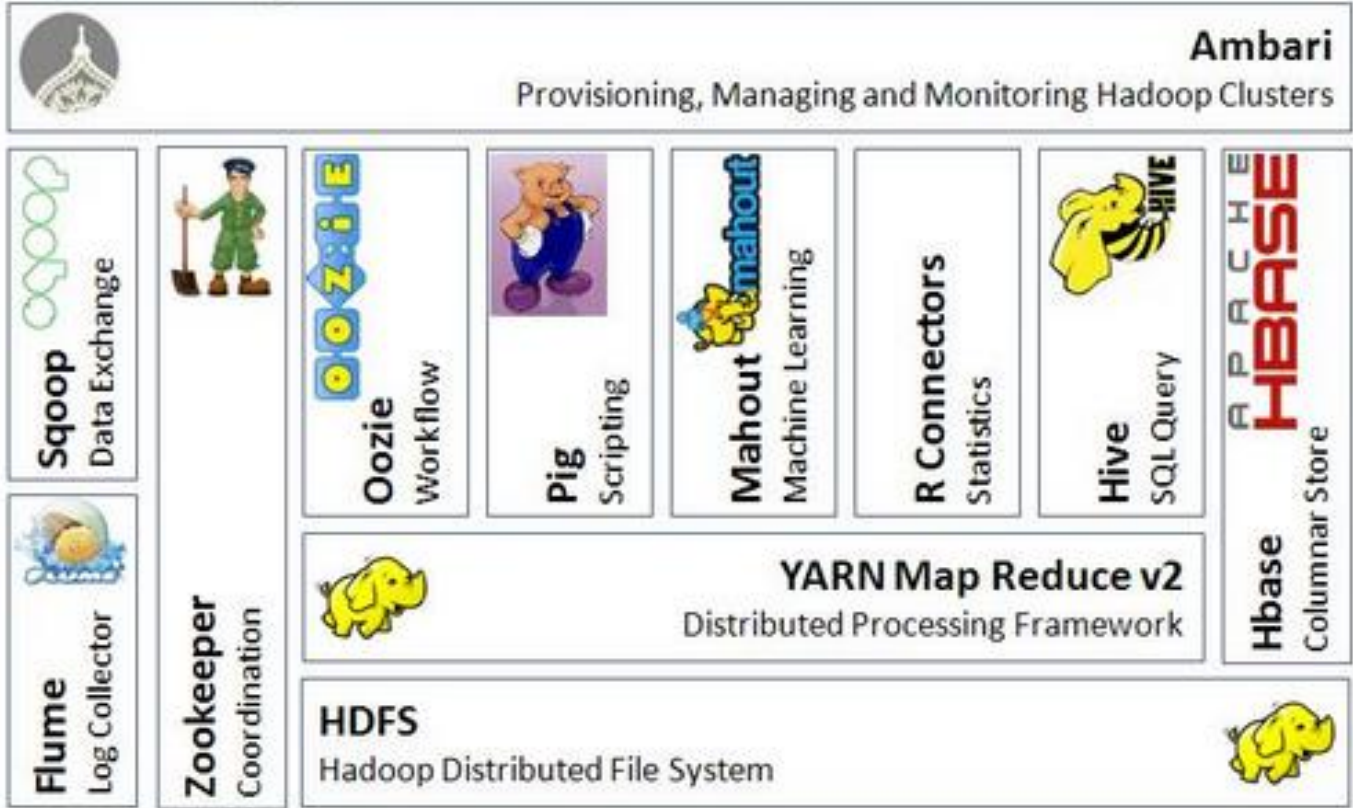
- Microsoft





# Hadoop architecture overview

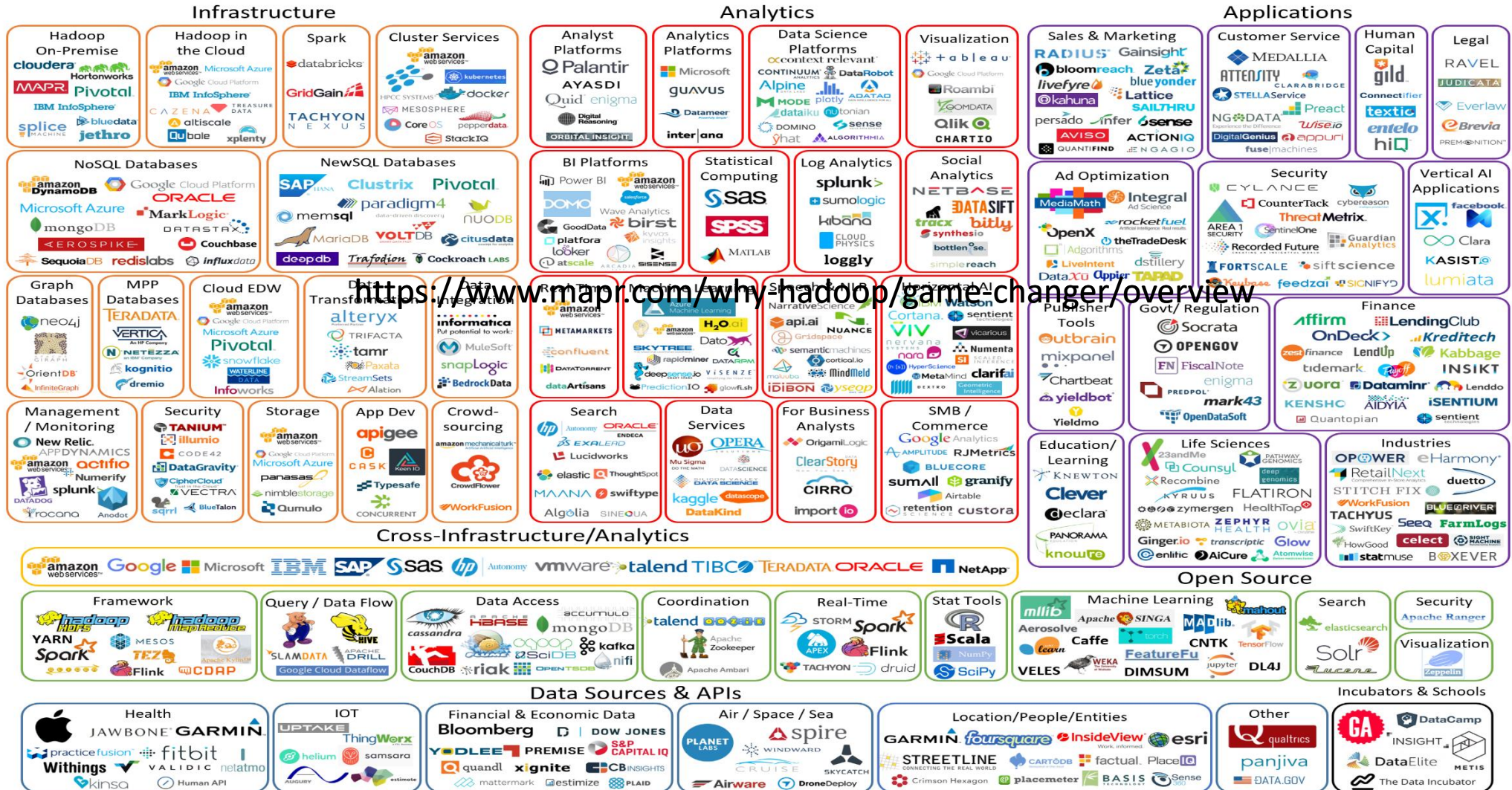
## Apache Hadoop Ecosystem





# Full picture

## Big Data Landscape 2016



# Our Architecture

## HADOOP

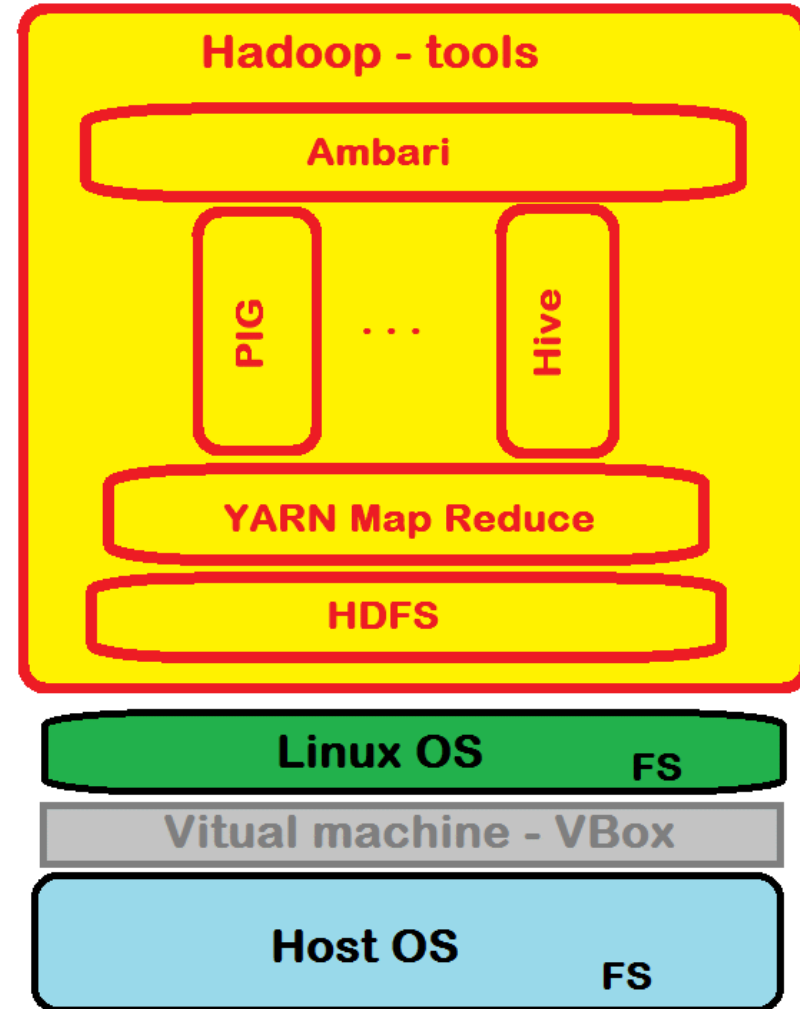
- \* User interface – Ambari
- \* Analyzing / extracting tool – PIG, Hive
- \* Map Reduce (processing file request)
- \* Hadoop File System

## System in sandbox

Linux (with a filesystem)  
eg. RedHat with ext2

## HOST OS (with a filesystem)

eg. Windows 10 with NTFS



# Hive extract / insert data

- HIVE data storage -> upon HDFS
- Hive shell -> 'command prompt' inside HIVE.
- Metastore -> inf about the DB e.g. the DB it self, tables,
- HiveQL -> SQL like interface
- Hive Clients e.g. ODBC, JDBC

# Hive Metastore

- Holds information of different storage e.g.
  - show database
  - show tables
  - describe "table" -> show design of table

# Hive Query Language -- HiveQL

- SQL like interface
- Create, insert, select
- Example

```
Select * from xxxx where yy < 10
```