

The Cloudera logo, featuring the word "cloudera" in a blue, lowercase, sans-serif font with a registered trademark symbol (®) to the upper right.The MapR logo, consisting of the word "MAPR" in a white, uppercase, sans-serif font, set against a solid red rectangular background with a registered trademark symbol (®) to the upper right.The Hortonworks logo, featuring three green silhouettes of elephants of increasing size walking to the right, positioned above the word "HORTONWORKS" in a bold, black, uppercase, sans-serif font with a registered trademark symbol (®) to the upper right.

HADOOP DISTRIBUTORS

CLOUDERA, HORTONWORKS, MAPR

What will you learn today?

- ✓ Introduction to Apache Hadoop
- ✓ Various Hadoop Distributions
- ✓ Cloudera Hadoop Distribution
- ✓ A closer look at Hortonworks and MapR
- ✓ How to choose a Hadoop Distribution



Where it all started – Apache Hadoop

The Apache Hadoop is an open source framework that allows distributed processing of large data sets across clusters of computers

Hadoop introduced a new way to simplify the analysis of large data sets, and in a very short time reshaped the big data market and have become the synonym for big data



A closer look at Apache Hadoop

Apache Hadoop includes following modules :

Hadoop Distributed File System (HDFS): A distributed file system that provides access to application data

Hadoop Common: The common utilities that support the other Hadoop modules

Hadoop YARN: A framework for job scheduling and cluster resource management

Hadoop MapReduce: A YARN-based system for parallel processing of large data sets



Popular Hadoop Distributions

cloudera®



MAPR

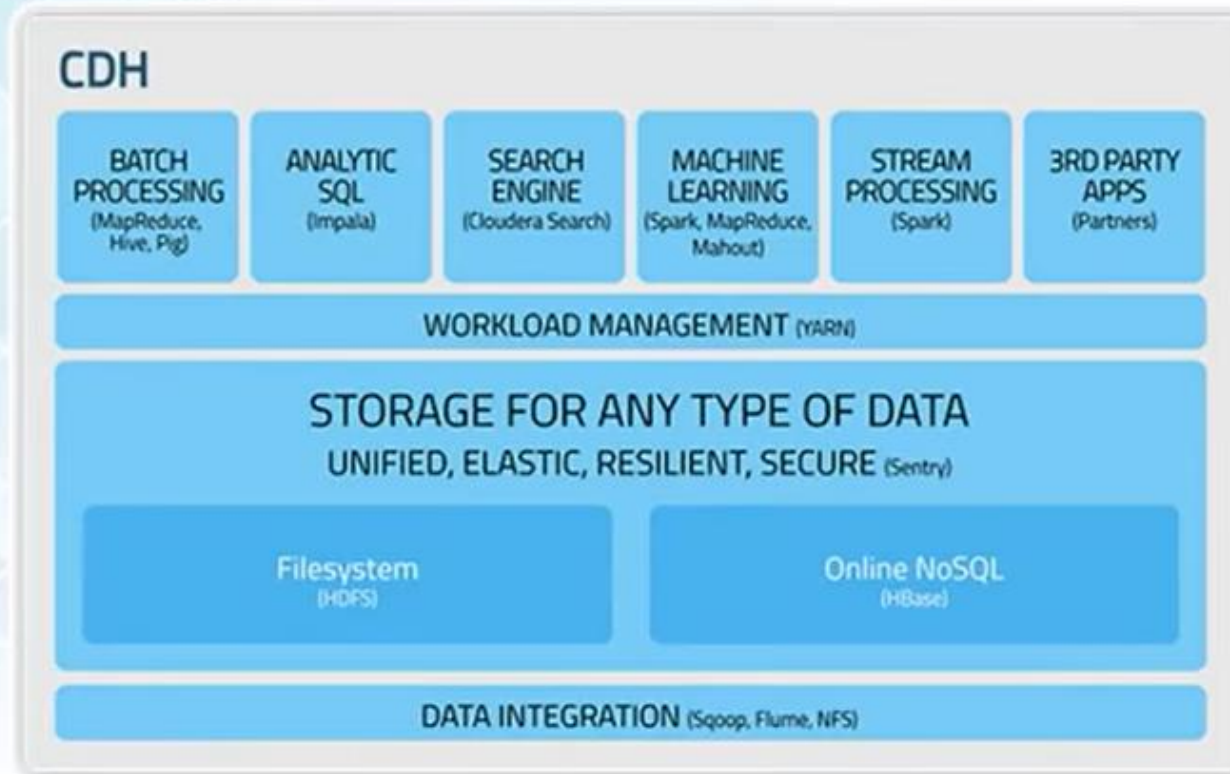
Popular Hadoop Distributions - Cloudera

Founded by a group of engineers from Yahoo, Google and Facebook [Cloudera ranks top](#) in the big data vendors list for making Hadoop a reliable platform for business use since 2008

The Cloudera logo, featuring the word "cloudera" in a bold, blue, sans-serif font, followed by a registered trademark symbol (®).

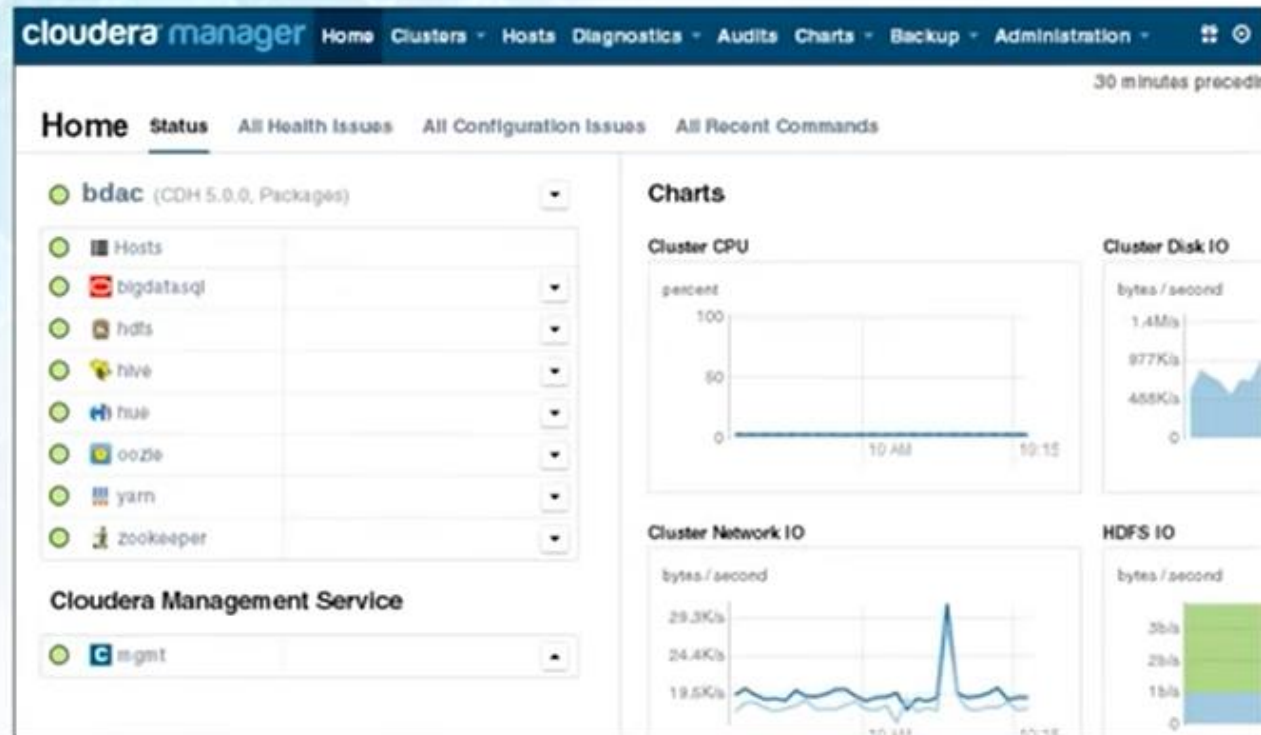
A closer look at Cloudera - CDH

Cloudera Hadoop (CDH) - CDH includes the core elements of Hadoop along with additional components such as a user interface, security, and integration with a broad range of hardware and software



A closer look at Cloudera – Cloudera Manager

Cloudera Manager makes administration of your enterprise data hub simple and straightforward, at any scale. With Cloudera Manager, you can easily deploy and centrally operate the complete Big Data stack



"Cloudera Manager takes much of the day-to-day burden of managing the environment and puts it into an easy-to-use, friendly web format. You can quickly identify where it elevates your understanding of what is going on 'under the hood.'"

– National Children's Hospital

A closer look at Cloudera – Other Products

Cloudera Express - Cloudera Express is a free download that combines CDH with Cloudera Manager, which provides robust cluster management capabilities like automated deployment, centralized administration, monitoring, and diagnostic tools

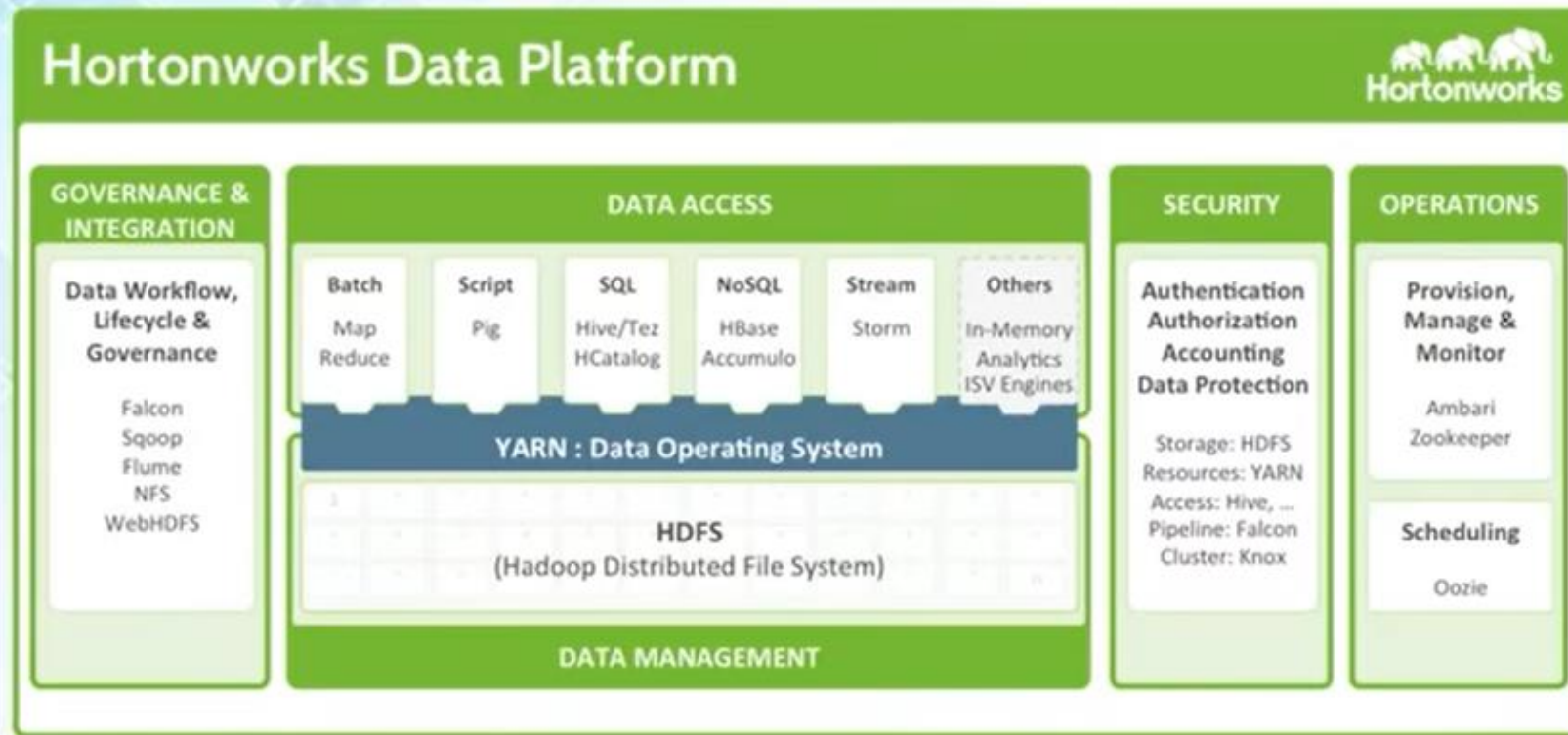
Cloudera Enterprise - Cloudera Enterprise includes CDH with advanced system management and data management tools plus dedicated support from Cloudera



Cloudera Director - Cloudera Director extends Cloudera's enterprise data hub architecture to the cloud, without compromising on security, management, and governance

Popular Hadoop Distributions - Hortonworks

Founded in 2011, [Hortonworks](#) has quickly emerged as one of the leading vendors of Hadoop



Hortonworks Data Platform

Hortonworks Sandbox

[Hortonworks Sandbox](#) lets you get started with Hortonworks Data Platform (HDP) . You can run Hortonworks Sandbox either in the cloud or on your personal machine.

HDP 2.3 on Hortonworks Sandbox

Runs on VirtualBox, VMware or Hyper-V

Try out the very latest features and functionality in Hadoop and its' ecosystem of projects with [HDP 2.3](#).

[Hortonworks Sandbox on a VM](#)

HDP 2.3 on Azure with Hortonworks Sandbox

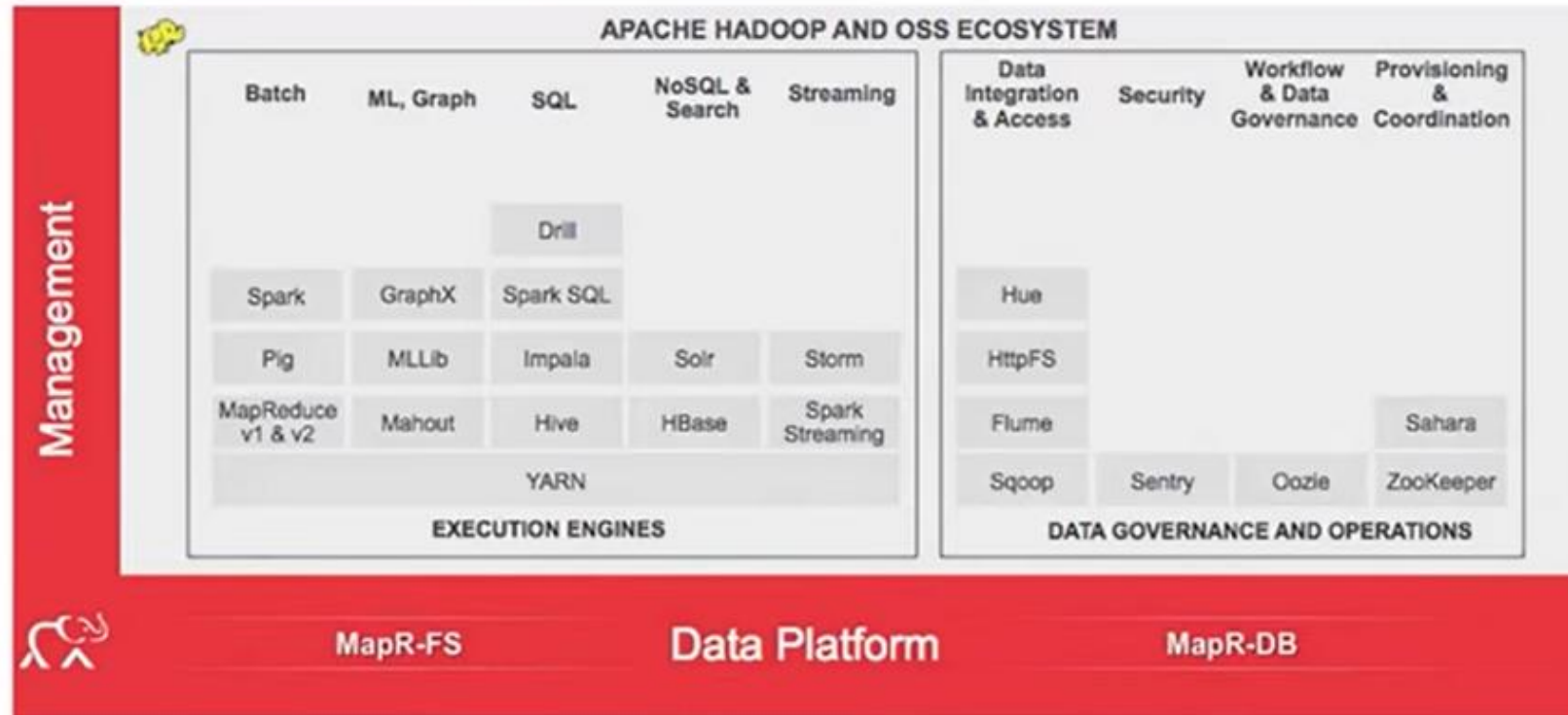
Cloud Hosted

Try Hortonworks Sandbox on Windows Azure. It's free for the the first month, and there's no need to download the VM!

[Hortonworks Sandbox in the Cloud](#)

Popular Hadoop Distributions - MapR

Compared to other Hadoop distributions e.g. Cloudera and Hortonworks, MapR takes a different approach as it uses its own proprietary file system [MapRFS](#)



[MapR Data Platform](#)

MapR Products

Two MapR Editions – M3 and M5

MAPR M3 FREE

- Control System
- NFS Access
- Performance
- Unlimited Nodes
- Free

MAPR M5 EDITION

- Control System
- NFS Access
- Performance
- High Availability
- Snapshots & Mirroring
- 24 X 7 Support
- Annual Subscription

Also Available through:



MAPR



Which one to choose ?

Before selecting the Hadoop Distribution ask yourself which problems you are trying to solve and what all features you need



Choosing a Hadoop Distribution



If you are looking for complete Hadoop stack with all features, then MapR is the way to go.

But note that MapR enterprise edition is not free and takes a different approach than Apache Hadoop

cloudera®



Choosing a Hadoop Distribution



If you are looking for complete Hadoop stack with all features, then MapR is the way to go.

But note that MapR enterprise edition is not free and takes a different approach than Apache Hadoop



Cloudera is based on 100% open source Apache Hadoop and has added its own proprietary tools

Similar to MapR, Cloudera also provides both free and paid distribution with extra features and support



Choosing a Hadoop Distribution



If you are looking for complete Hadoop stack with all features, then MapR is the way to go.

But note that MapR enterprise edition is not free and takes a different approach than Apache Hadoop



Cloudera is based on open source Apache Hadoop but has added its own proprietary tools.

Similar to MapR, Cloudera also provides both free and paid distribution with extra features and support



Hortonworks is the only commercial vendor to provide complete open source Hadoop.

Hortonworks intentionally not developed proprietary software and uses open source softwares e.g. Ambari, Stinger and Solr

Why not try them all ?

1

All Hadoop Distribution vendors provide free (community edition) version, its not a bad idea to try them all and get an idea how each one of them is different from others

Survey

Your feedback is vital for us, be it a compliment, a suggestion or a complaint. It helps us to make your experience better!

Please spare few minutes to take the survey after the webinar.



Thank you!