

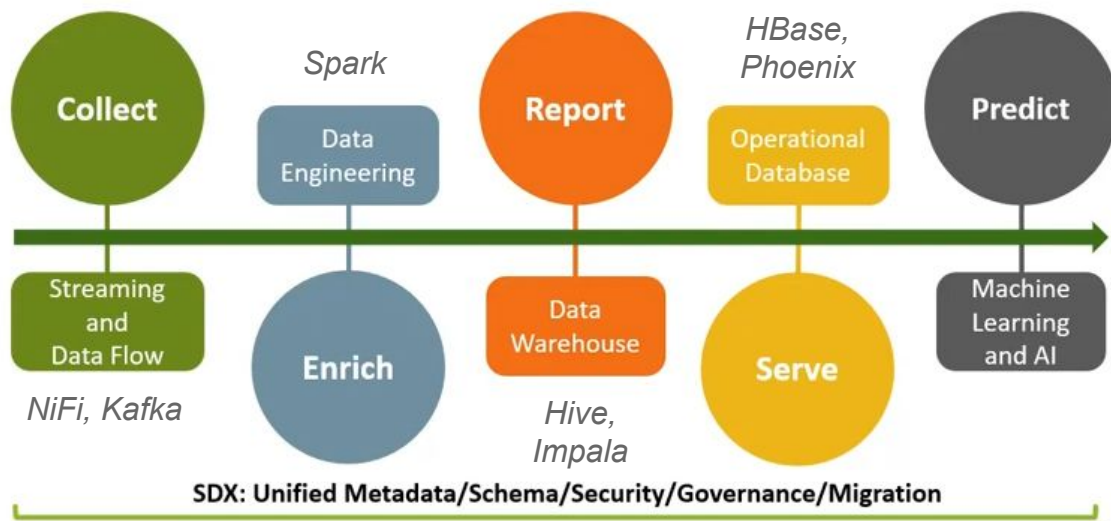
CLOUDERA

Cloudera Data Platform

Petra KAFERLE DEVISSCHERE

What is CDP?

- Cloud computing platform for businesses
- Tools to analyze and centralize data
- Security and governance at the corporate level
- Private, public and hybrid cloud

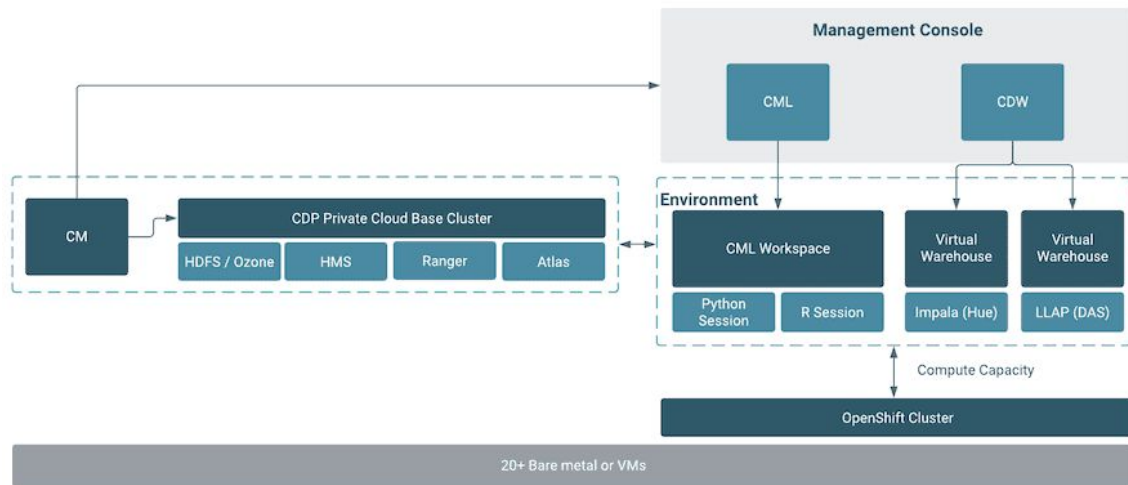


CDP Public Cloud

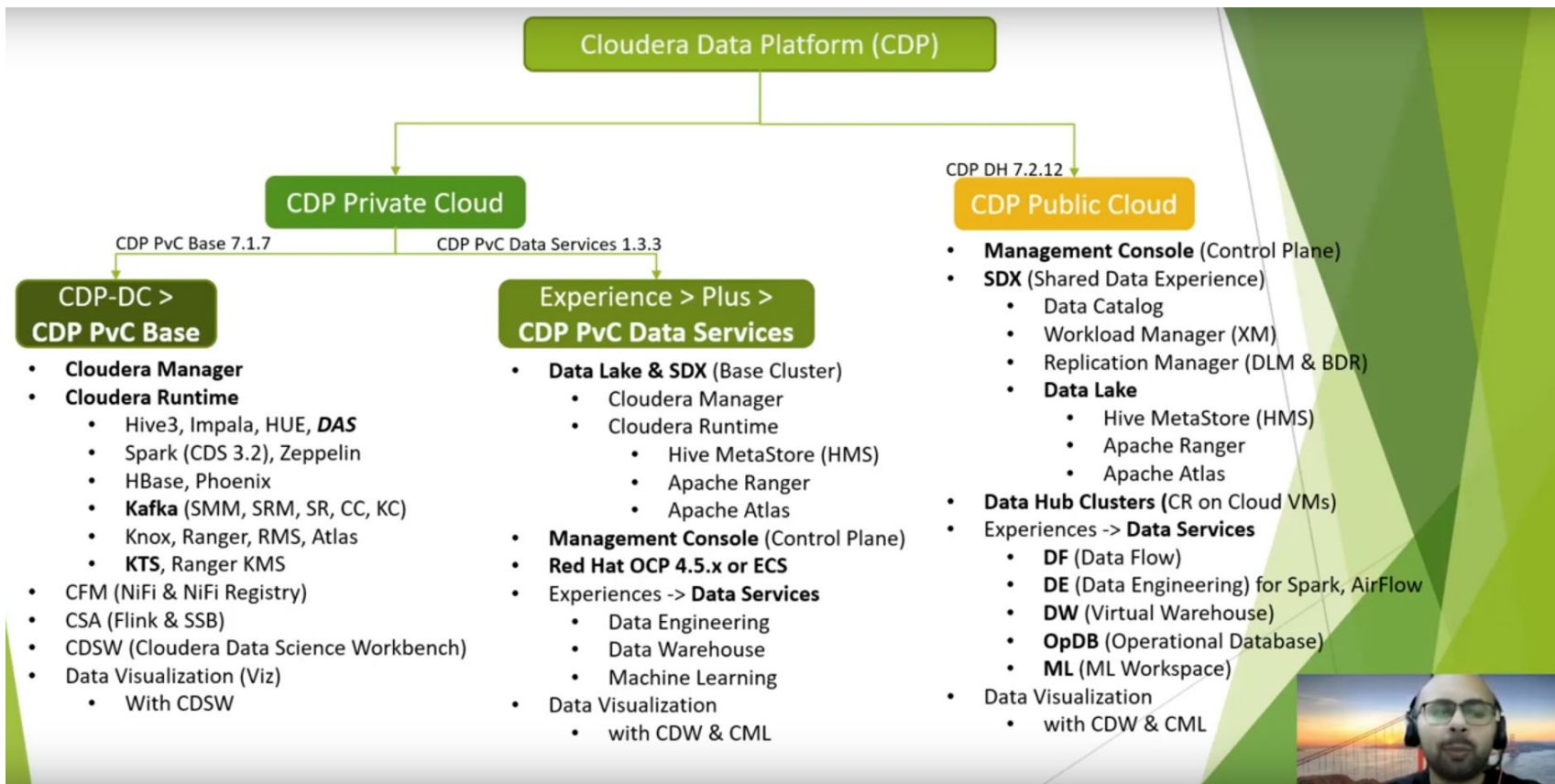
- Platform-as-a-service (PaaS)
- Could be deployed on all three public clouds and transferred among them
- It can automatically scale workloads and resources up and down in order to enhance performance and lower costs

CDP Private Cloud

- Designed for hybrid cloud deployment
- Computing and storage are decoupled, enabling clusters of these two to scale independently
- The components are much more flexible since the user has more control over the cluster deployment



CDP Private And Public Cloud



CDP Console

CLUSTERA Data Platform



A planned maintenance is coming, starting 2022-12-10 06:00:00 PM CET, through 2022-12-10 09:00:00 PM CET. During this maintenance window interactive workloads, API and CLI based integration and automation and DataHub autoscaling may be affected.

Data Services



DataFlow



Data Engineering



Data Warehouse



Operational Database



Machine Learning

Data Management



Data Hub Clusters



Data Catalog



Replication Manager



Workload Manager



Management Console

Lab Overview



1. bring the data from the Alpha Vantage to the Cloudera Data Platform with **NiFi**
 - NiFi - automates the flow of data between systems
2. store them to S3 bucket
3. create the **Iceberg** table with the help of **HUE Editor**
 - Iceberg - data lakehouse solution
 - HUE Editor - SQL Assistant for Databases & Data Warehouses
4. process data with **Spark** (jar) and ingest it into the Iceberg table
 - Spark - distributed compute engine
5. create new dataset in **Impala** based on the Iceberg table
 - Impala - low latency and high concurrency for BI/analytic queries
6. create dashboard based on the new dataset using Data Visualization tool
7. add data for a new stock and re-run the whole pipeline
8. check the Iceberg table history

