



ระบบข้อมูลขนาดใหญ่ ด้านการแพทย์และสุขภาพ กระทรวงสาธารณสุข (Big Data)



M.S. Information Technology and Management

ทำความเข้าใจ Big Data และ Analytic



รู้จัก Big Data จาก HDC



วิธีการทำให้ถึงจุดหมาย Big Data : Approach

Big Data management and Analytic

CURRENT STATE

BI/DW CHALLENGES

- Batch oriented, inflexible, high latency
- Brittle & labor intensive
- Performance & scalability challenged
- Aggregated, silo'd, structured data

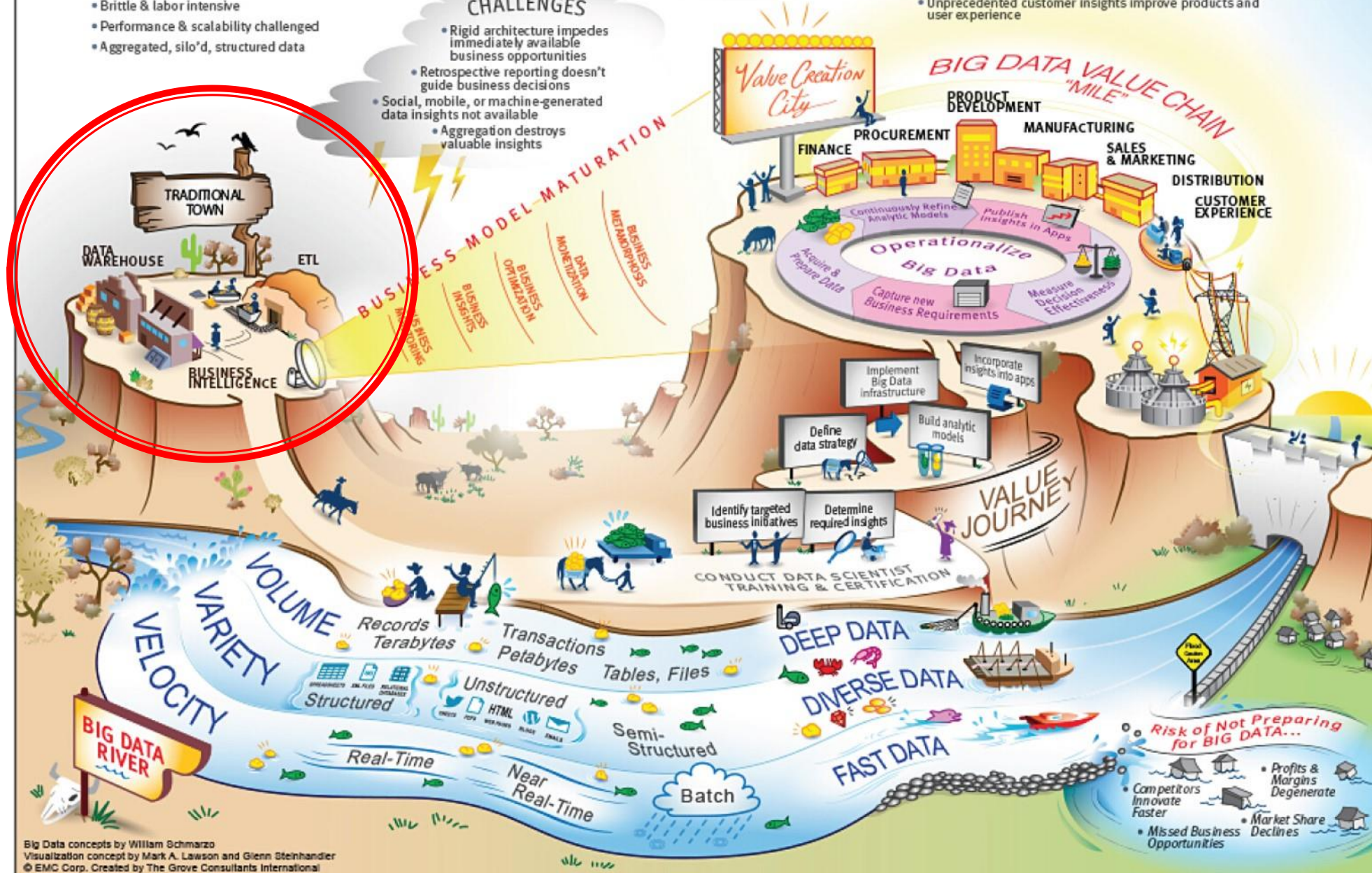
BUSINESS CHALLENGES

- Rigid architecture impedes immediately available business opportunities
- Retrospective reporting doesn't guide business decisions
- Social, mobile, or machine-generated data insights not available
- Aggregation destroys valuable insights



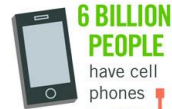
FUTURE STATE

- Big Data analytics continuously drive business innovations
- Real-time insights optimize operations and boost profits
- Unprecedented customer insights improve products and user experience



40 ZETTABYTES

[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005



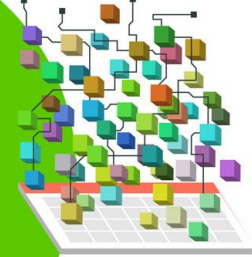
6 BILLION PEOPLE
have cell phones

WORLD POPULATION: 7 BILLION

Volume SCALE OF DATA

It's estimated that
2.5 QUINTILLION BYTES

[2.3 TRILLION GIGABYTES]
of data are created each day



Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

Velocity ANALYSIS OF STREAMING DATA

The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session

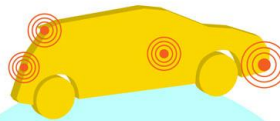


By 2016, it is projected
there will be
18.9 BILLION NETWORK CONNECTIONS

— almost 2.5 connections
per person on earth



Modern cars have close to
100 SENSORS
that monitor items such as
fuel level and tire pressure



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data,
with 1.9 million in the United States



As of 2011, the global size of
data in healthcare was
estimated to be

150 EXABYTES
[161 BILLION GIGABYTES]



**30 BILLION
PIECES OF CONTENT**
are shared on Facebook
every month



Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated
there will be

**420 MILLION
WEARABLE, WIRELESS
HEALTH MONITORS**



**4 BILLION+
HOURS OF VIDEO**
are watched on
YouTube each month



400 MILLION TWEETS
are sent per day by about 200
million monthly active users



**1 IN 3 BUSINESS
LEADERS**

don't trust the information
they use to make decisions



Poor data quality costs the US
economy around
\$3.1 TRILLION A YEAR



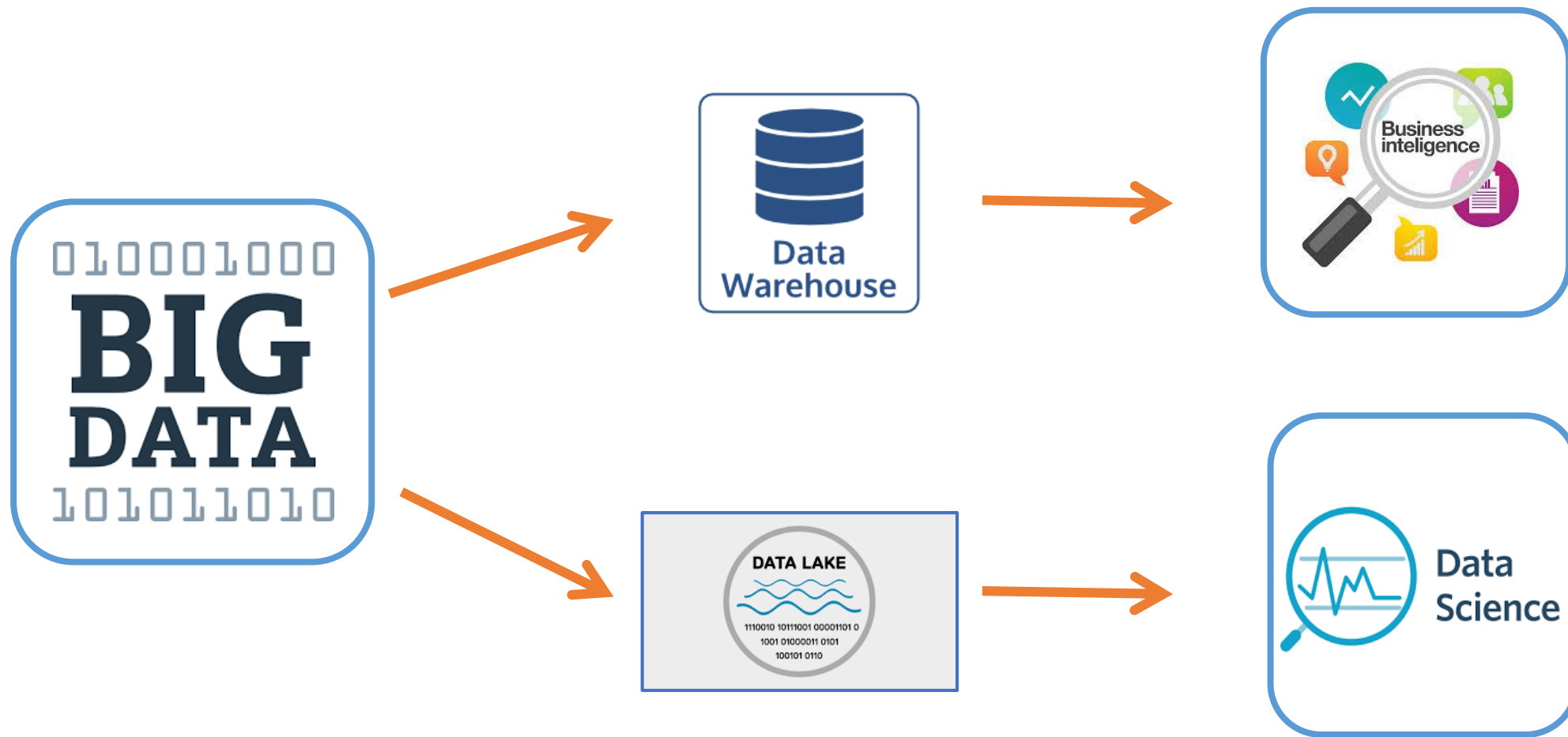
**27% OF
RESPONDENTS**

in one survey were unsure of
how much of their data was
inaccurate

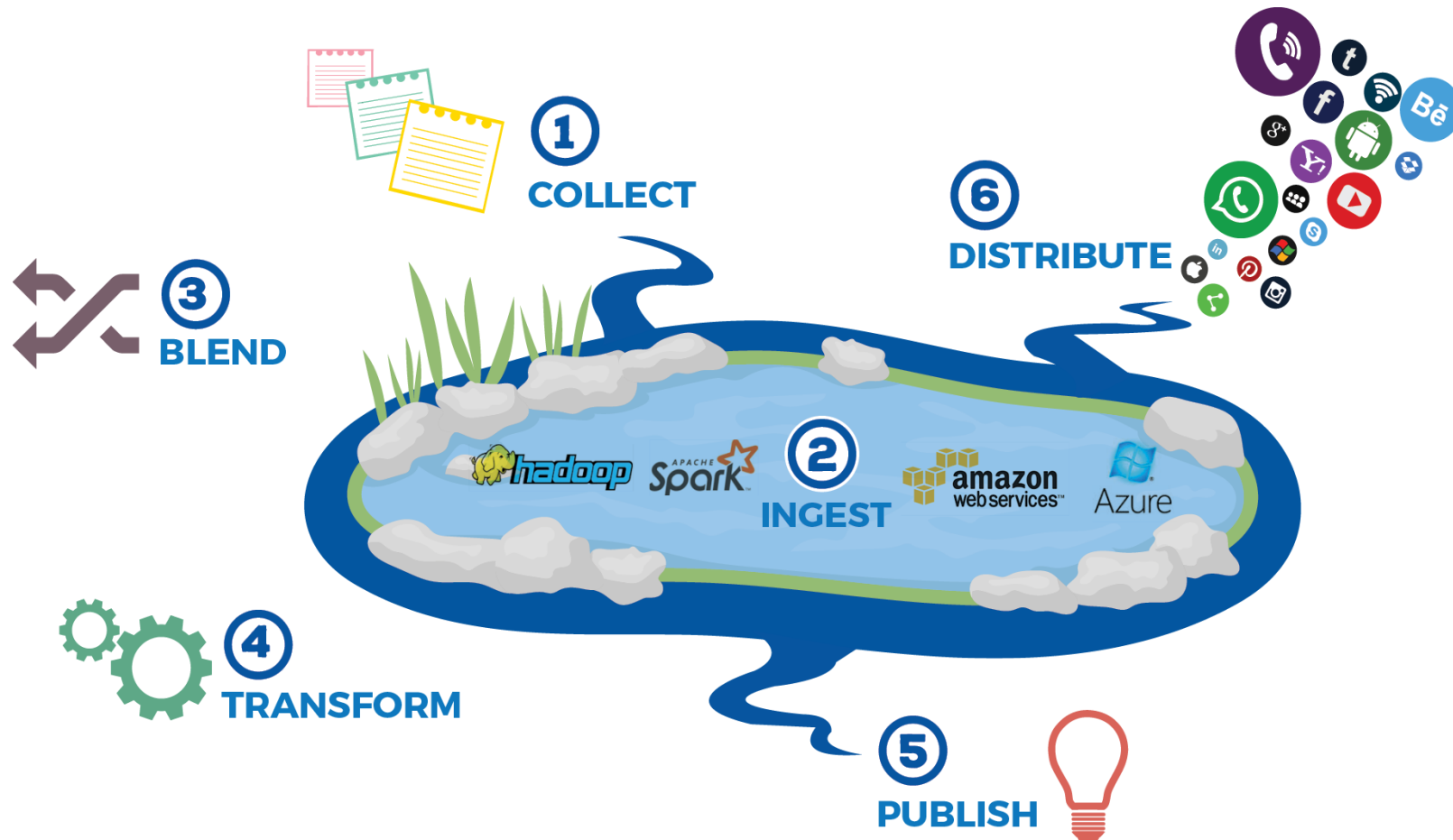
Veracity UNCERTAINTY OF DATA



Big Data



Data Lake

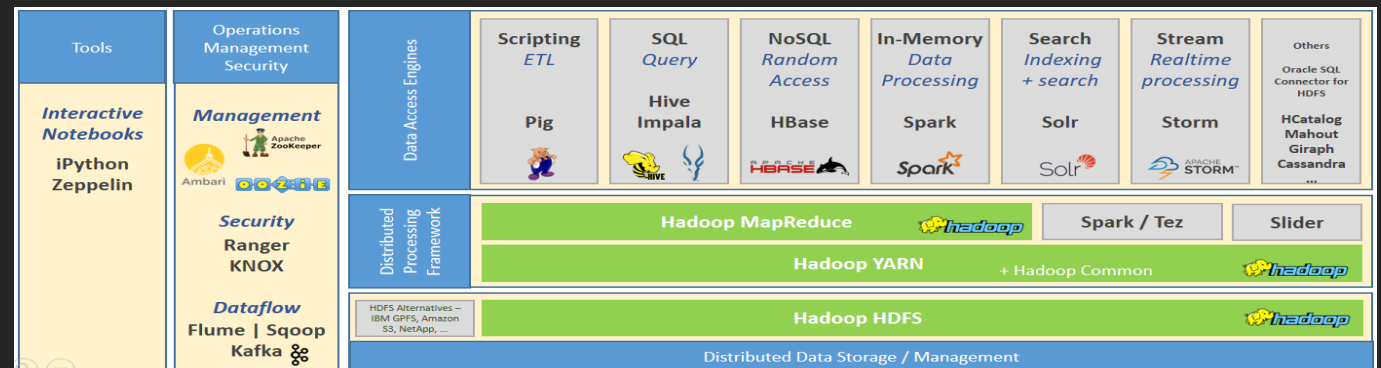


เทคโนโลยีสำหรับประมวลผล Big Data

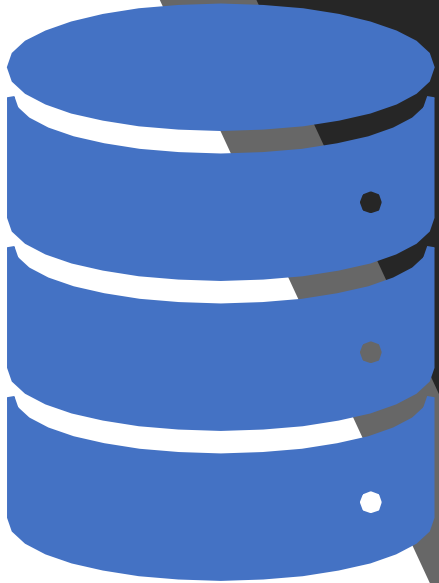


1) กลุ่มยุกบูกเบิก (กลุ่มที่ใช้ Hadoop และ พันธมิตร)

ระบบการจัดการเครื่องแม่ข่ายในลักษณะคลัสเตอร์ เข้าถึงและดึงข้อมูลอย่างรวดเร็วด้วยวิธี MapReduce (Map และ Reduce) ระบบคอมพิวเตอร์ที่จะรองรับการทำงานของ Hadoop จะเป็นกลุ่มเครื่องแม่ข่ายขนาดเล็กหลายๆ เครื่องที่มีหน่วยจัดเก็บข้อมูลขนาดใหญ่ต่อเชื่อมกันผ่านระบบเครือข่าย (Local Area Network) หรือเครือข่ายระยะไกล (Wide Area Network)



เทคโนโลยีสำหรับประมวลผล Big Data

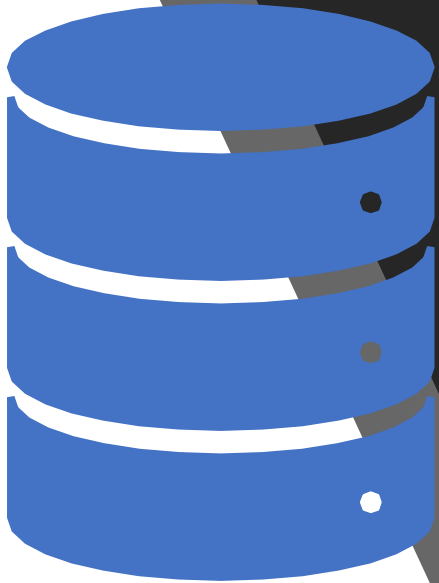


2) ระบบฐานข้อมูลที่ไม่ใช้ภาษา SQL (NoSQL Database)

ระบบฐานข้อมูลที่ไม่ใช้ภาษา SQL มีความสามารถที่รวดเร็ว สามารถรองรับข้อมูลแบบ Semi-Structured และ Unstructured ได้ รองรับการขยายตัวในแนวนอน (Horizontal Scaling) ซึ่งสอดคล้องกับสถาปัตยกรรมของ Hadoop ตัวอย่างผลิตภัณฑ์ทางด้าน NoSQL Database ที่เป็นที่นิยม ได้แก่ Cassandra, CouchBase, HBase, [MongoDB](#)



เทคโนโลยีสำหรับประมวลผล Big Data

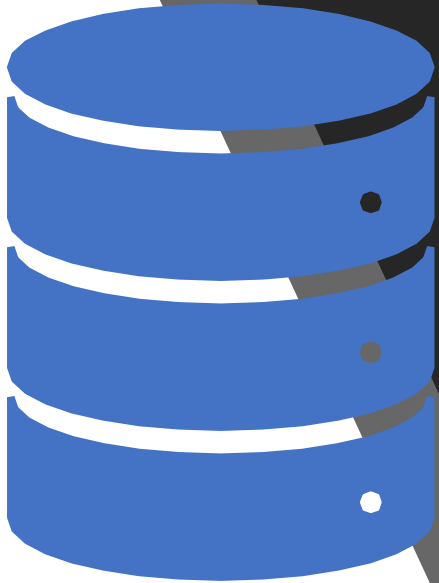


3) Data Visualization Tools

เครื่องมือที่จะช่วยแปลงข้อมูล “Big Data” ที่ได้รับการกลั่นกรองแล้ว มาแสดงในรูปของแผนภาพ ง่ายต่อการเข้าใจ และนำไปสู่การตัดสินใจในขั้นถัดไป บทบาทของเครื่องมือกลุ่มนี้จะอยู่ในระดับปฏิบัติการ ให้ติดตามสถานะของระบบและการแก้ปัญหาได้ง่าย



เทคโนโลยีสำหรับประมวลผล Big Data



4) กลุ่ม Analytic Database

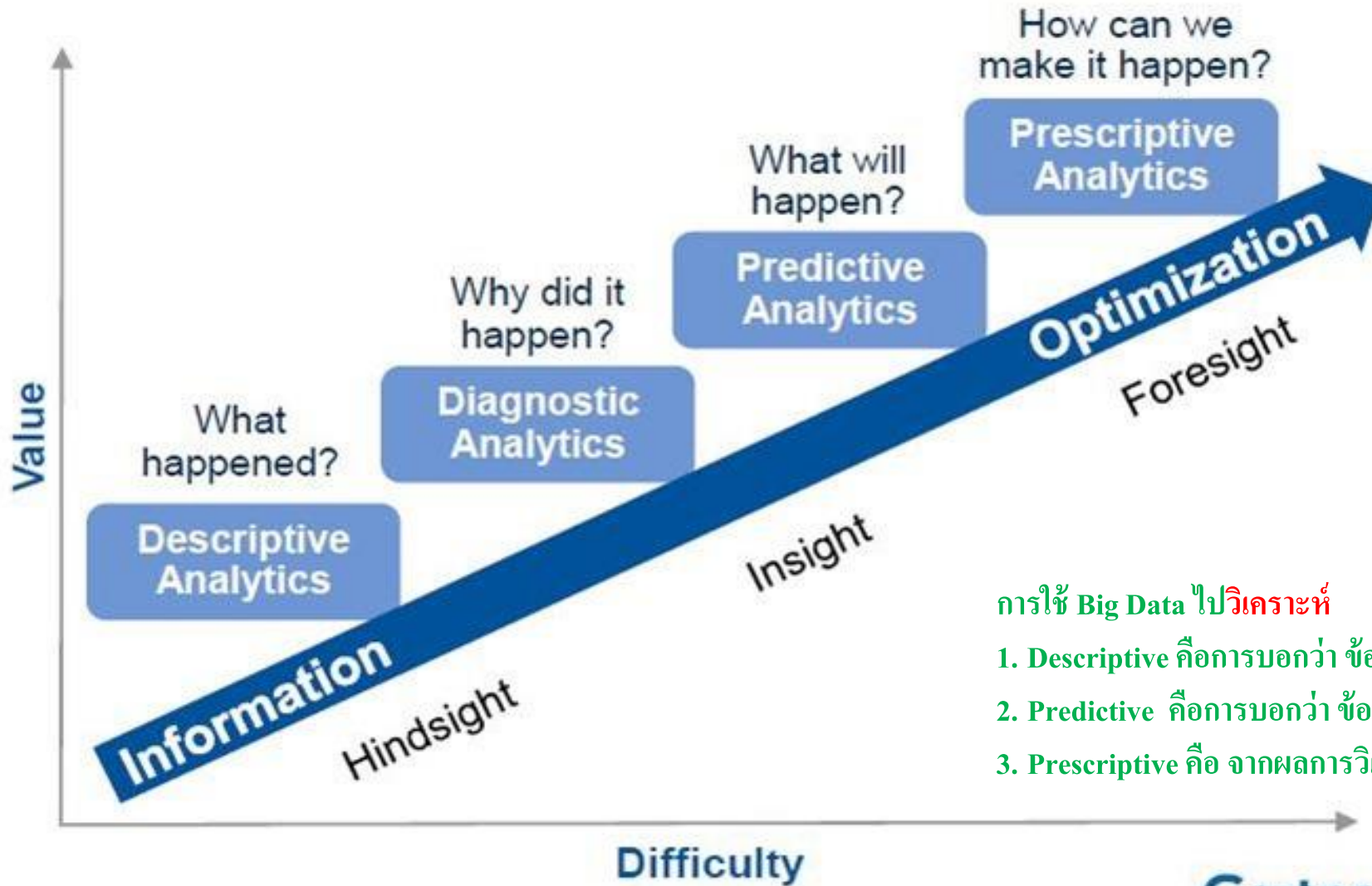
ผลิตภัณฑ์ฐานข้อมูล ค่าย ยักษ์ใหญ่ ที่ออกผลิตภัณฑ์ มาให้ตอบโจทย์ตลาด Big Data มีการใช้เทคนิคในการทำงานแบบต่างๆ เพื่อตอบโจทย์ด้านความเร็ว ไม่ว่าจะเป็นการประมวลผลใน หน่วยความจำ (In-memory Computing) การประมวลผลในระบบฐานข้อมูล (In-database Computing)

ตัวอย่างผลิตภัณฑ์ในกลุ่มนี้ได้แก่ Aster Data (Teradata), Exadata (Oracle), Greenplum (EMC) Netezza (IBM), **Vertica (HP)**



DATA WAREHOUSE	vs.	DATA LAKE
structured, processed	DATA	structured / semi-structured / unstructured, raw
schema-on-write	PROCESSING	schema-on-read
hierarchically archived	STORAGE	object-based, no hierarchy
less agile, fixed configuration	AGILITY	highly agile, configure and reconfigure as needed
mature	SECURITY	maturing
business professionals	USERS	data scientists et. al.

Data Analytics



การใช้ Big Data ไปวิเคราะห์

1. Descriptive คือการบอกว่า ข้อมูลที่มีอยู่ มีอะไรอยู่บ้าง
2. Predictive คือการบอกว่า ข้อมูลนั้น สามารถคาดการณ์ หรือ ทำนาย อะไรได้บ้าง
3. Prescriptive คือ จากผลการวิเคราะห์ข้อมูล จะบอกว่า เราควรทำอะไรบ้าง

Gartner.

Traditional Analytics (BI)

vs

Big Data Analytics

Focus on

- Descriptive analytics
- Diagnosis analytics

- **Predictive analytics**
- **Data Science**

Data Sets

- Limited data sets
- Cleansed data
- Simple models

- Large scale data sets
- More types of data
- Raw data
- Complex data models

Supports

Causation: what happened,
and why?

Correlation: new insight
More accurate answers

ความเข้าใจใหม่/ข้อมูลเชิงลึก
คำตอบที่ถูกต้องมากขึ้น

สาเหตุ VS ความสัมพันธ์

The Differences

Hortonworks

Hortonworks offers Apache Foundation certified software

Hortonworks Embeds Hadoop into existing data platforms

HDP a Native component

No Proprietary Software

Open Source License

Free

Cloudera

Cloudera sells Commercial software

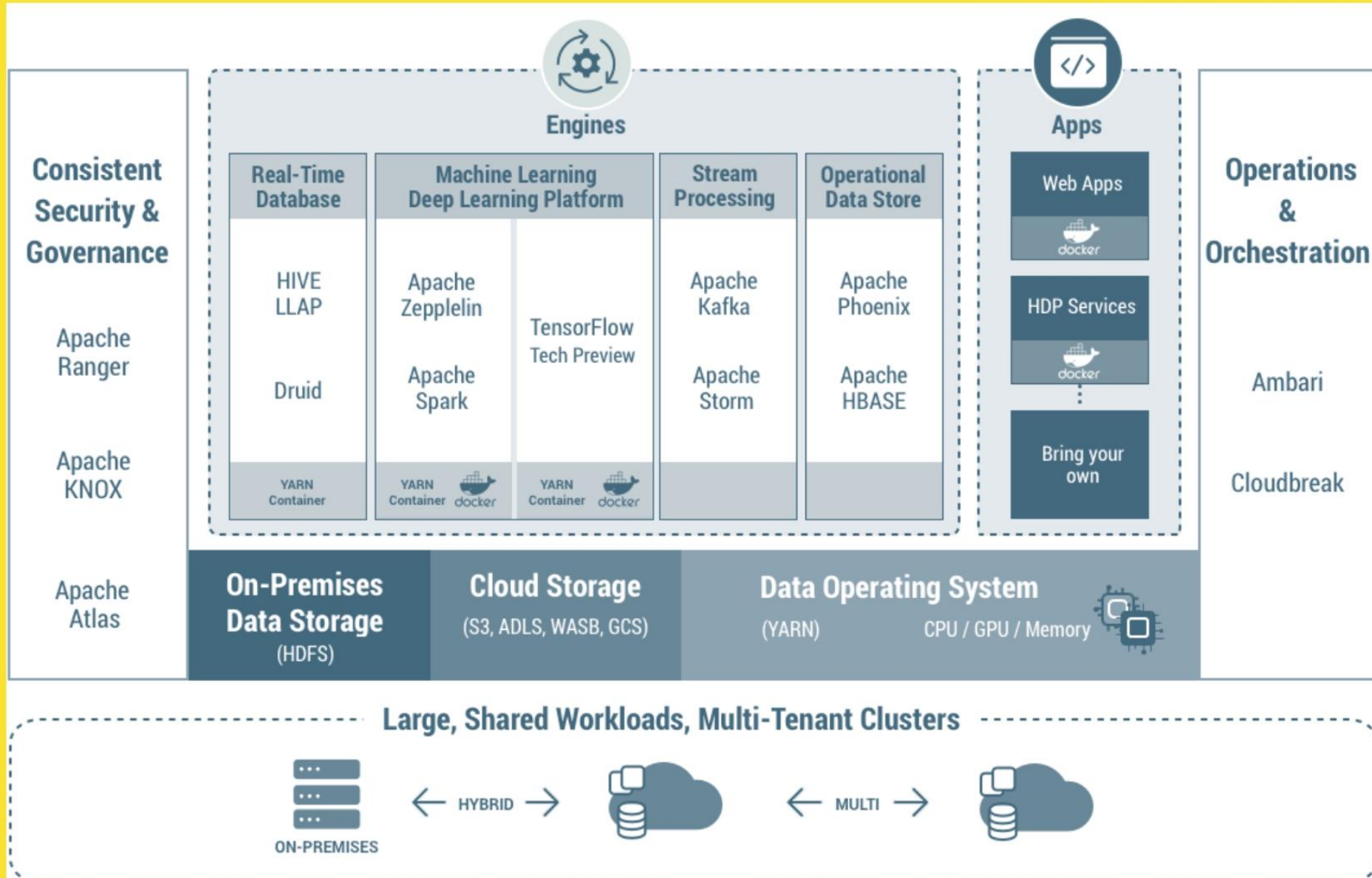
Cloudera Embeds with other commercial software providers

Cloudera CDH is Not a native component

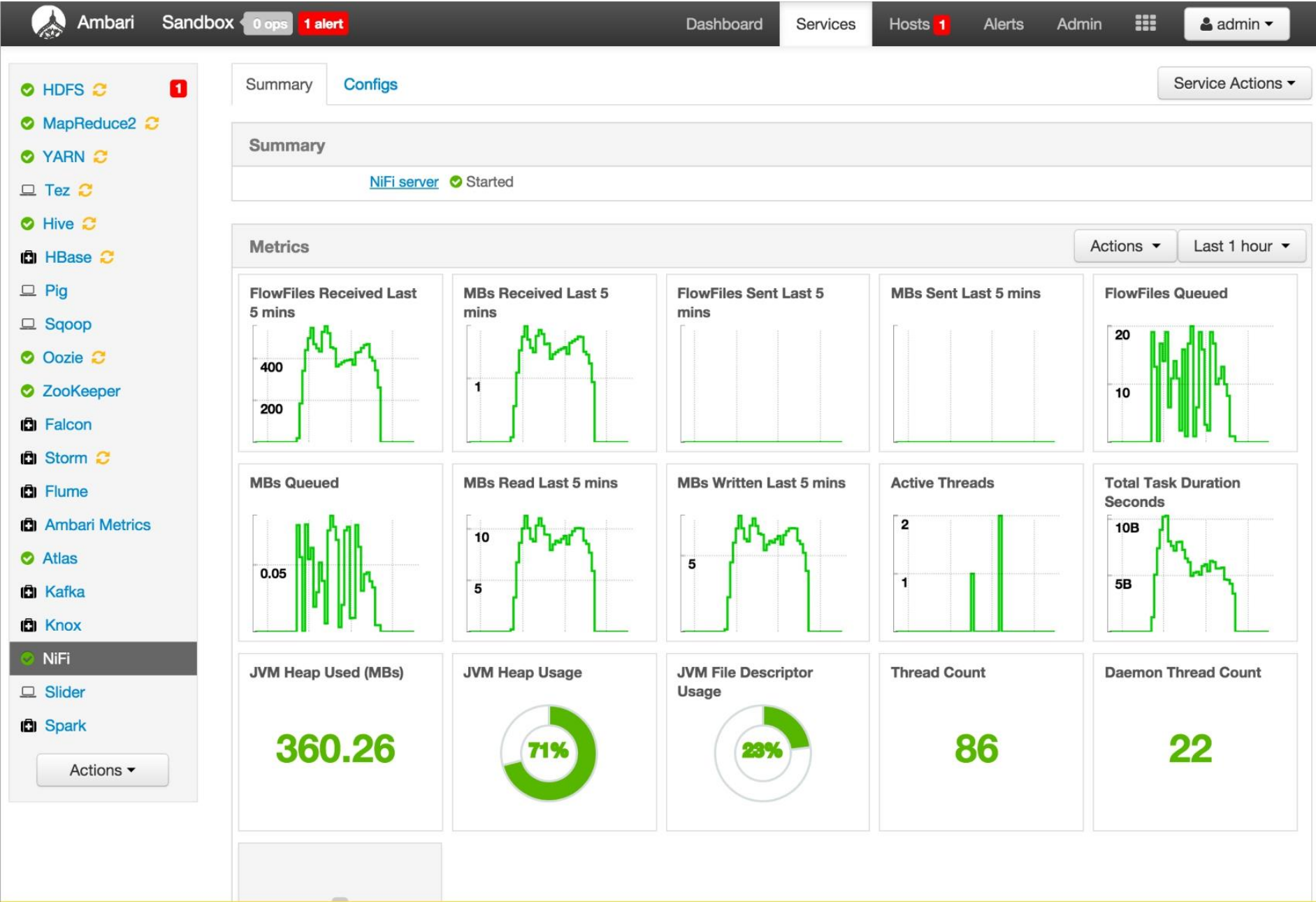
Proprietary Software

Commercial License

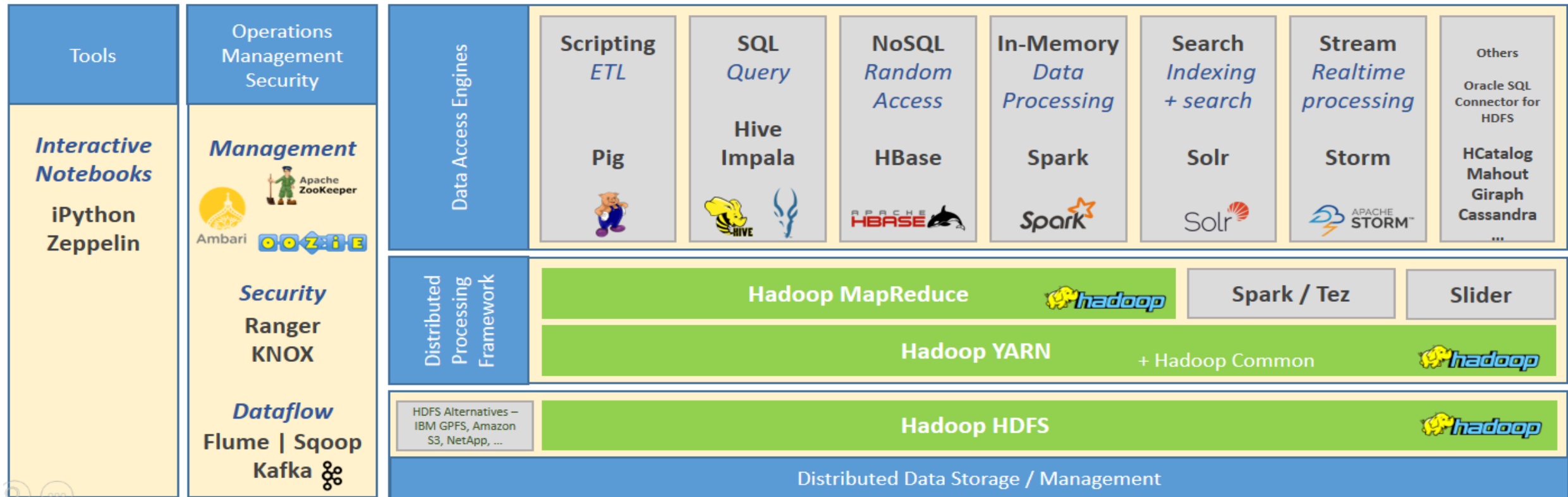
Paid



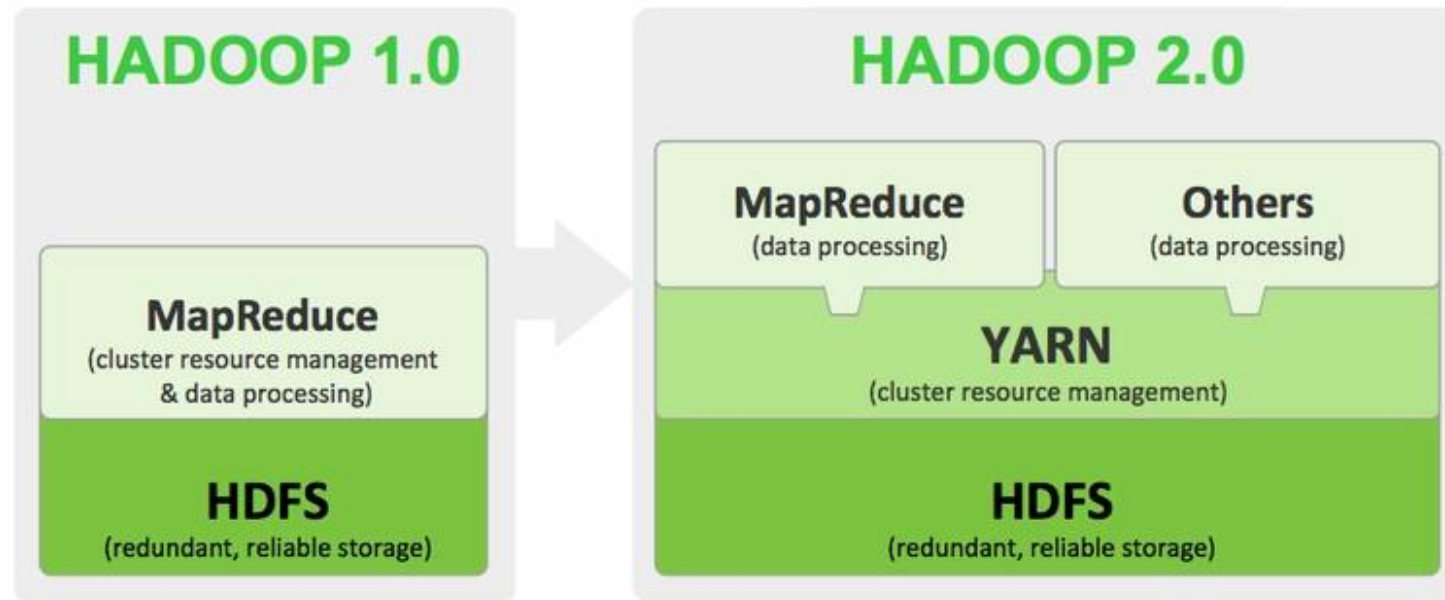
Hortonworks Manager



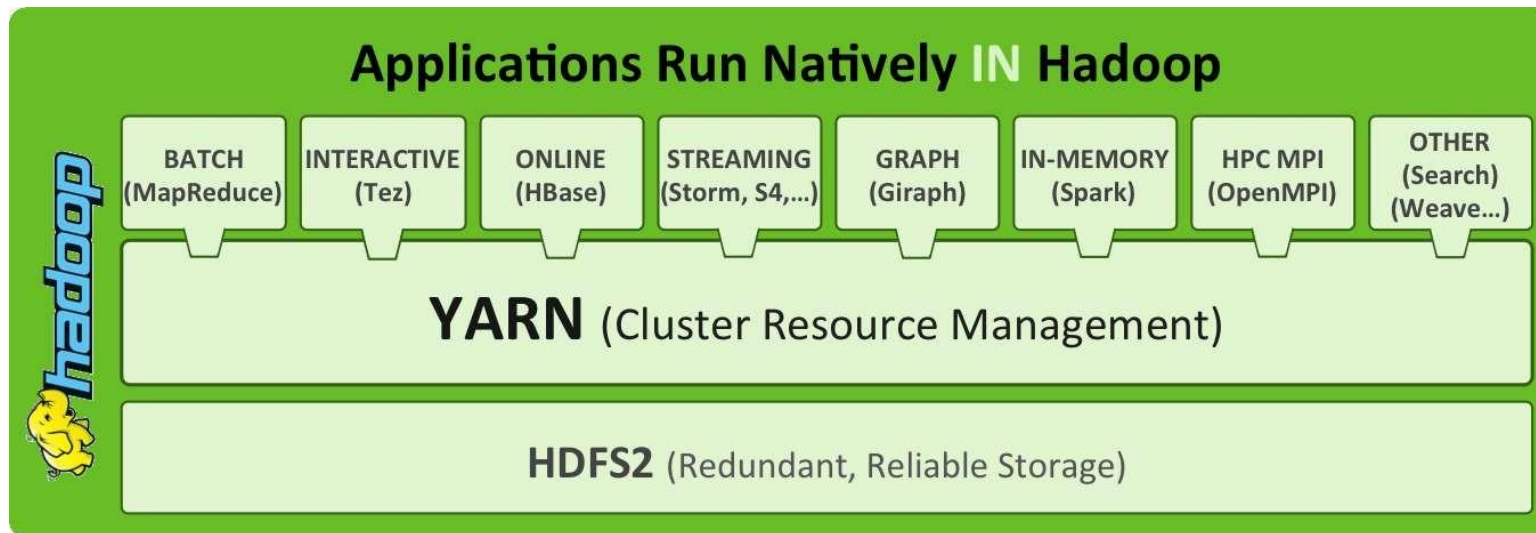
Cloudera Hadoop



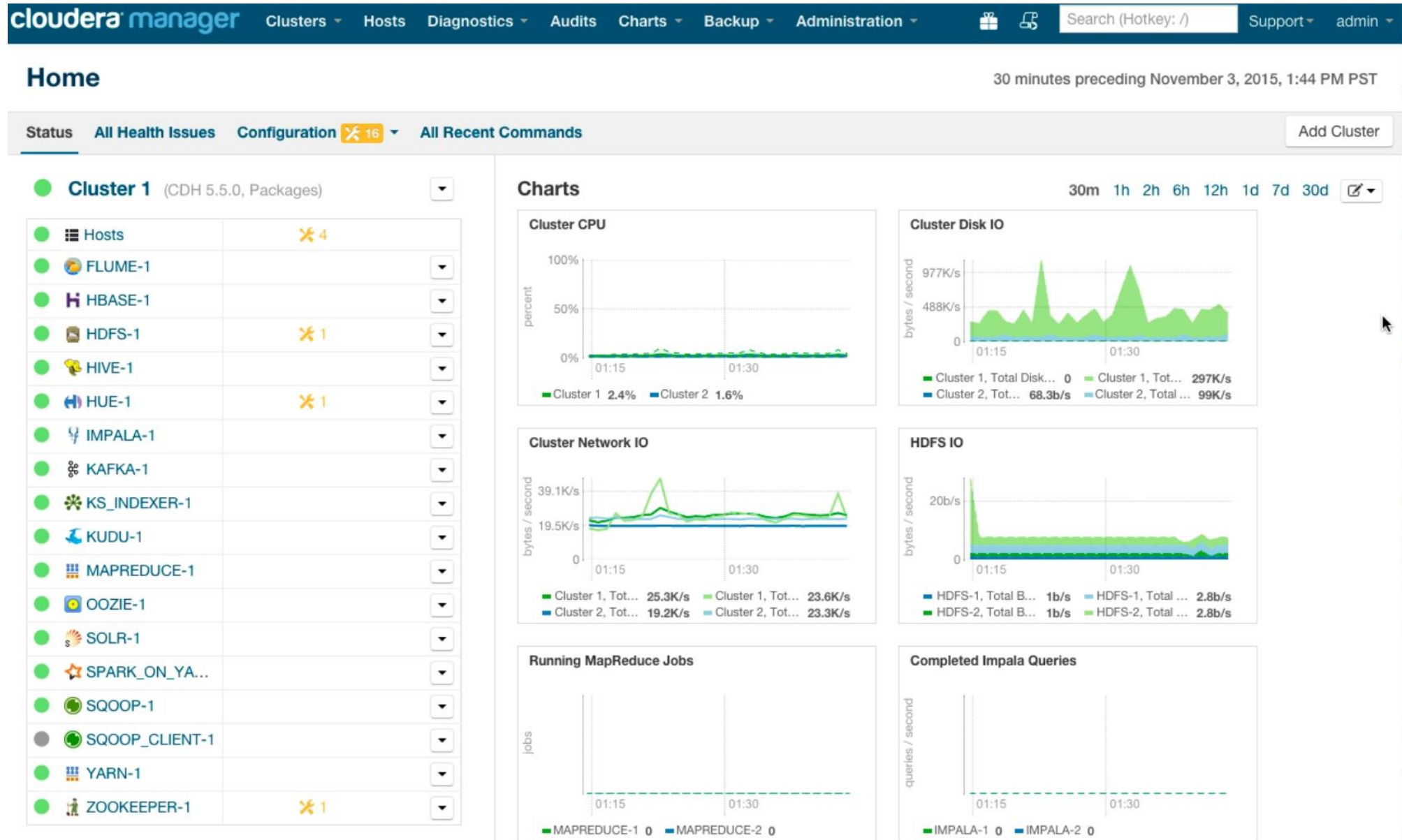
Hadoop v2 (YARN)



YARN (Hadoop OS)



Cloudera Manager



ทำความเข้าใจ Big Data และ Analytic

รู้จัก Big Data จาก HDC

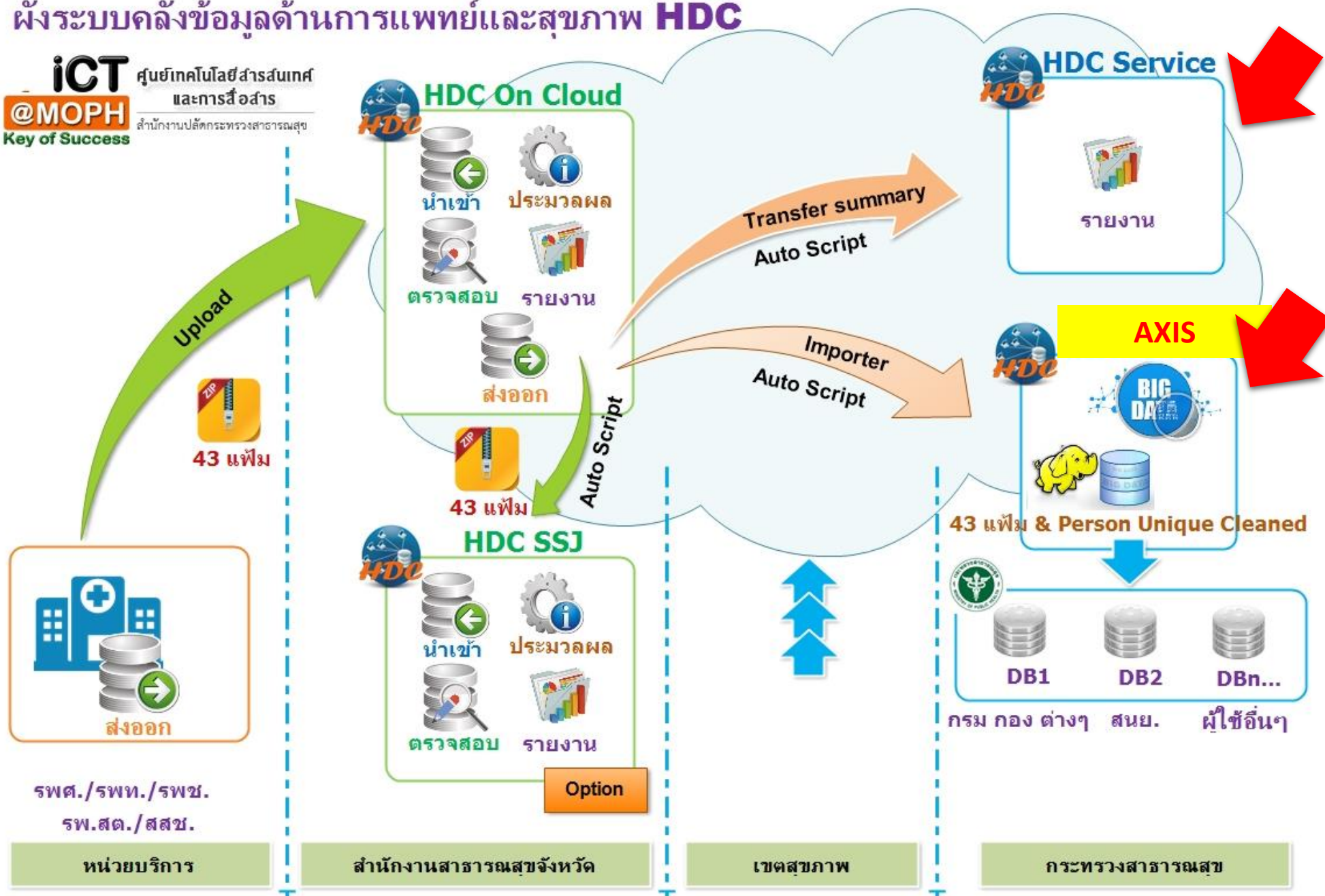
วิธีการทำให้ถึงจุดหมาย Big Data : Approach

Big Data management and Analytic



ผังระบบคลังข้อมูลด้านการแพทย์และสุขภาพ HDC

ICT ศูนย์เทคโนโลยีสารสนเทศ
และการสื่อสาร
@MOPH สำนักงานปลัดกระทรวงสาธารณสุข
Key of Success



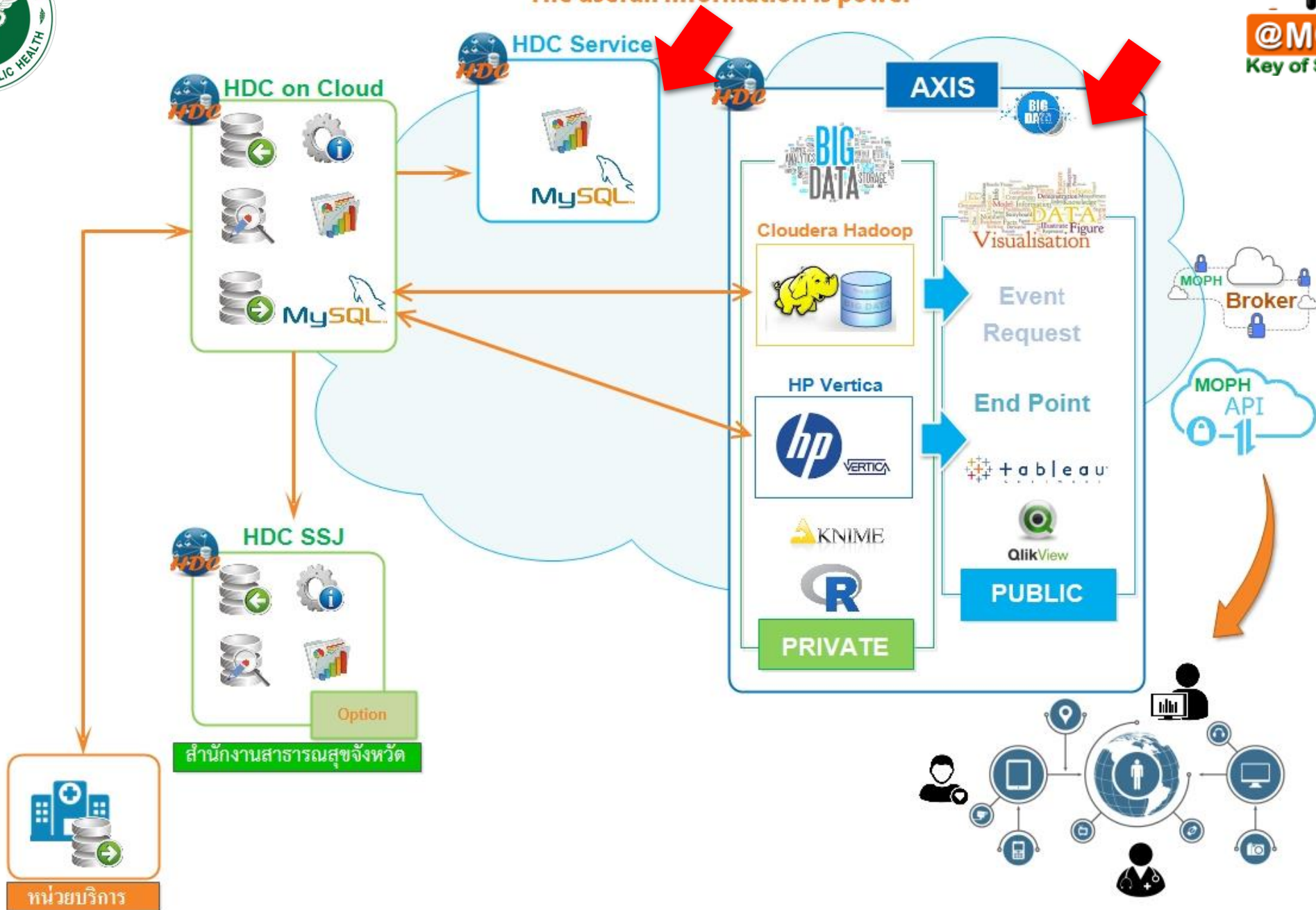


Health Data Center in Health 4.0

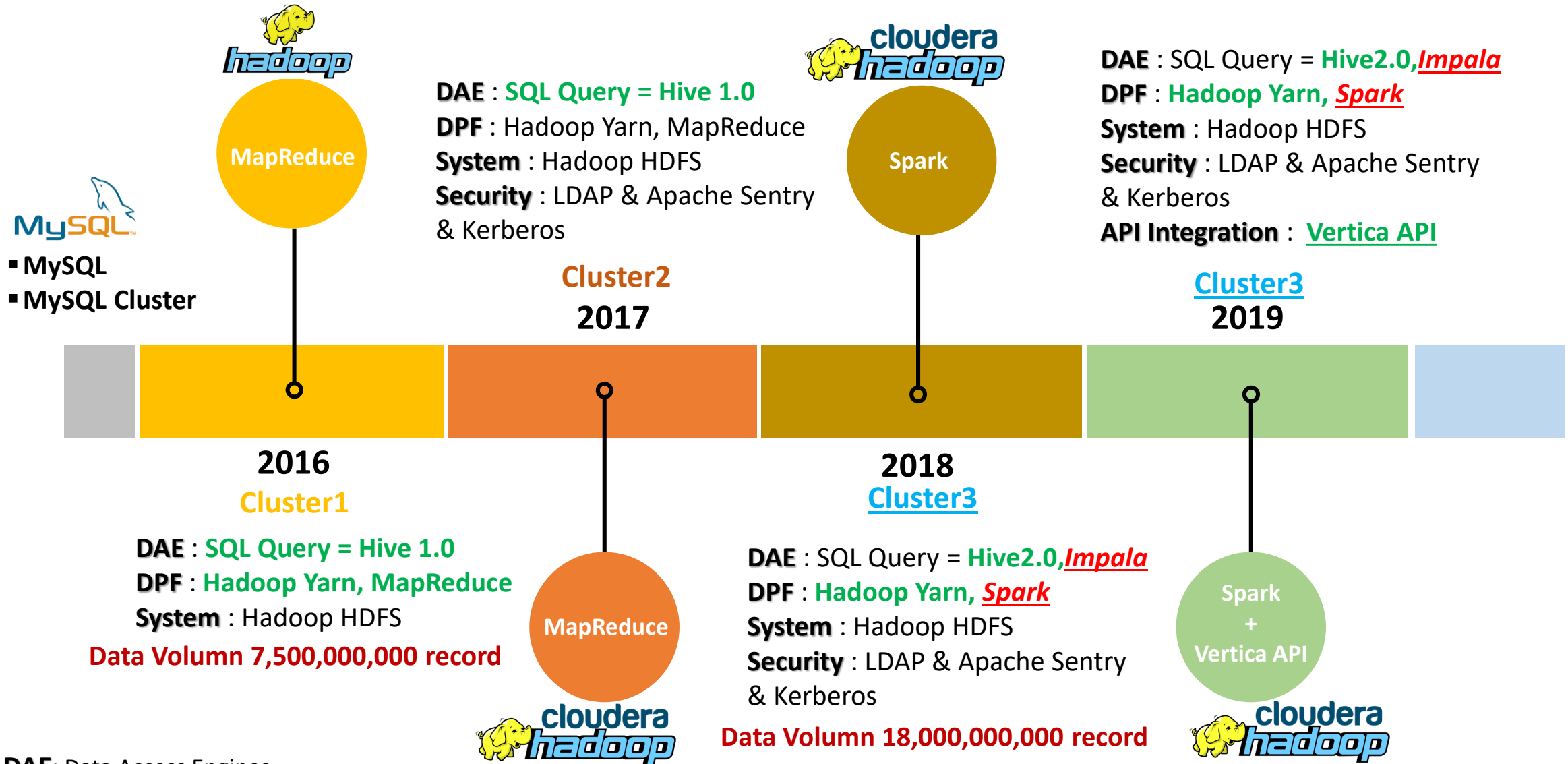
"The usefull information is power"

iCT
@MOPH
Key of Success

ศูนย์เทคโนโลยีสารสนเทศ
และการสื่อสาร
สำนักงานปลัดกระทรวงสาธารณสุข



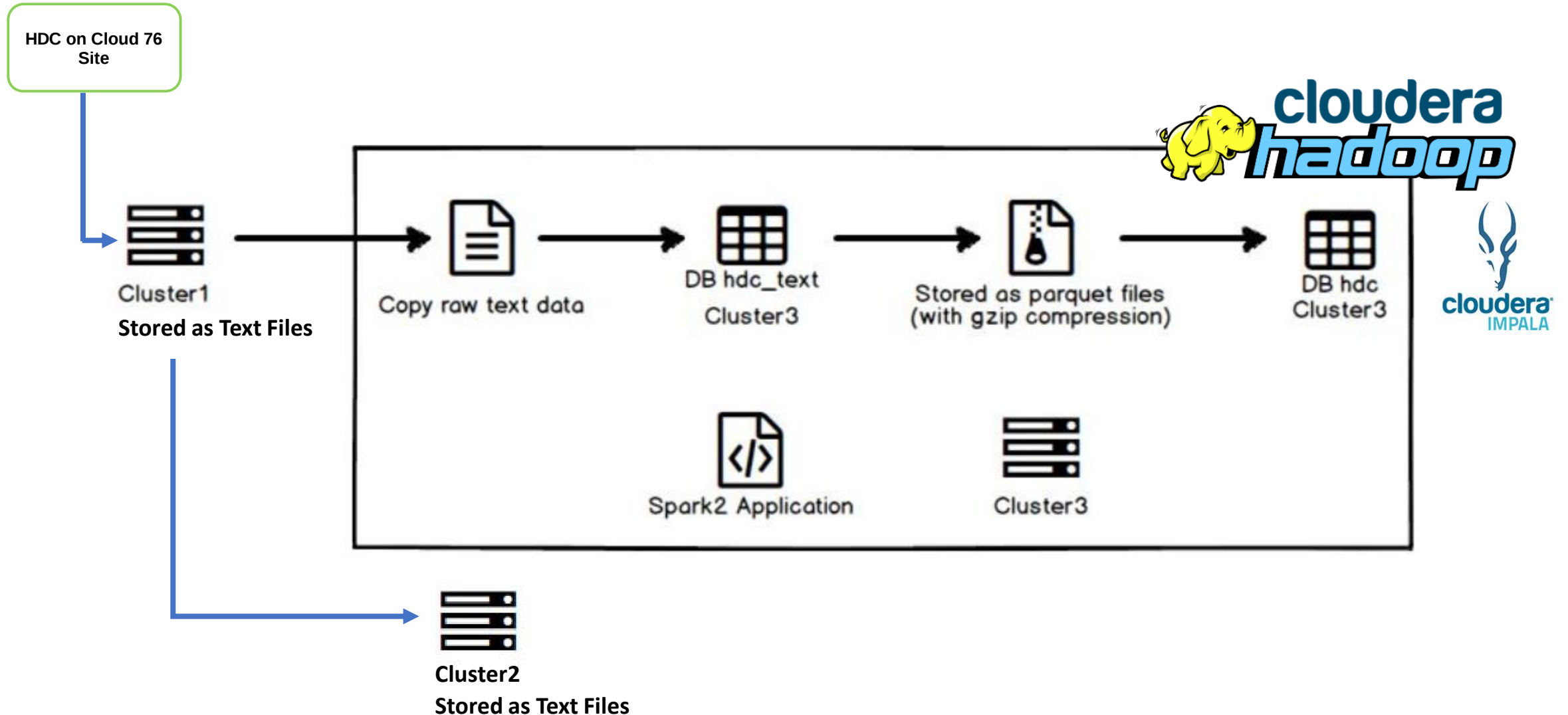
HDC Big Data Time line



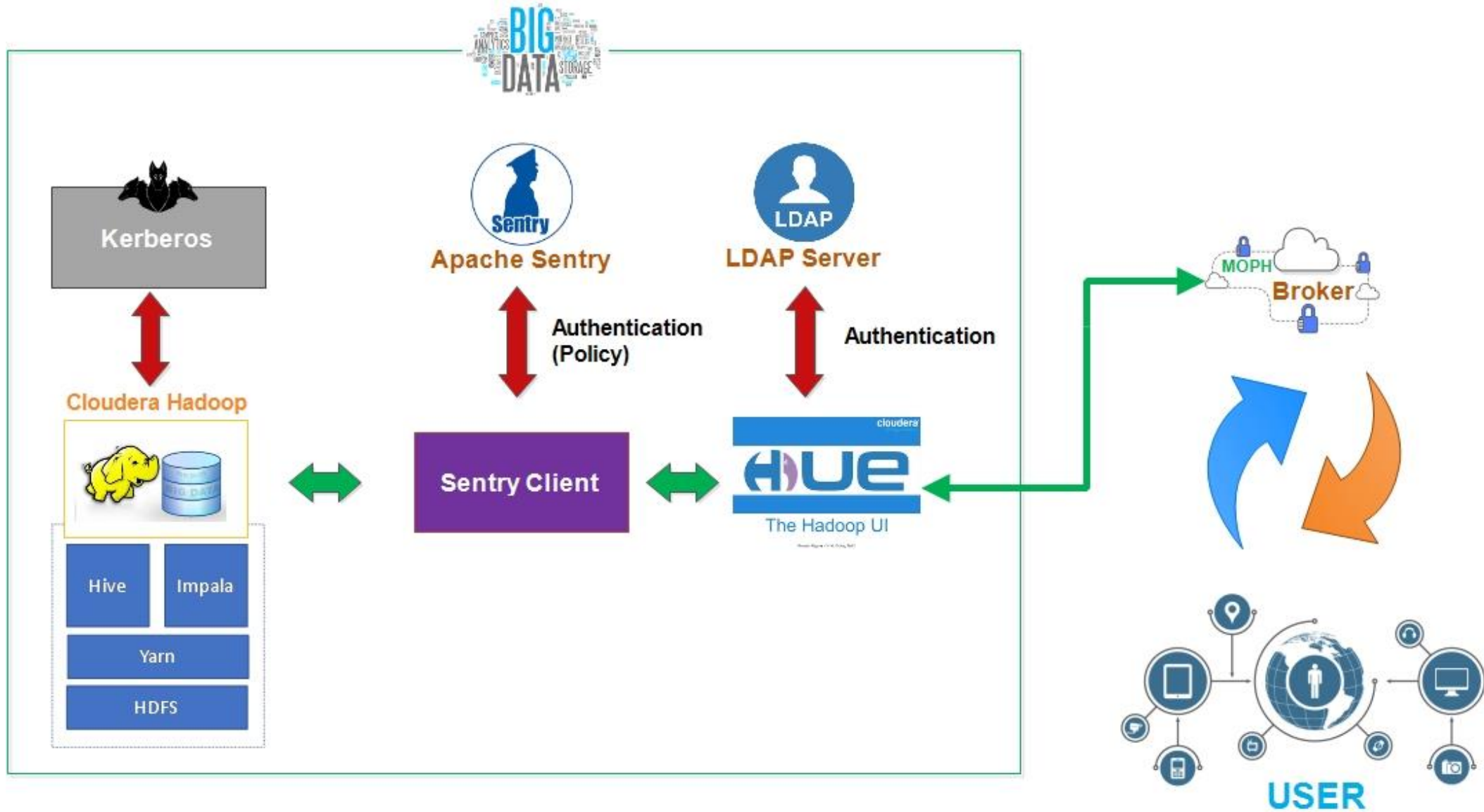
DAE: Data Access Engines

DPF: Distributed Processing Framework

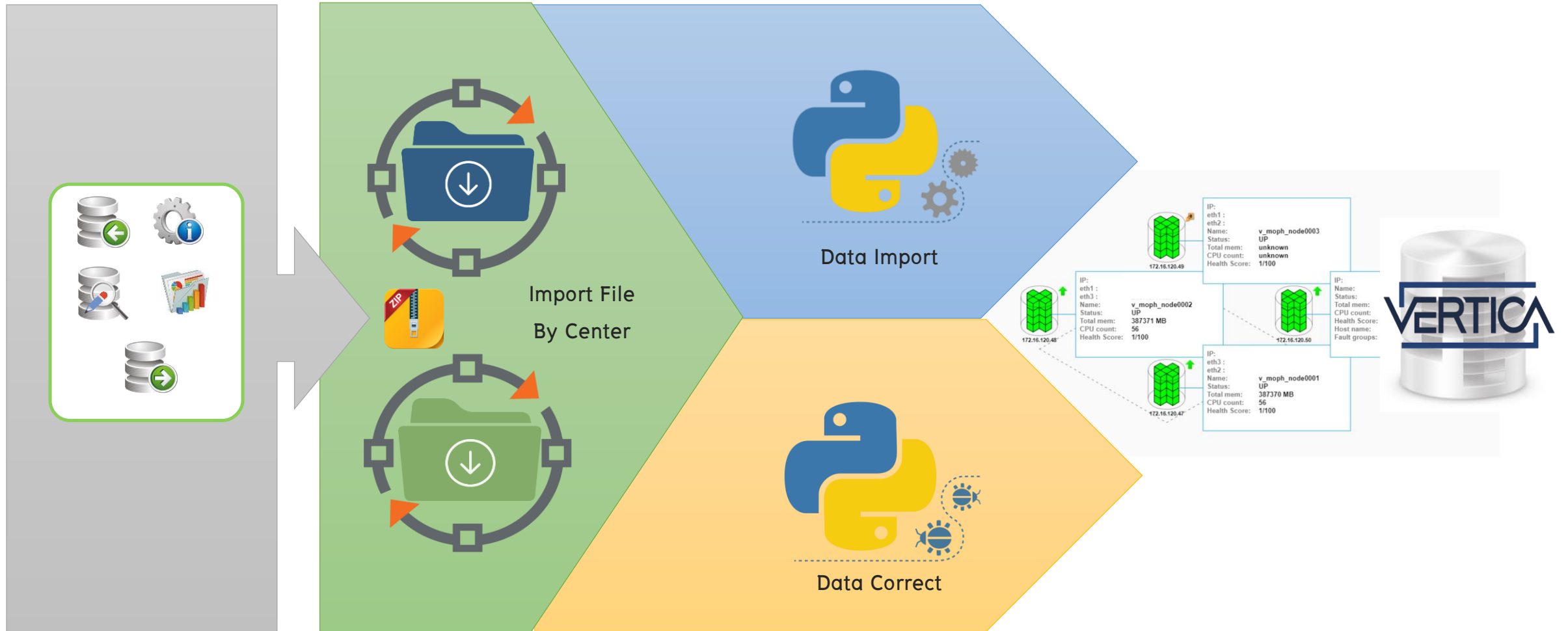
Hadoop Data Import and Stored



Hadoop Security Process

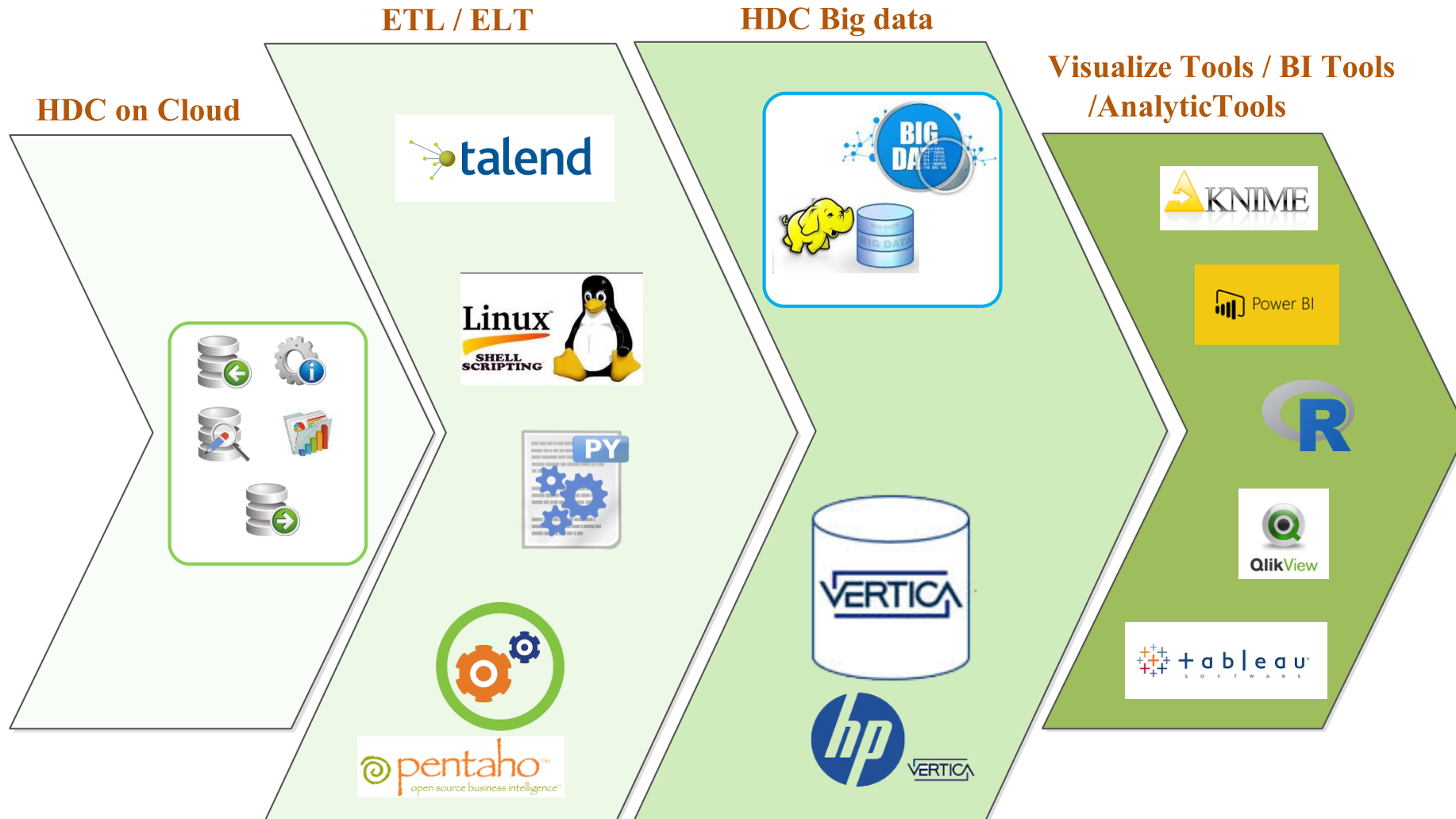


Vertica Data Import and Stored



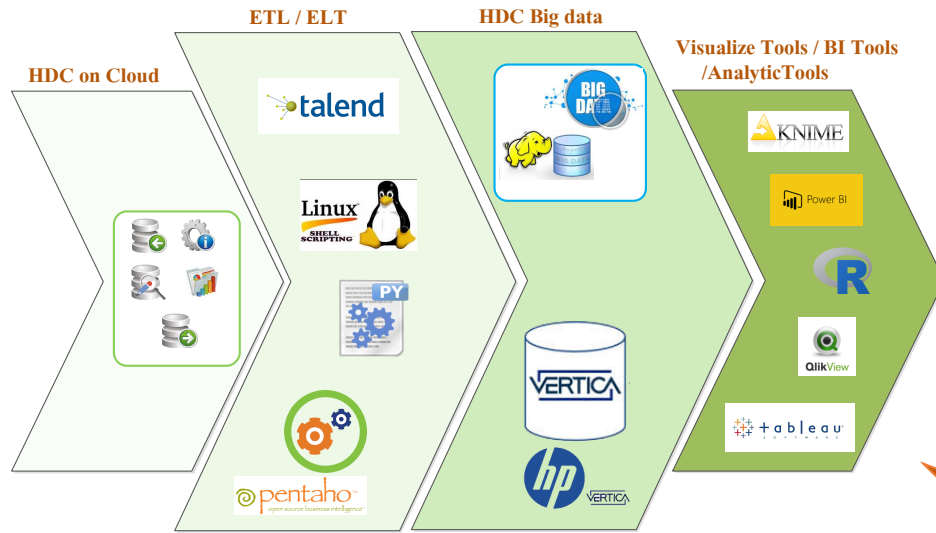


HDC Big Data





HDC Big Data Service



HDC Big Data



Data Mart

Place

Age

Sex

Time

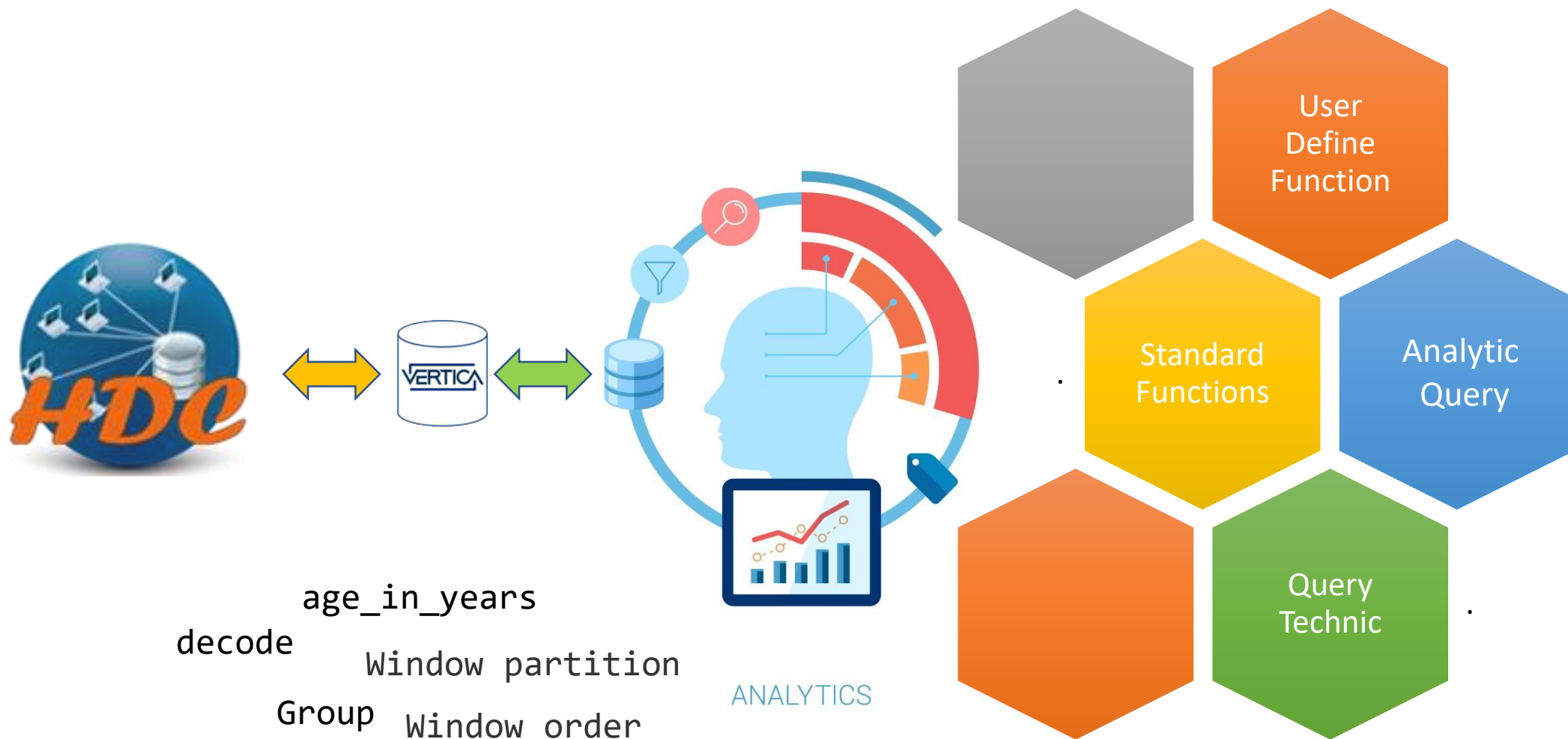
Diseases

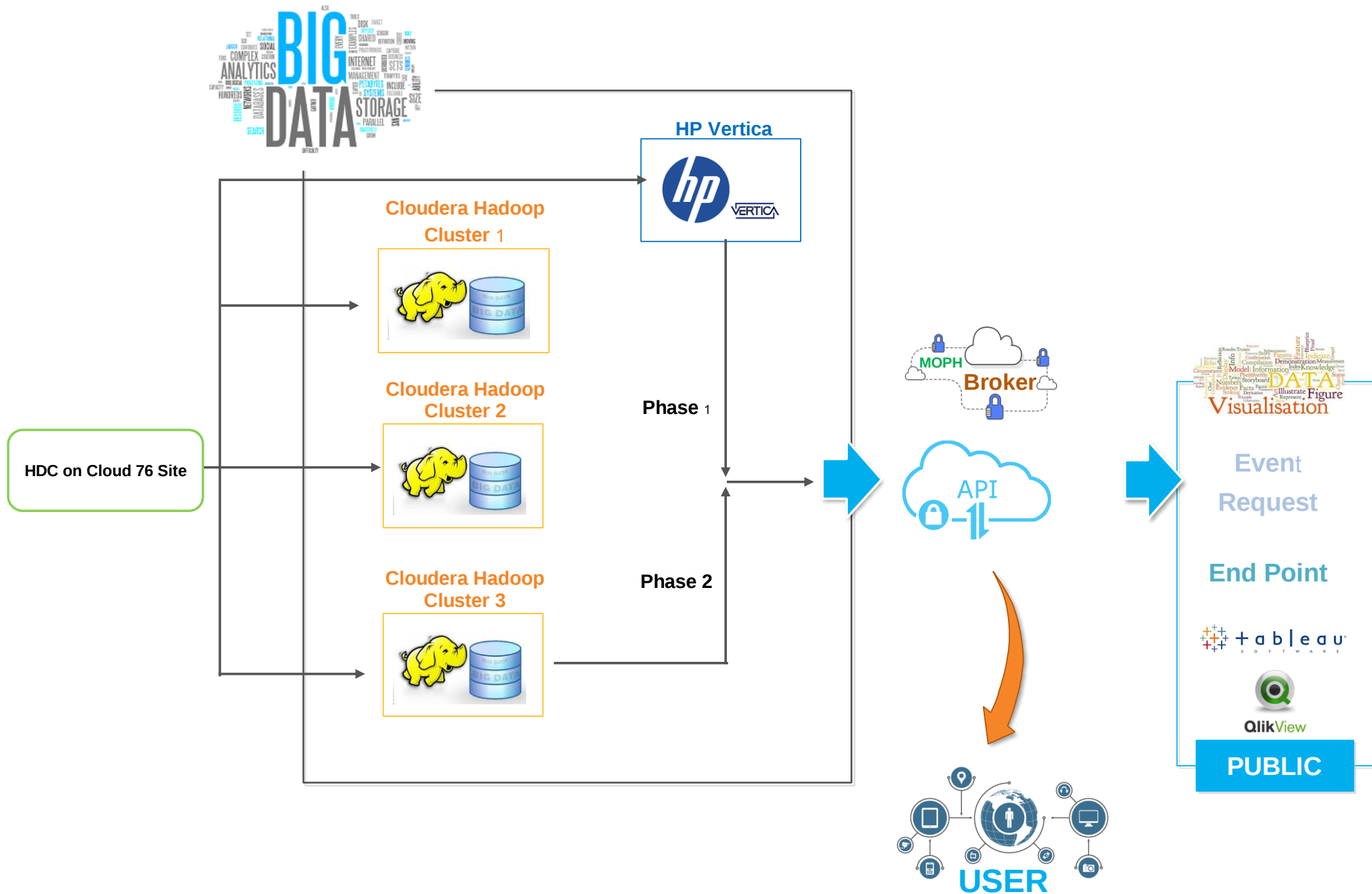
Death

Risk/Spatial PP

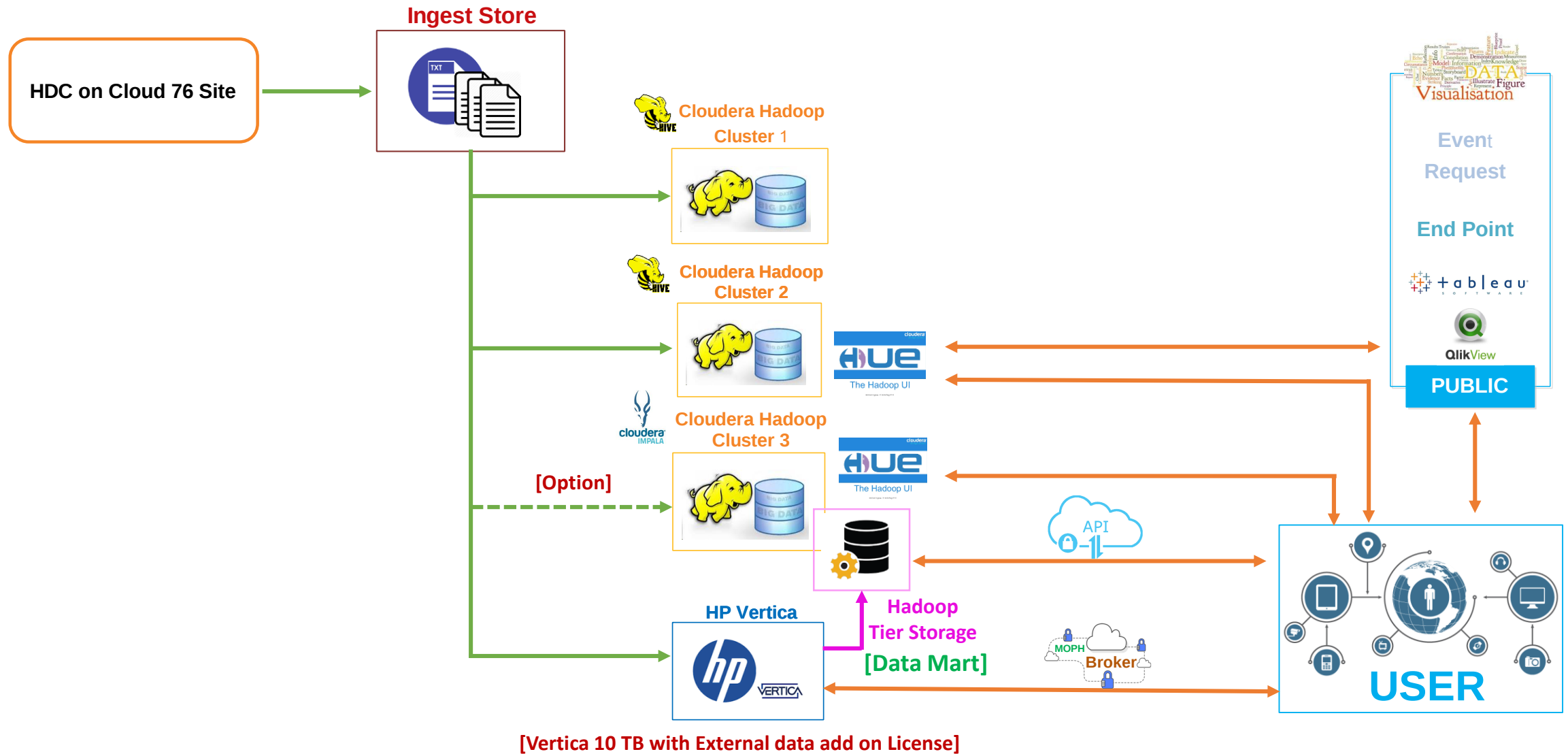


HDC ANALYTICS





HDC Big data architecture : Managing Data Lifecycle



ทำความเข้าใจ Big Data และ Analytic

รู้จัก Big Data จาก HDC

วิธีการทำให้ถึงจุดหมาย Big Data : Approach

Big Data management and Analytic

ต้องคำนึงถึงอะไรบ้าง?

ก่อนจะมี Big data

Mind set

- Big data/Data lake/ Data warehouse/Data Mart
- Query/Analytic

Technic/Technology

- Partition / Docker
- Explain Query
- Front end / Connector

Skill

- Data dictionary/Data Structure/ Data Definition
- Relation

Security

- SSL
- Tunnel

ต้องทำอะไรบ้าง ?

เมื่อจะมี **Big data**

สร้างระบบ

- ออกแบบ
- ติดตั้งและตั้งค่าระบบ

บริหารจัดการระบบ

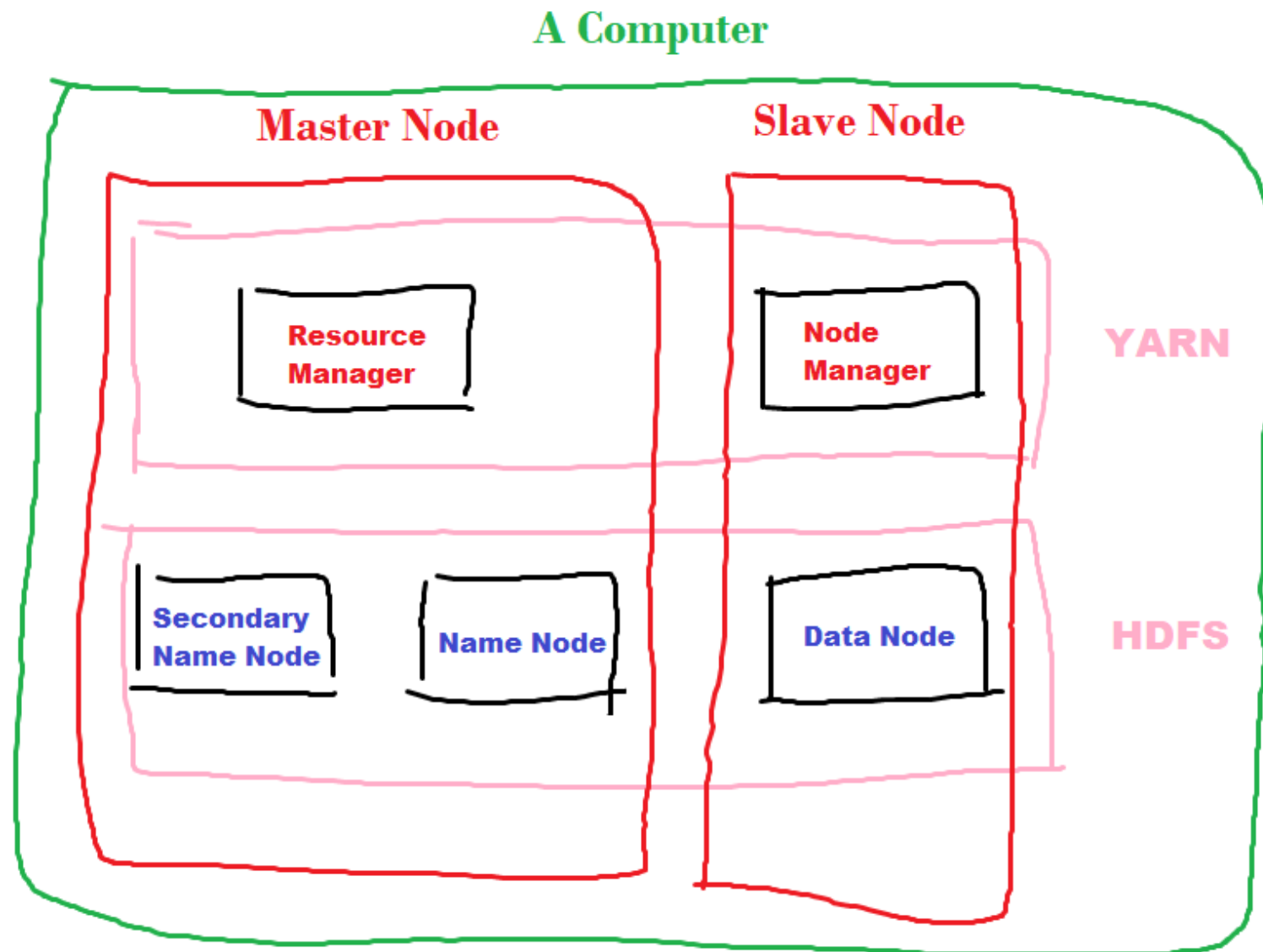
- Monitor / Maintenance
- Help desk

ใช้งานระบบ

- Preparation (ETL) [Cleansing, Mart]
- Query / Analytic
- Visualize

Hadoop Single Node Cluster

- Docker



HP vertica
Single node

- Docker



Vertica Cluster

