

CLOUDERA

EDGE TO AI VISION AND JOURNEY

Johannes (Joep) Muselaers | Solutions Engineer Nordics

COMBINED STRENGTH

Where we coming from

cloudera

**DATA
ANALYTICS**



**DATA
IN MOTION**

**DATA
AT REST**

NEW COMPANY

Cloudera at a glance

20000+ 30000+ 30000+

Customers across all
markets

[Meet our data heroes](#)

Solution and service
partners

[Find a partner](#)

Employees doing business
in 28 countries

[Join us](#)

LEADING IN TOP INDUSTRIES

8/10

TOP
GLOBAL



BANKING

10/10

TOP
GLOBAL



TELCO

40+

GOVERNMENT
CUSTOMERS



PUBLIC

10/10

TOP
GLOBAL



AUTOMOTIVE

9/10

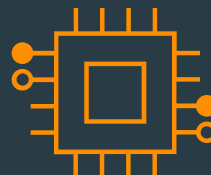
TOP
GLOBAL



PHARMA

8/10

TOP
GLOBAL



TECHNOLOGY

NEW COMPANY LOGO

CLOUDEERA

OPPORTUNITY



“AI, IoT and Augmented Reality are about to reinvent how industries utilize data—and could drive an era of productivity growth in our cities, farms, factories and hospitals.”

Morgan Stanley Research

REALITY

91%

ORGANIZATIONS STRUGGLE TO REACH DATA MATURITY

Gartner

WE **EMPOWER PEOPLE** TO TRANSFORM COMPLEX DATA INTO CLEAR AND ACTIONABLE INSIGHTS

GROW BUSINESS

Marketing automation
Personalized marketing
Customer experience
Churn prevention
Customer retention



CONNECT BUSINESS (IoT)

Predictive maintenance
Connected vehicles
Smart cities
Healthcare analytics
Industrial IoT



PROTECT BUSINESS

Regulatory compliance
Risk modeling & analysis
Financial crime prevention
Fraud detection
Cybersecurity



WE BELIEVE **DATA** CAN MAKE WHAT IS IMPOSSIBLE TODAY,
POSSIBLE TOMORROW

IMPOSSIBLE?
POSSIBLE.



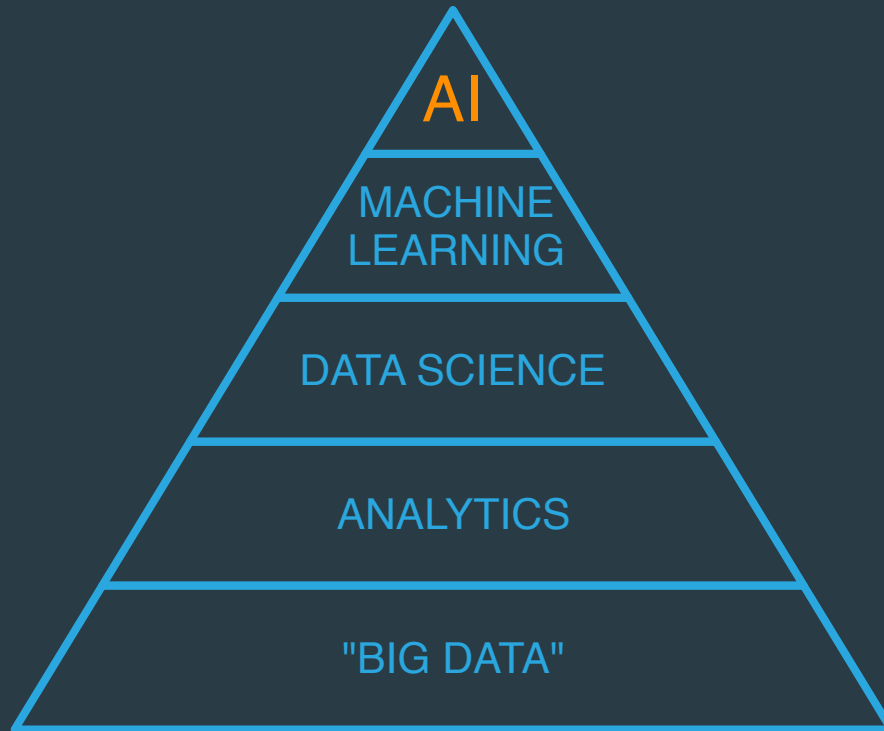
DIGITAL TRANSFORMATION TOP PRIORITY

- Enterprises become data driven



HIERARCHY OF NEEDS FOR THE DATA-DRIVEN ENTERPRISE

The “AI Ladder”



DATA SCIENCE EXPERTS BEST PRACTICE

- You can't do ML without doing data science
- You can't do data science without analytics
- You can't do analytics without Big Data
- They are all dependent on each other, and **must operate on the same platform**

CLOUDERA

THE PATH TO ENTERPRISE DATA CLOUD

**THE
ENTERPRISE
DATA
CLOUD
COMPANY**

WHERE WE ARE **TODAY**: DATA FROM ANYWHERE, FROM THE **EDGE TO AI**

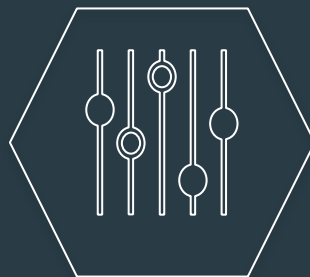
IoT, Ingest &
Streaming



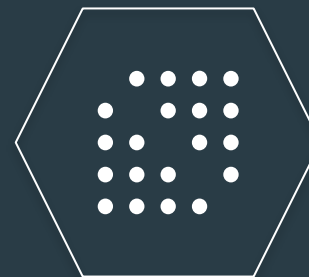
Data
Engineering



Data
Warehouse



Operational
Database



AI, ML &
Data Science



Edge | Private | Public | Multi | Hybrid

WHAT'S NEEDED FOR THE FUTURE DATA ARCHITECTURE?

AGILE INTERACTION MODEL

- Each data application runs on its own virtual private cluster (while sharing the data)
- Strong isolation of tenants and workloads (mitigate noisy neighbors)
- Independent upgrade cycles for components associated with each application

EFFICIENT/FLEXIBLE COMPUTE MODEL

- Run on virtualized & elastic compute infrastructure (primarily Kubernetes based)
- Scale compute separate from storage (given a solid network backbone architecture)

UNIFIED SECURITY, MANAGEMENT & METADATA

- Single pane of glass to manage 1000s of applications with a common data substrate
- Automation to drive dozens of application patterns and lifecycles
- Common security, governance, compliance, and metadata catalog

EASILY MOVE DATA & APPS BETWEEN ON PREMISES AND CLOUD

- Migration tools to easily move data and workloads
- Automatically synchronize metadata/security policies (and unified compliance tools)

THE ANSWER IS AN ENTERPRISE DATA CLOUD ARCHITECTURE

- Multi-function and multi-tenant analytics
- Hybrid and multi-cloud (migration/synchronization)
- Unified security and governance across workloads and cloud deployments on the same data
- Open platform

IOT, INGEST &
STREAMING

DATA
WAREHOUSING

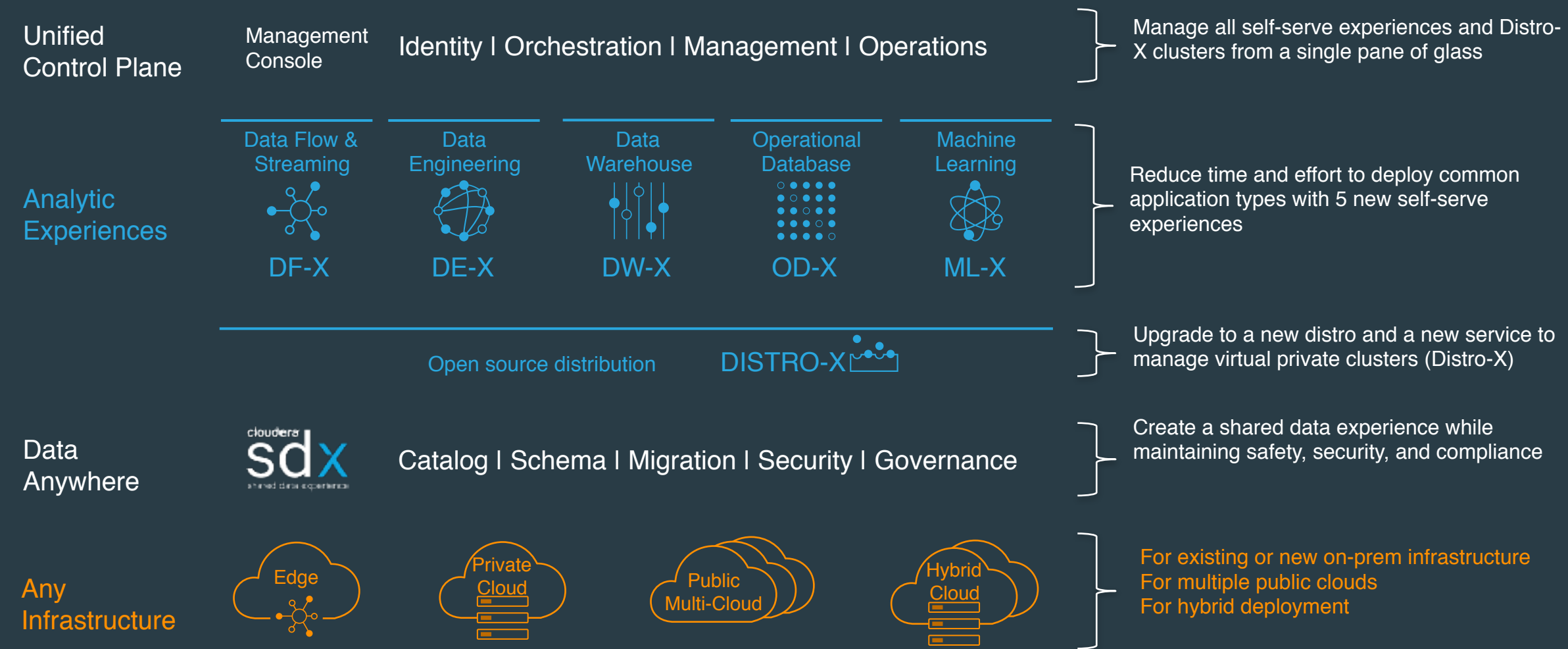
AI/ ML
DATA SCIENCE

SECURITY & GOVERNANCE

PUBLIC CLOUDS
compute & storage

DATACENTER
compute & storage

OUR ANSWER: THE **CLOUDERA DATA PLATFORM** (CDP)



THANK YOU

CLOUDERA

EARLY VIEW OF TECHNOLOGIES DEPRECATED IN CDP

- Apache Ambari → Replaced by Cloudera Manager
- Apache Sentry → Replaced by Apache Ranger
- Cloudera Director → Replaced by Cloudbreak
- Hive on Spark → Replaced by Hive on Tez

SIDE BY SIDE COMPARISON OF CDH AND HDP

Shared Tech	CDH 6.1	HDP 3.1
Hadoop	3.0.0	3.1.1
Spark	2.4	2.3
Hive	2.1.1	3.1.0
Oozie	5.0.0	4.3.1
HBase	2.1.1	2.0.2
Accumulo	1.9.2	1.7.0
Kafka	2.0.0	2.0.0
Sqoop	1.4.7	1.4.7
Solr	7.4	7.0
CDSW	1.5	1.5
CDF	3.2	3.2
Phoenix	5.0.0 (soon)	5.0.0

Non-Shared Tech	CDH Only	HDP Only	CDP
Governance	Navigator	Atlas / DSS	<i>Merge</i>
SQL	Impala	Hive LLAP	<i>Keep both</i>
File Format	Parquet	ORC	<i>Keep both</i>
Updatable Storage	Kudu	Hive ACID, Druid	<i>Keep both</i>
Query Interface	Hue	DAS Lite	<i>Merge</i>
Workload Mgmt	WXM	DAS	<i>Merge</i>
Backup	BDR	DLM	<i>Merge</i>