

cloudera

MACHINE LEARNING [Open Source] IN THE ERA OF DIGITAL TRANSFORMATION

Eynar Espinoza Pérez

Tech Lead for the Spanish Speaking Americas

Resumen



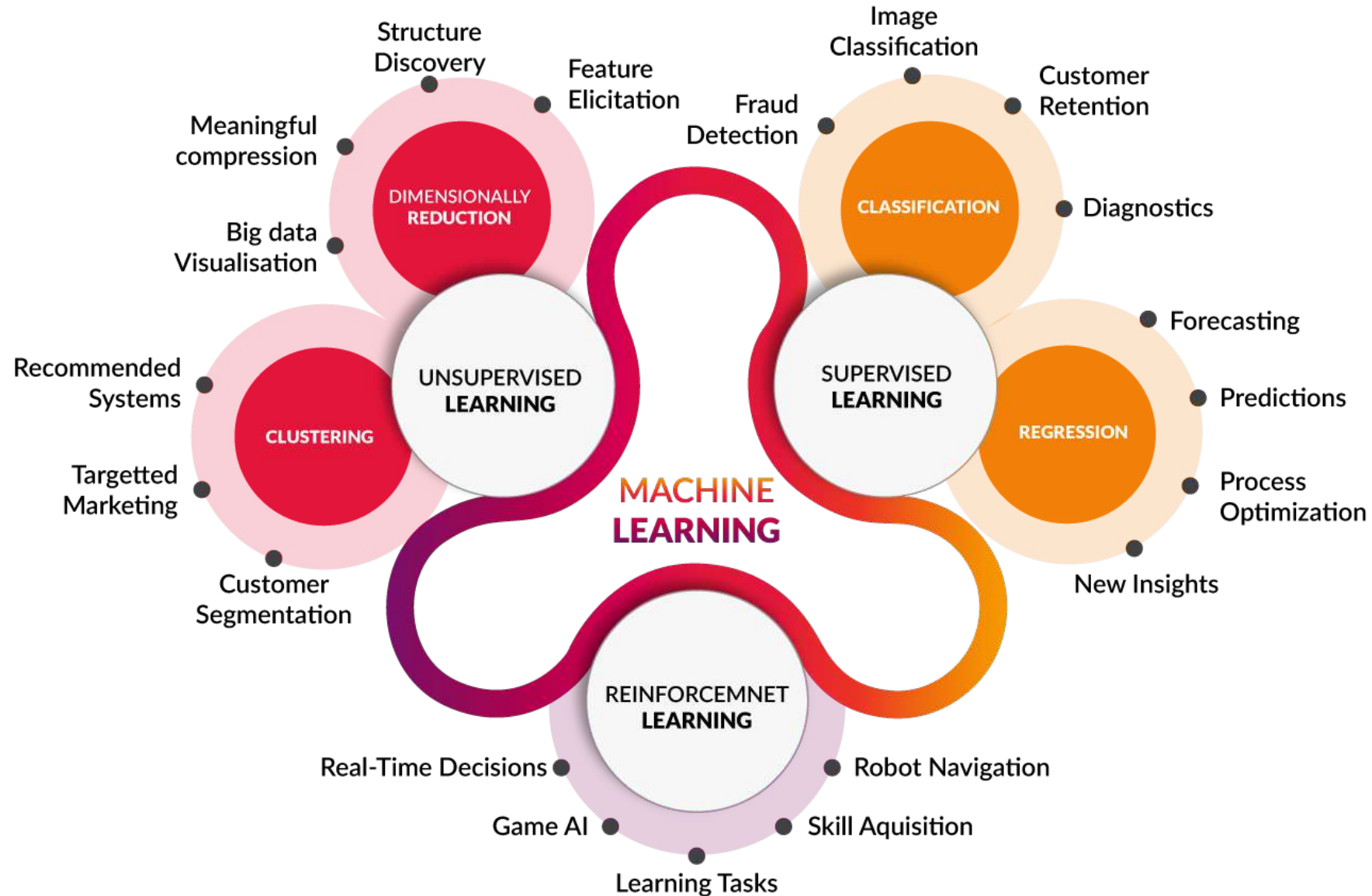
... a quick definition

MACHINE LEARNING

*“... machine learning **is an application** of artificial intelligence that provides systems with the ability to **automatically learn** and improve from experience, **without** being explicitly programmed ..”*

Arthur Samuel

MACHINE LEARNING APPROACH



... a quick overview

DIGITAL TRANSFORMATION

... *Digital Transformation: practices to success*

Understanding and Leadership

... more than a Strategy

Became Customer Centric

... a real attitude

From Data to Information

... in real time

Stimulate Predictive Sale/Offers

... let the power to the models

Process / Decision Automation

... from think together to collective thinking

Innovate in new business models

... use the knowledge acquired

New Leadership Model aligned with DT

... change management

Build new functions and promote new skills

... critical thinking, technological creativity

... a big ¿?

BIG DATA

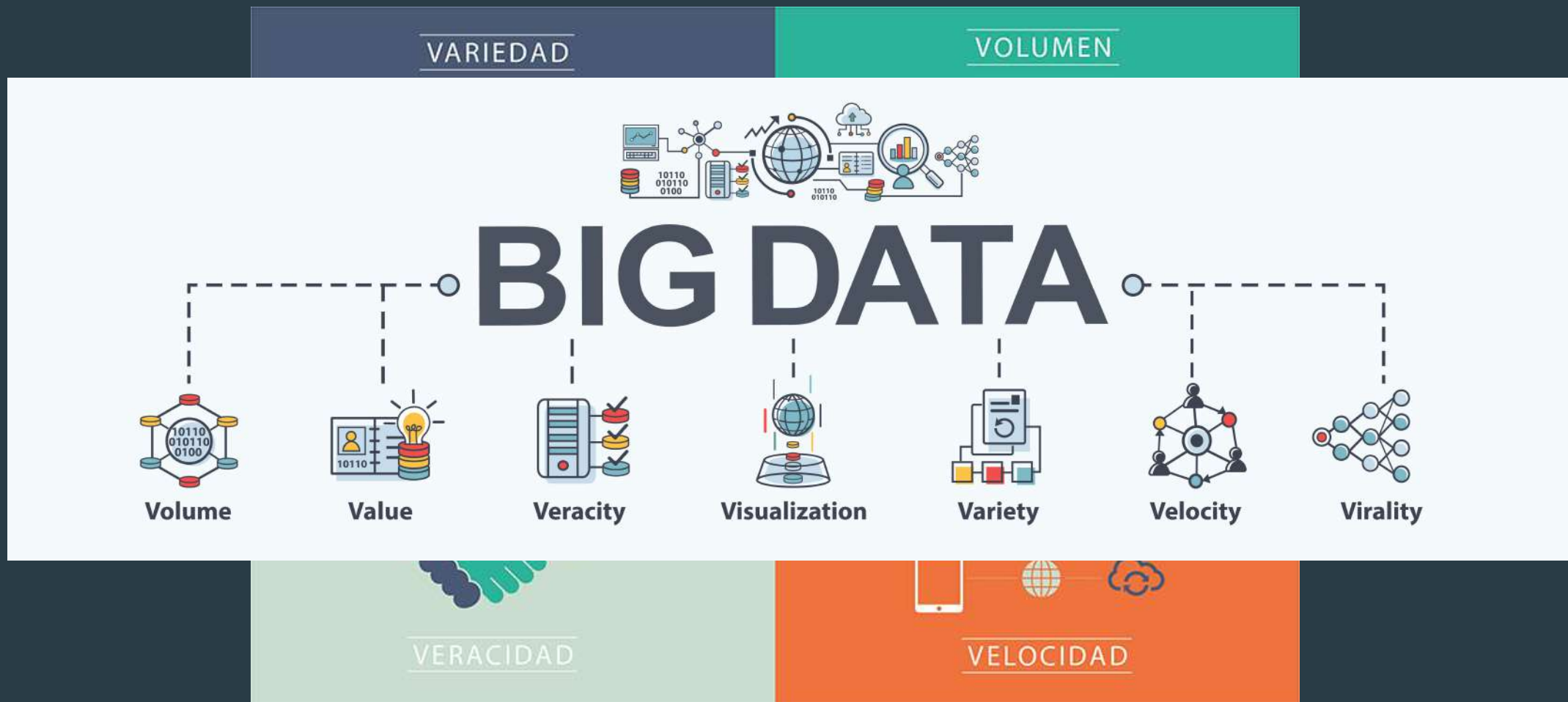
... a big ¿?



... a big ¿?



... a big ¿?



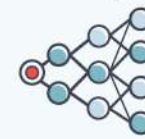
... a big ¿?



Volume



ty



Virality

... a big ¿?



BIG DATA



ML IN THE ENTERPRISE

MACHINE LEARNING IS A GROWTH ENGINE

Machine learning refers to **algorithms and methods** to extract useful patterns from data. When we say machine learning, we mean **broad, transformational data capabilities**.



DRIVE
customer insights



CONNECT
products &
services (IoT)



PROTECT
business

It's enabling **entirely new businesses**, not just modernizing existing systems.

THE CHALLENGE

Balance these needs

DATA SCIENCE

- Access to granular data
- Flexibility
 - Preferred open source tools
- Elastic provisioning
 - Compute
 - Storage
- Reproducible research
- Path to production



VS.

DATA MANAGEMENT

- Security
- Governance
- Standards
- Low maintenance
- Low cost
- Self-service access

THE TYPICAL SOLUTION

“If I can’t use my favorite tools, I’ll...”

- Copy data to my laptop
- Copy data to a data science appliance
- Copy data to a cloud service

Why this is a problem:

- Complicates security
- Breaks data governance
- Adds latency to process
- Makes collaboration more difficult
- Complicates model management and deployment
- Creates infrastructure silos

WHY NOW?

Disruptive trends create opportunities for change

CLOUD

Access to massive, heterogeneous **compute** just a click away.

BIG DATA

Affordable to store and access any amount or type of **data**.

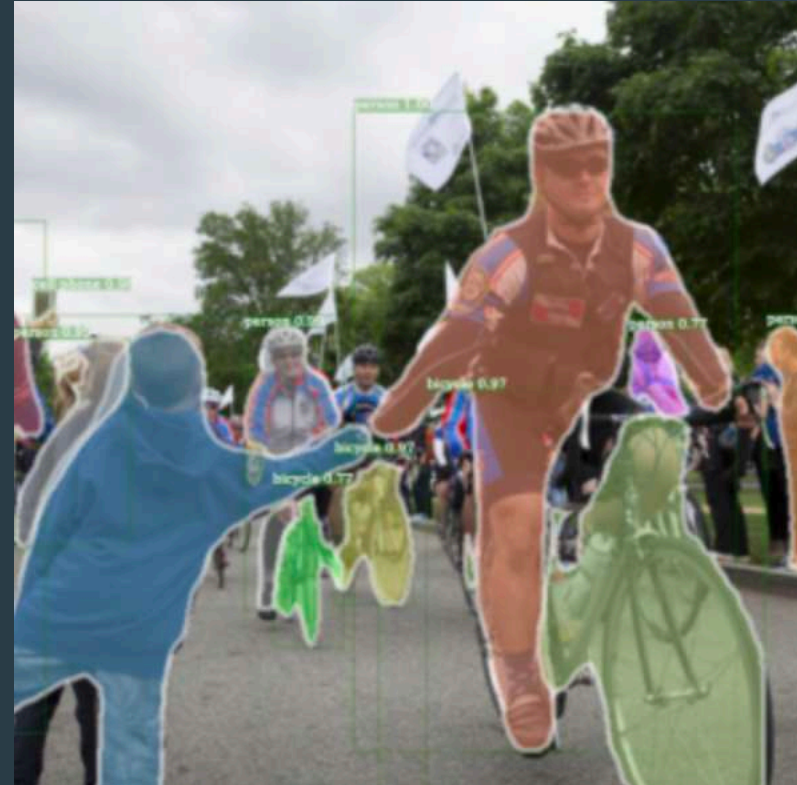
OPEN SOURCE

Machine learning **algorithms** are available for free, and frequently updated.

COMPUTER VISION

Detectrum (Facebook Research)

FAIRs research platform for object detection research, implementing popular algorithms like Mask R-CNN and RetinaNet

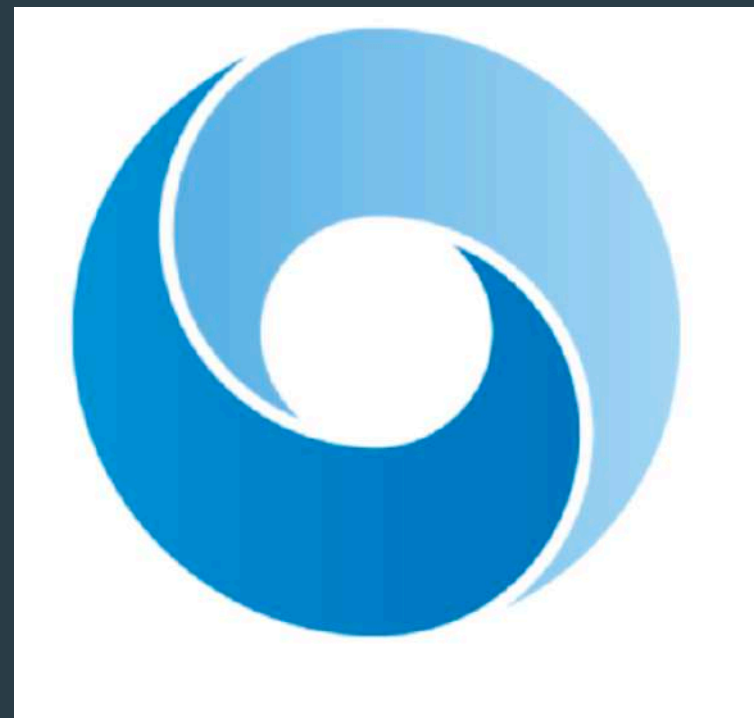


<https://github.com/facebookresearch/Detectron>

REINFORCEMENT LEARNING

TRFL

A library of useful building blocks for writing reinforcement learning (RL) agents in TensorFlow

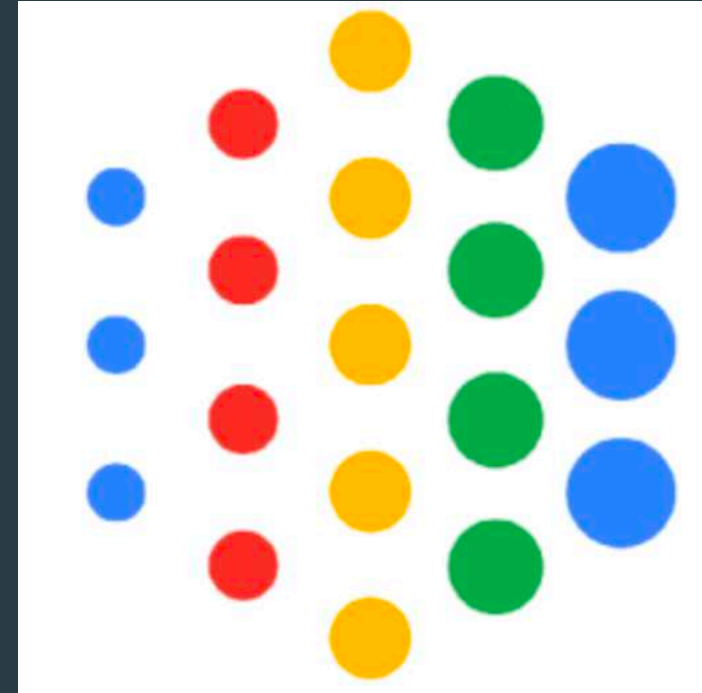


<https://github.com/deepmind/trfl>

NLP

BERT (Google research)

TensorFlow code and pre-trained models for BERT



<https://github.com/google-research/bert>

GAN

DeOldify

A Deep Learning based project for colorizing and restoring old images



<https://github.com/jantic/DeOldify>



BONUS

REINFORCEMENT LEARNING

MAMEToolkit

Arcade Game Reinforcement
Learning Python Library

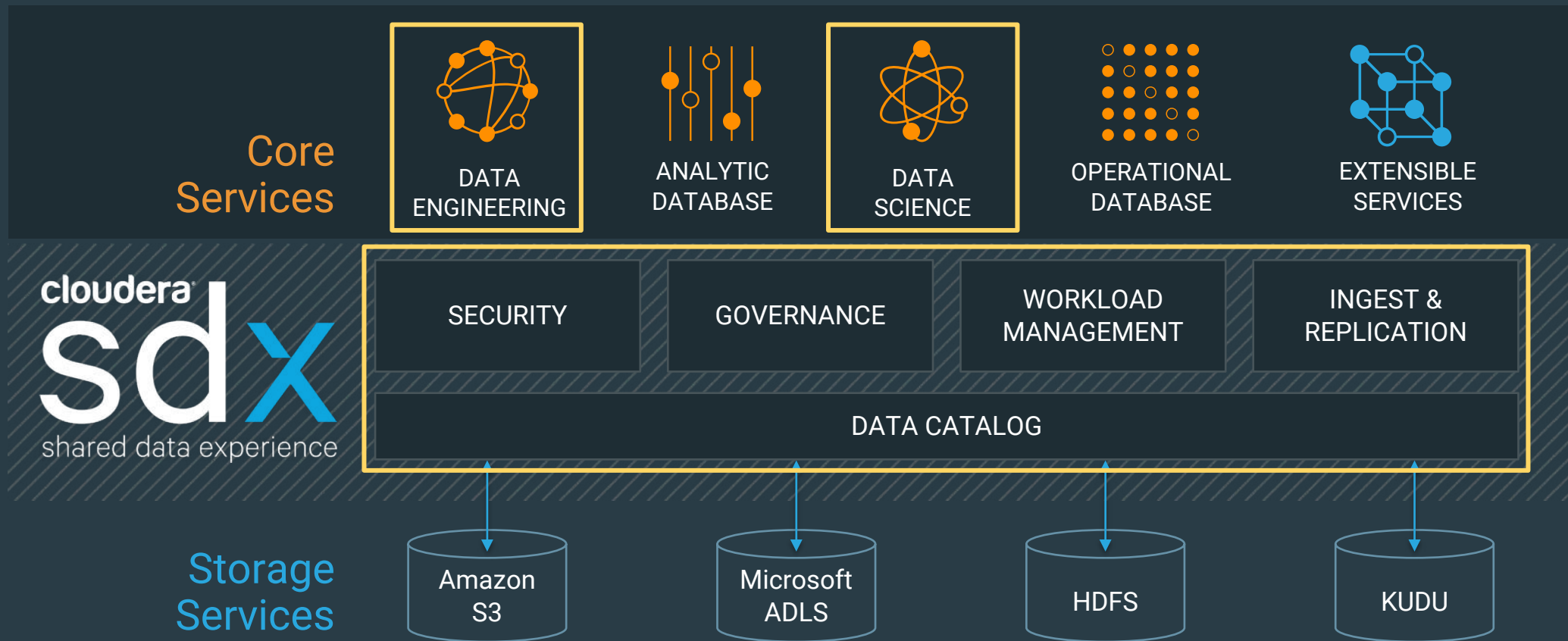


<https://github.com/M-J-Murray/MAMEToolkit>

PLATFORM

MACHINE LEARNING IS BUILT ON DATA MANAGEMENT

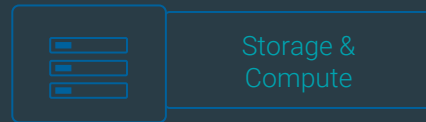
Integrated data, workflows, metadata, security, governance, ...



NEW CLUDERA DATA PLATFORM

CDP Management Console

SDX – security, governance, discovery, replication, and workload analysis



CDH & HDP



CDP Public Cloud



CDP Private Cloud

THE ENTERPRISE DATA CLOUD



COMPONENTS

Traditional Platform Consumption:

- Data Hub Clusters

New analytic experiences:

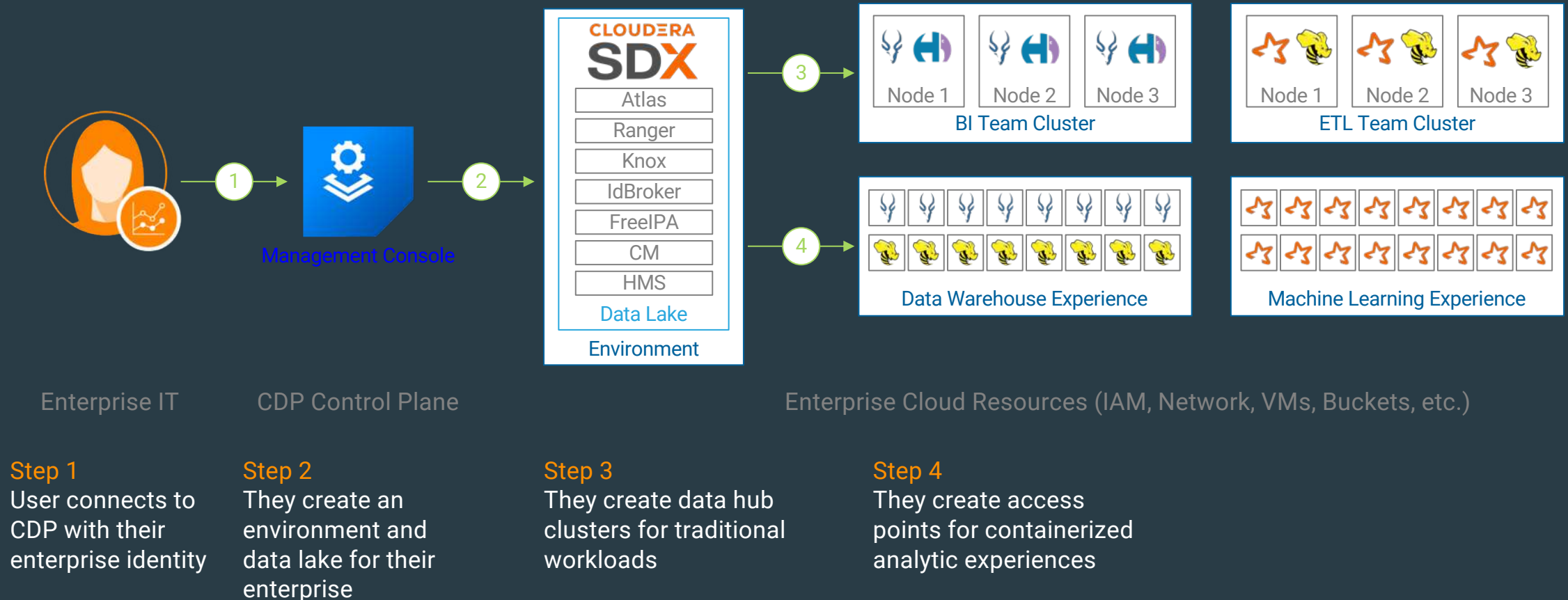
- Data Warehouse
- Machine Learning
- More to come

Control Plane services:

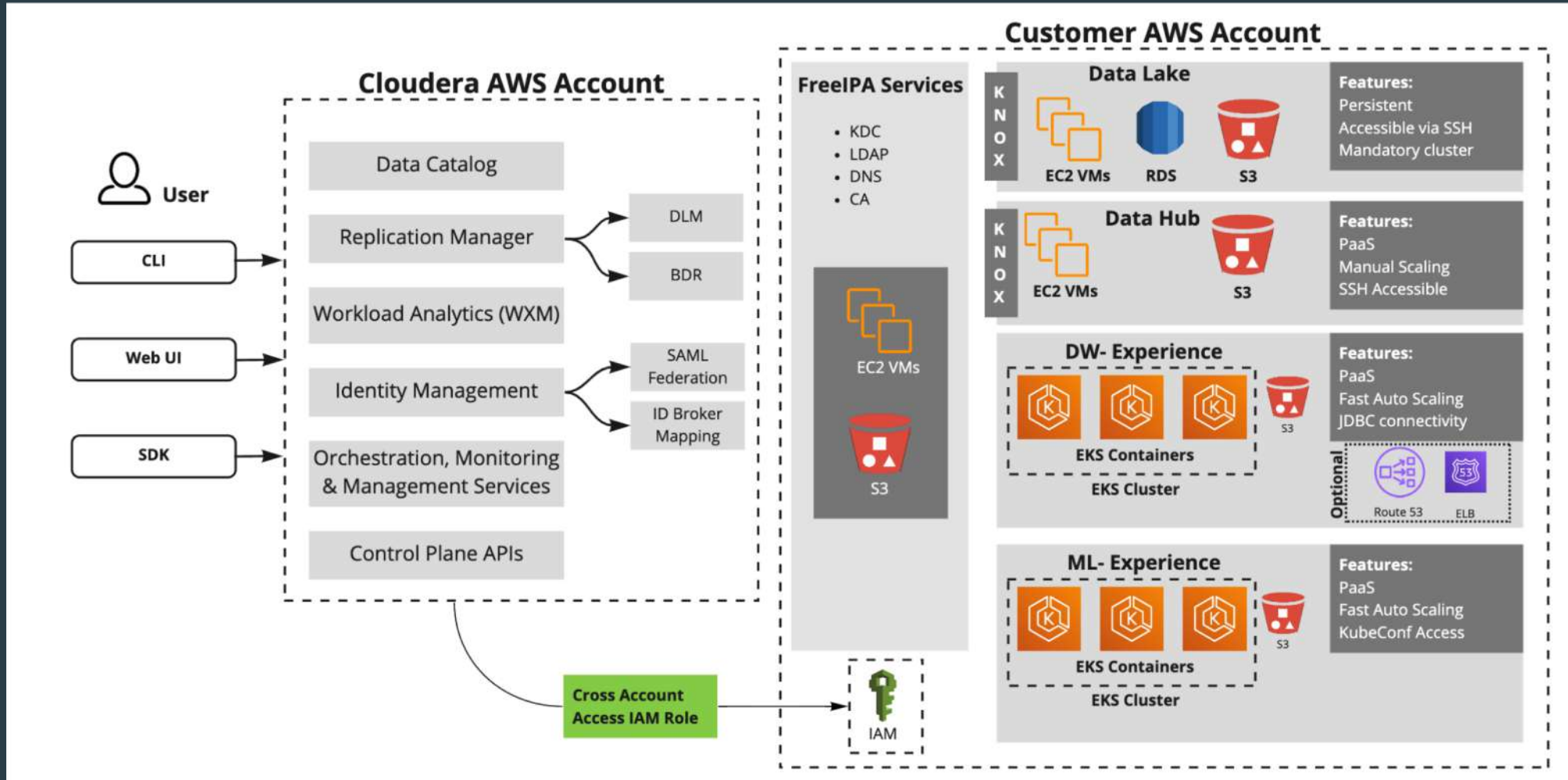
- Workload Manager
- Replication Manager
- Data Catalog
- Management Console

KEY CONCEPTS & COMPONENTS

Typical user flow



AWS high level architecture



THANK YOU

cloudera