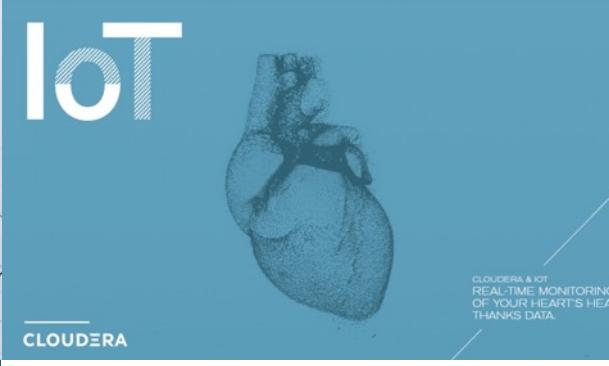




# ENTERPRISE DATA CLOUD

CLOUDERA



CLOUDERA

CLOUDERA

CLOUDERA & IoT  
REAL-TIME MONITORING  
OF YOUR HEART'S HEALTH.  
THANKS DATA.



# EDGE 2AI

CLOUDERA



CLOUDERA



CLOUDERA

The Era of Enterprise Data Cloud with

# CLOUDERA DATA PLATFORM

Open Source Lisbon 2020

Alex Campos | Solutions Engineer

[ac@cloudera.com](mailto:ac@cloudera.com)

---

# AGENDA

- Enterprise Data Cloud
- Cloudera Data Platform
- SDX
- Recap

# ENTERPRISE DATA CLOUD

# CLOUD DATA LAKES REQUIRE ...



Identities

How do we easily connect to **corporate** identity?



Schema

How do multiple personas **find** and **share** datasets across different compute engines?



Policy

Is there a central place to protect sensitive data at a **field** and **row** level?



Audits

Can I comply to the regulatory requirements with comprehensive **audits** and **lineage**?



Privacy, compliance & regulation



# ... AND ACTIVE MANAGEMENT



## Data Profiling & Stewardship

Can I detect **sensitive** data automatically, **tag** and **curate** for easy discovery and security controls?

State is required for all but the simplest use cases



## Replication

What happens to metadata and lineage, when I move data **across environments**?

- Multi-tenant
- Multi-workload



## Workload Management

How do I track workloads for **troubleshooting** and optimization with elastic computes?

- Sensitive data
- Regulated use case

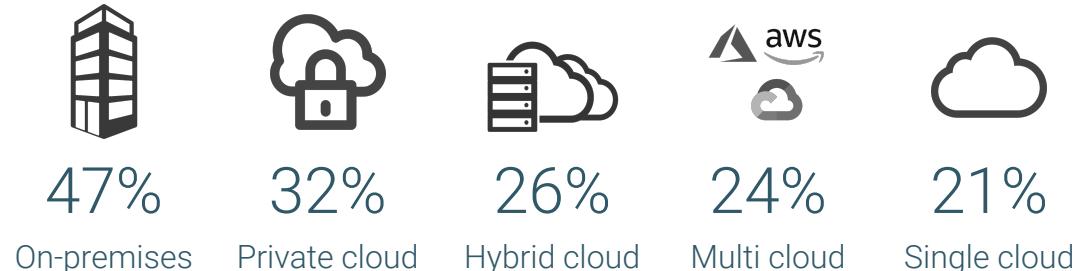


## Encryption

Is the data **protected** at rest and in-transit?

- Long Running Apps

DATA MANAGEMENT IS SPREAD ALL OVER



Gartner recently warned that “Data and analytics leaders must prepare for the complexities of multi cloud and intercloud deployments to avoid potential performance issues... unplanned cost overruns and ... difficulties with integration efforts.”

# CLOUDERA

# DATA

# PLATFORM

# CLOUDERA DATA PLATFORM

## Industry's First Enterprise Data Cloud

Any Cloud | All Data



Multiple public  
Hybrid  
Private  
Data center  
Edge

Data Lifecycle



Streaming  
Data engineering  
Data warehousing  
Machine learning & AI

Secure & Governed



Data & metadata  
Fine grained security  
Lineage and provenance  
Data & workload migration

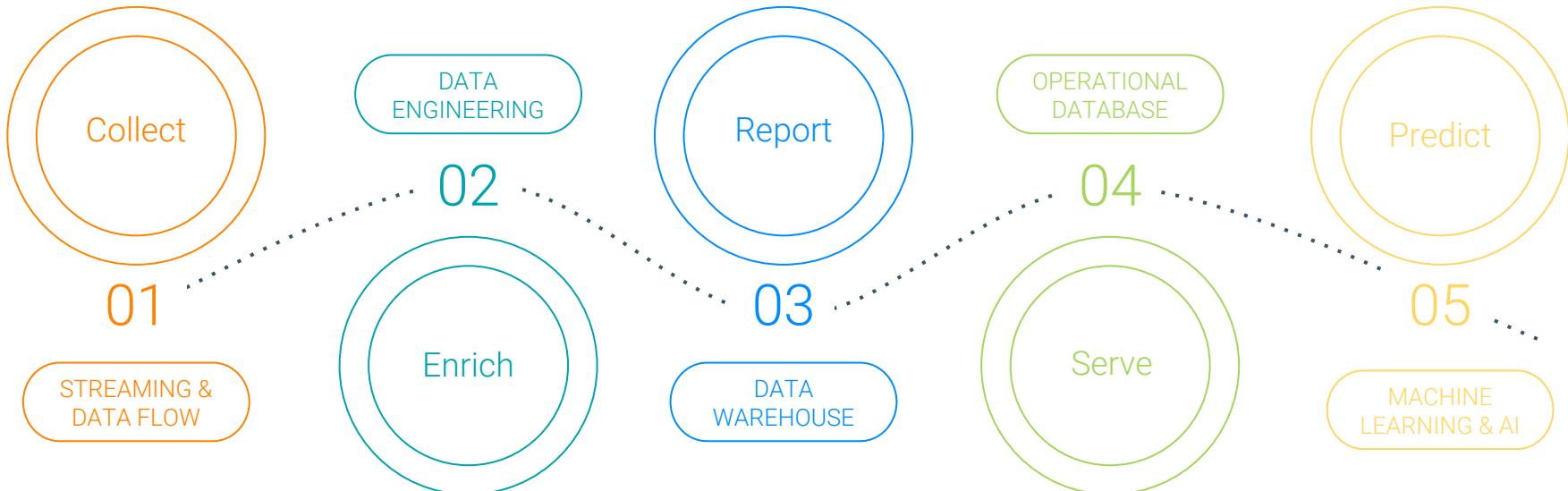
Open + Standards



100% Open source  
Open data formats  
Open storage & compute  
Open APIs

# CLOUDERA - THE ENTERPRISE DATA CLOUD COMPANY

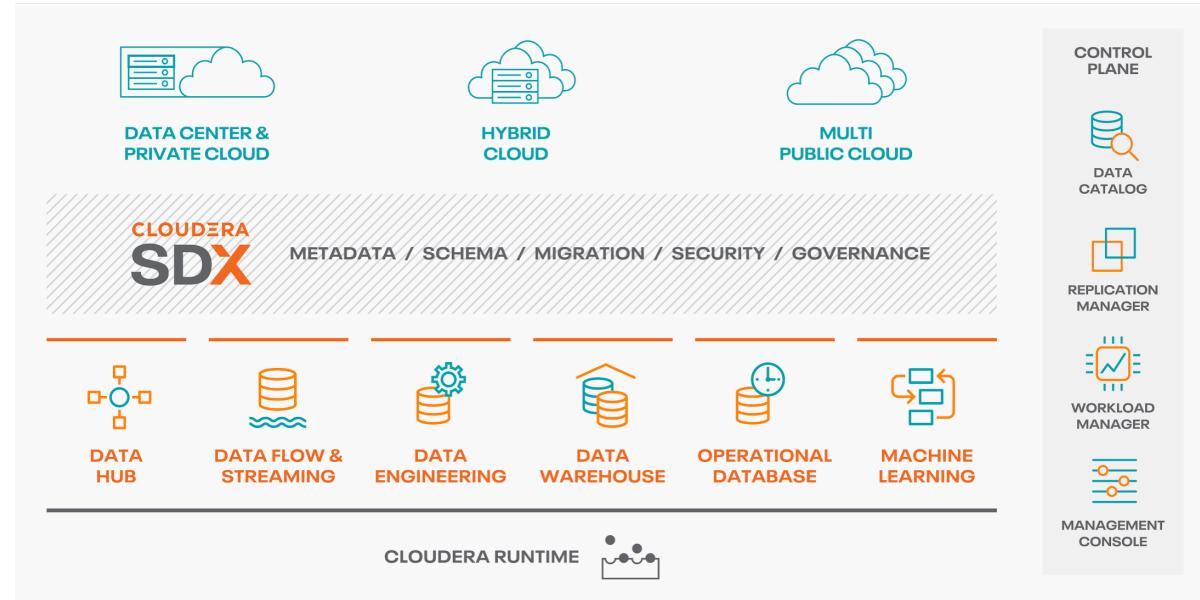
Manage and secure the data lifecycle in any cloud or datacenter



SECURITY | GOVERNANCE | LINEAGE | MANAGEMENT | AUTOMATION

# CLOUDERA DATA PLATFORM

- Control cloud costs with auto scale, suspend and resume
- Optimize workloads based on analytics and machine learning
- View data lineage across any cloud and transient clusters
- Use a single pane of glass across hybrid and multi-clouds
- Scale to petabytes of data and 1,000s of diverse users



# ONE PLATFORM – TWO FORM FACTORS

CDP Public Cloud  
(platform-as-a-service)

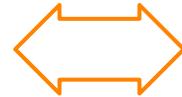
Control Plane



CDP On-Prem  
(installable software)



CLOUDERA  
**SDX**



Cloudera Runtime

# CDP ON-PREMISES



## CLOUDERA RUNTIME

Best of both new distribution  
50+ open source projects  
Foundational for CDP on-premises & CDP public cloud



## CDP PVC Base

Best of both product - "Unity"  
SDX Data Lake security  
Traditional Data Layer (HDFS) and On-prem object store (Ozone)  
Foundation for Private Cloud Experiences



## CDP PVC Experiences

New Management Console  
Compute & storage separation  
New containerized cloud-native analytic experiences (DW, ML, etc.)

# CDP PUBLIC CLOUD



## CLOUD INFRASTRUCTURE

IaaS and PaaS consumption  
Cloud vendor commitment/agreement  
Services required: VMs, Containers, Object Storage, Databases



## CLOUD OPTIMIZED ARCHITECTURE

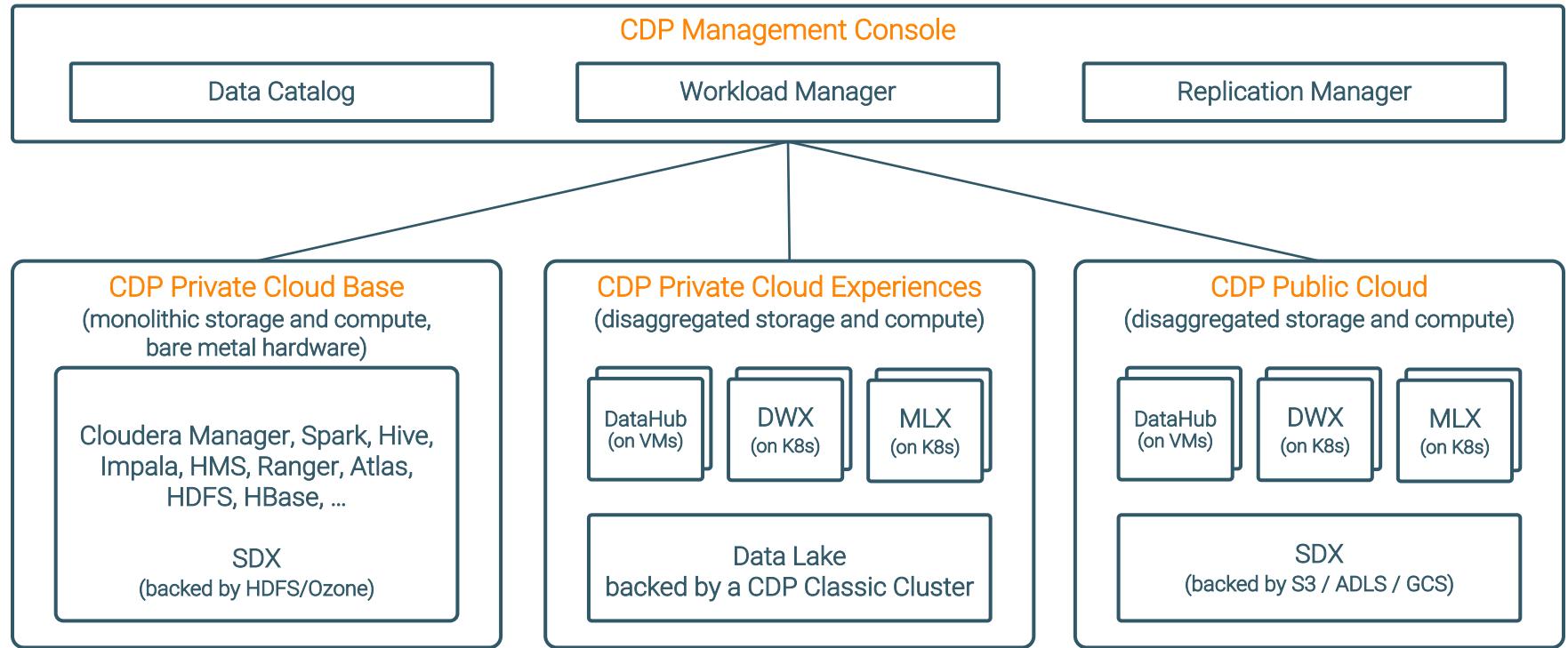
SDX Data Lake security data  
Data State  
Data Layer backed cloud object storage (S3, ADLS, GCS)  
Compute and storage segregation



## PLATFORM SERVICES

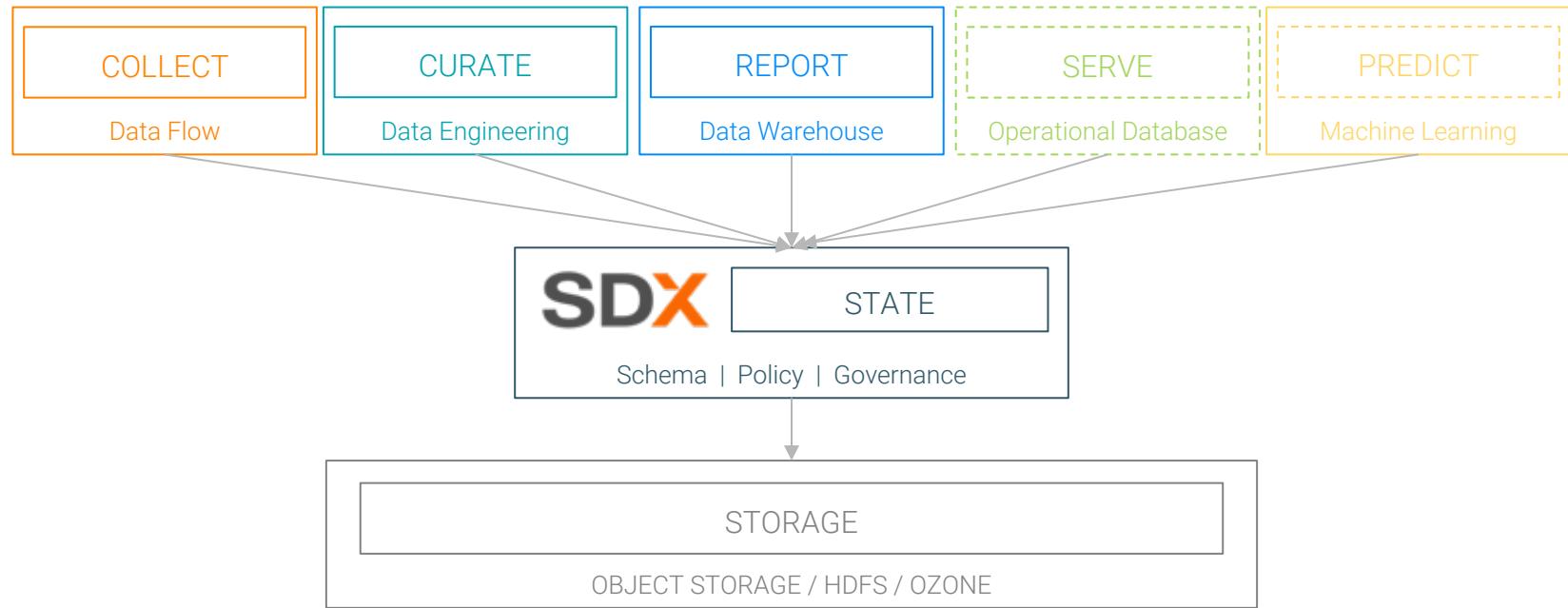
Data Hub, traditional cluster deployment  
Self-services Experiences, based on container processing resources  
On-demand consumption

# TECHNOLOGY: HYBRID, MULTI-CLOUD DEPLOYMENT ON CDP



# SDX

# CDP ARCHITECTURE WITH **SDX**



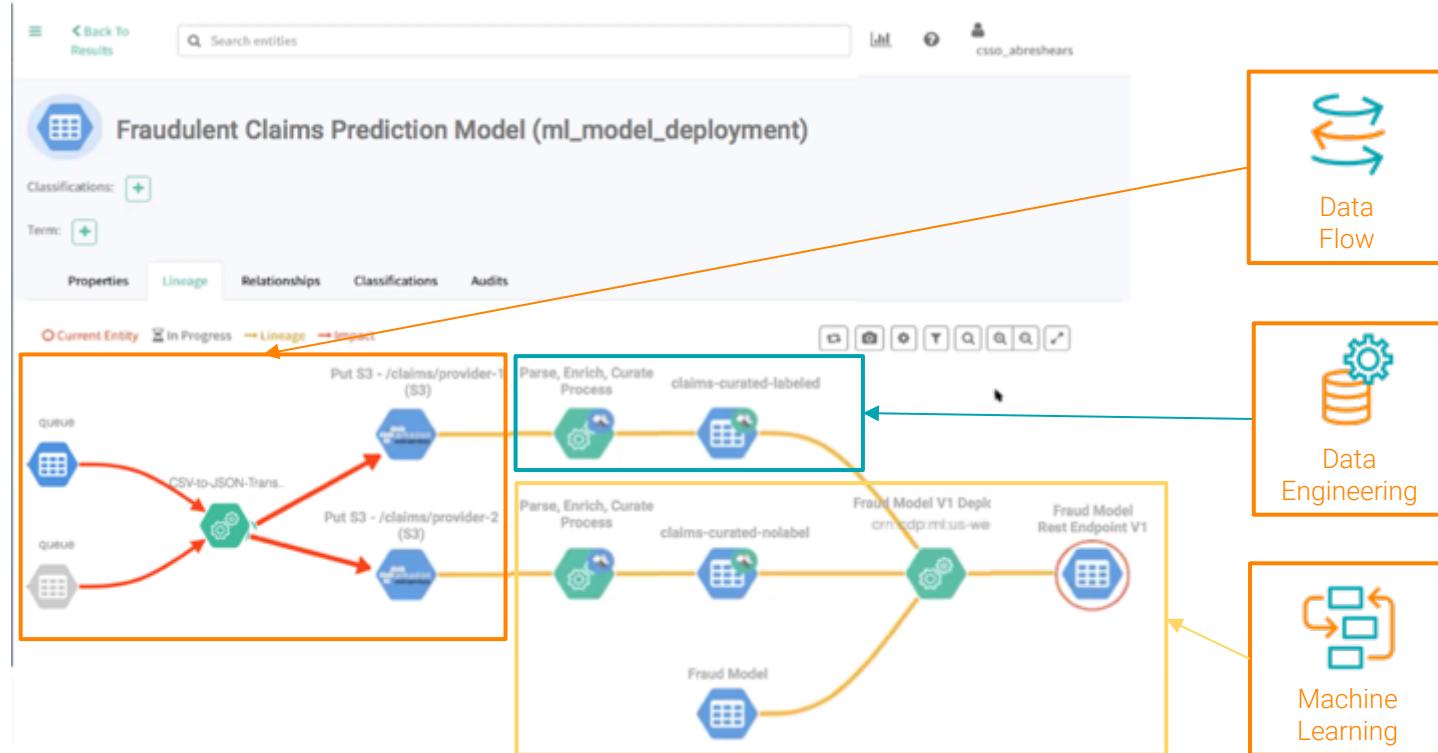
# CONSISTENT SECURITY AND GOVERNANCE

Built for multi-functional analytics anywhere



- **Data Catalog:** a comprehensive catalog of all data sets, spanning on-premises, cloud object stores, structured, unstructured, and semi-structured
- **Schema:** automatic capture and storage of any and all schema and metadata definitions as they are used and created by platform workloads
- **Replication:** deliver data as well as data policies there where the enterprise needs to work, with complete consistency and security
- **Security:** role-based access control applied consistently across the platform. Includes full stack encryption and key management
- **Governance:** enterprise-grade auditing, lineage, and governance capabilities applied across the platform with rich extensibility for partner integrations

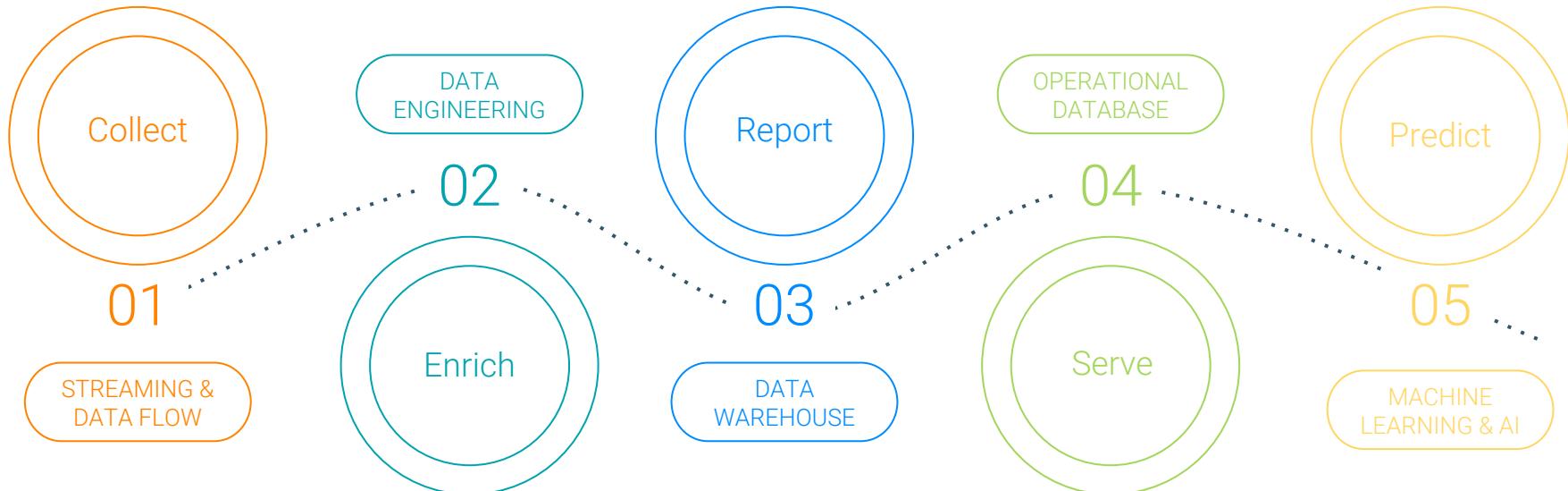
# SDX ENABLES GOVERNANCE FOR THE DATA LIFECYCLE



# RE CAP

# CLOUDERA - THE ENTERPRISE DATA CLOUD COMPANY

Manage and secure the data lifecycle in any cloud or datacenter



SECURITY | GOVERNANCE | LINEAGE | MANAGEMENT | AUTOMATION

# CLOUDERA DATA PLATFORM - UNIQUE CAPABILITIES

Enterprise Needs for Data Management and Analytics

DATA LIFECYCLE      vs      SINGLE FUNCTION

HYBRID & MULTI      vs      CLOUD ONLY

SECURE & GOVERNED      vs      BASIC SECURITY

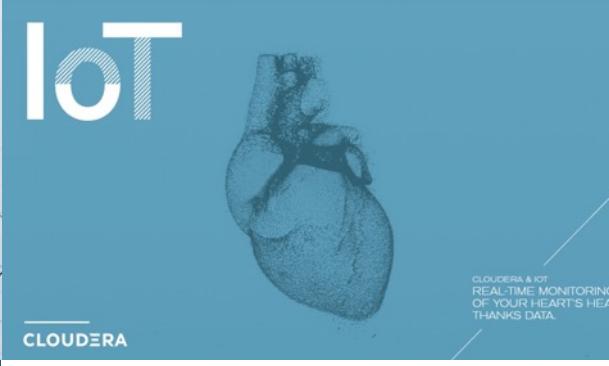
OPEN SOURCE      vs      PROPRIETARY LOCK-IN

PLATFORM      vs      POINT SOLUTION



# ENTERPRISE DATA CLOUD

CLOUDERA



CLOUDERA

CLOUDERA



CLOUDERA



# EDGE 2AI

CLOUDERA



# THANK YOU