



# CLOUDERA DATA PLATFORM

Oct 2021

# Cloudera Team



Suresh MR  
Director - Channels & Alliances



Vinay Rayker  
Partner Technology Leader

---

# AGENDA

- Introductions
- Hybrid Data Cloud for Large Enterprises
- What is Cloudera Data Platform (CDP) ?
- Why CDP?
- What's in it for you?
- Q&A

---

# Introductions

# CLOUDERA

THE ENTERPRISE DATA CLOUD COMPANY

---

We believe that **data** can make what is impossible today, possible tomorrow

---

We empower **people** to transform complex data into clear and actionable insights

---

We deliver cloud-native services to manage and secure the **data lifecycle** in any cloud or data center

# CLOUDERA FINANCIAL POSITION

## Q4 FY21 Results

**190**

Customers over  
\$1M in ARR

**\$778M**

ARR  
10% growth YoY

**90%**

Subscription  
software gross  
margin

**\$773M**

Cash, cash equivalents,  
marketable securities  
and restricted cash

**1,005**

Customers over  
\$100K in ARR

**\$206M**

Subscription  
software revenue,  
up 14% YoY

**17%**

Operating margin

**\$50M**

Operating  
income

Q4FY21 Earnings Release - March 10, 2021

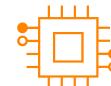
Notes: All numbers are from our Q4-FY21 financial results - Non GAAP.

# HOW DO CUSTOMERS USE CLOUDERA?

Every business use case is a data lifecycle use case



BANKING



TECHNOLOGY



TELCO



LIFE SCIENCES

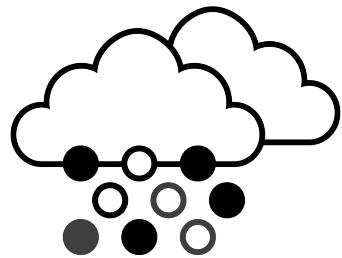


MANUFACTURING

USE CASES	Fraud detection • Anti-money laundering • Spend analytics	Customer analytics • Threat detection • Predictive support	Churn analysis • Customer care • Network optimization	Patient care (IoT) • Genomics research • Regulatory compliance	Predictive maintenance (IoT) • Supply chain optimization • Remote monitoring
KEY CUSTOMERS	• Barclays • Citi • Santander UK	• Cisco • Intel • Reef Technology	• Globe Telecom • Deutsche Telecom • Robi Axiata	• GlaxoSmithKline • Clearsense • Cerner	• Navistar • Micron • Sikorsky

# CLOUDERA

THE ENTERPRISE DATA CLOUD COMPANY

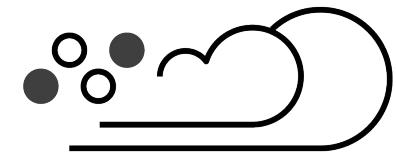


Any Cloud



Data Lifecycle

CLOUDERA  
**SDX**



Secure & Governed

Open

---

# Hybrid Data Cloud

# DATA MANAGEMENT IS SPREAD ALL OVER

# **CRITICAL SUCCESS FACTORS TO ACHIEVE A BETTER ENTERPRISE DATA STRATEGY IN A MULTI-CLOUD ENVIRONMENT**

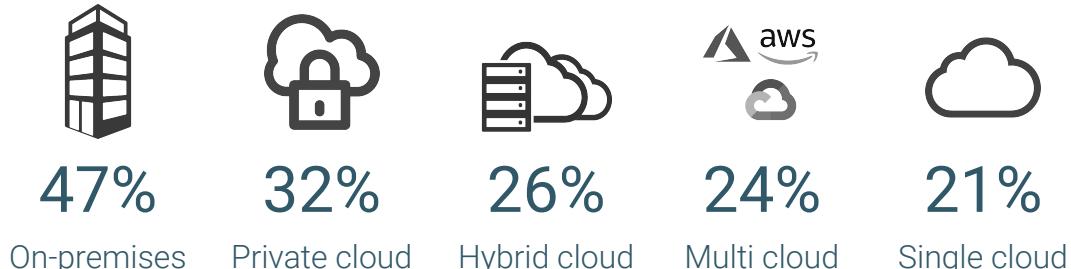
Data may be the fuel that powers the modern enterprise, but it is only as useful as the ability to explore, analyze, and transform a deep well of information into intelligence and action. The most exciting organizations today bring an insight-driven approach to everything they do—from evolving their business models to embedding analytics capabilities across the organization and incentivizing corporate culture to become more galvanized by data and the interpretation of the story it tells.

To achieve the promise of an enterprise data strategy, most organizations plan to increase dramatically their use of analytics and cognitive capabilities for everything from business intelligence (BI) to machine learning to artificial intelligence (AI). A Harvard Business Review Analytics Services survey of 185 global executives across a wide range of industries finds that the vast majority of executives (73%) believe that their data resources hold the key for creating business value—including 33% who give it the highest possible rating: 10 out of 10 in terms of importance. More than

"Our data platform is like electricity," says Mike Flynn, principal at PwC. "We want to make it super easy and super fast, for all of our teams around the world to plug

The survey finds, however, that there are a number of challenges to realizing the full potential value of enterprise data that may be distributed across private, hybrid, and multiple public-cloud environments. Data siloes, data lock-in, and lack of interoperability are among the greatest strategic impediments that organizations face, according to respondent—all problems that can intensify without a strategy to manage data spread across multiple clouds.

[View more about our access management solutions](#)



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.”

# ENTERPRISES ARE EMBRACING PRIVATE CLOUD

IDC Research - Cloud Growth, Migration, and Repatriation Continue to Gain Momentum

67%

Of enterprise workloads  
run on public and private  
cloud implementations

84%

Of enterprises report  
repatriating some workloads  
from public cloud

52%

Of repatriated workloads  
move to private clouds

*"As enterprise customers gain cloud expertise they're placing investments in **private cloud solutions** for increased security, compliance, performance, control and cost savings.  
Private clouds often act as a stepping-stone in the hybrid cloud journey."*

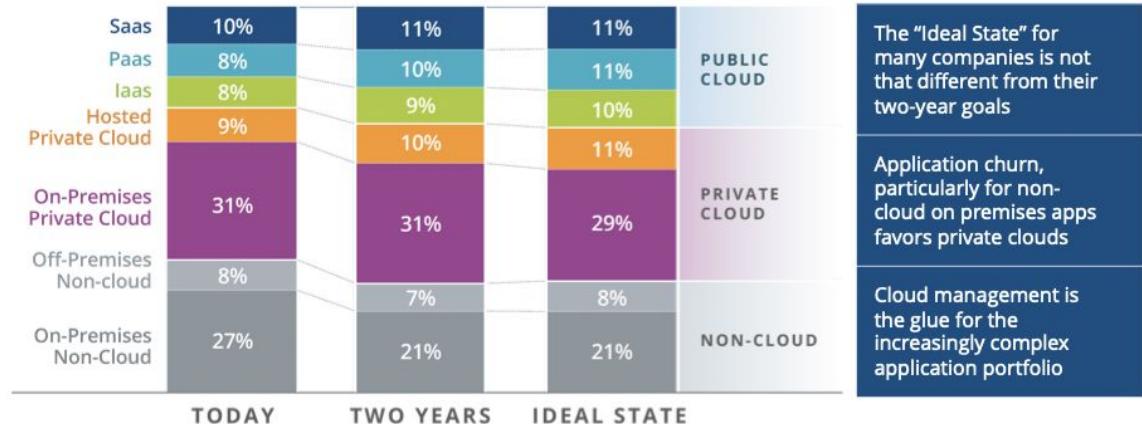
IDC, [Cloud Growth, Migration, and Repatriation Continue to Gain Momentum](#), Michelle Bailey, Chris Kanthan, March 2020  
IDC, [Cloud Pulse 1Q20: Survey Findings](#), Doc # US46396720, May 2020

# IDC Cloud Pulse Survey

40% of apps in private cloud is ideal state

Hybrid cloud management is key

## Customers Default to Multi-cloud Environments Private Cloud a strong focus for on- and off-premises solutions



Source: IDC's Cloud Pulse Q119, June 2019, n=2211

© IDC

8

# WHAT F500 COMPANIES ARE STRUGGLING WITH?

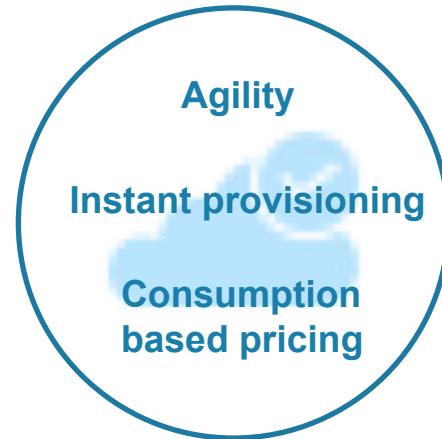
Cloud brings very attractive benefits, however it brings some strong challenges to overcome

## Lack of unified Security & Governance

Each cloud native product re-creates its own silo of metadata

## Lack of support

Managed services ensure that the service spins up and down, not that your app will be running



## LOBs going to the Cloud in an uncontrolled manner

e.g. a data scientist using \$700 K in a quarter due to clusters not spun up at a Telco in APAC

## Cost and delay to re-write the apps

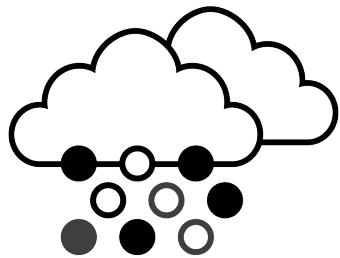
\$2.0 – 2.5 M per app to re-write them for a cloud native service

## Cloud vendor lock-in

Once the apps are written for a specific cloud-native service

# CLOUDERA

THE ENTERPRISE DATA CLOUD COMPANY

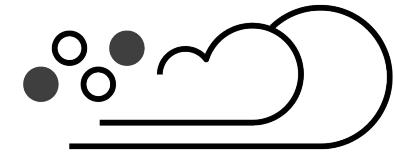


Any Cloud



Data Lifecycle

CLOUDERA  
**SDX**



Secure & Governed

Open

---

# What is CDP?

# CLOUDERA - THE ENTERPRISE DATA CLOUD COMPANY

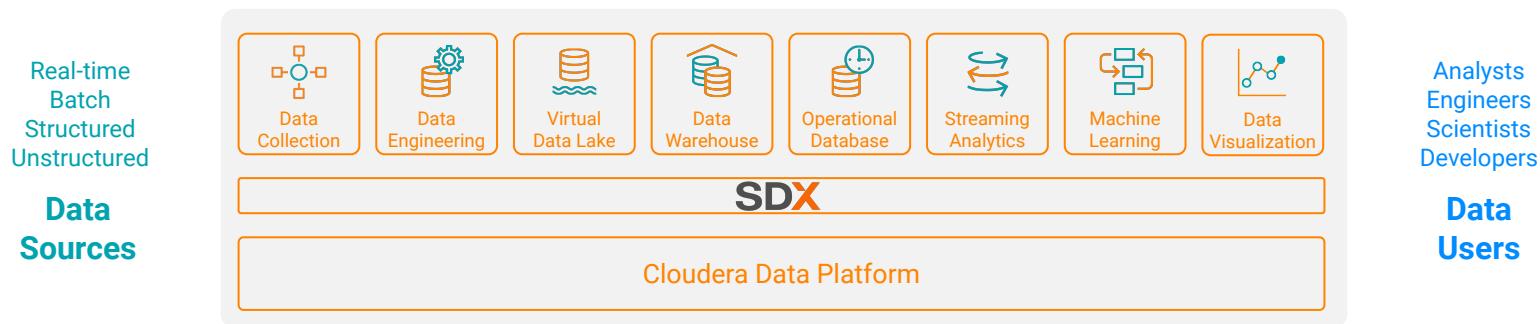
Manage and secure the data lifecycle in any cloud or datacenter



**CLOUDERA**  
**SDX**

SECURITY | GOVERNANCE | INTELLIGENCE | METADATA | CATALOG

# A HYBRID / MULTI-CLOUD DATA PLATFORM AND AN INTEGRATED SUITE OF SECURE ANALYTIC APPS



**Data Lifecycle**  
Integration for better user productivity and faster time to value



**Hybrid & Multi-Cloud**  
to leverage existing investments and reduce risk



**Secure & Governed**  
to simplify data protection, sharing and compliance



**Open & Extensible**  
to support more use cases faster and at lower cost

# ONE PLATFORM – TWO FORM FACTORS

## CDP Public Cloud

(platform-as-a-service)

## CDP Private Cloud

(platform as installable software)

### Control Plane



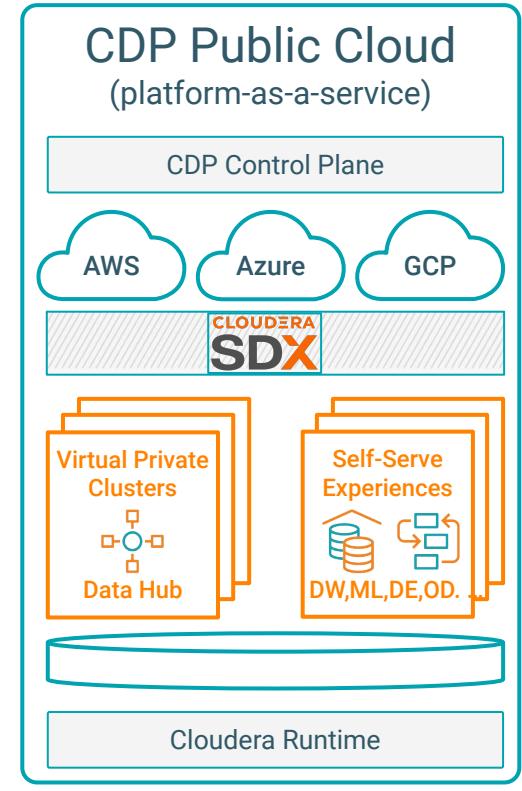
**SDX**



### Cloudera Runtime

# CLOUDERA DATA PLATFORM – PUBLIC CLOUD

- Available on AWS, Azure & GCP
- VM-based Data Lake and Data Hub clusters
- Containerised workloads:
  - Cloudera Data Warehouse (CDW)
  - Cloudera Machine Learning (CML)
  - Cloudera Data Engineering (CDE)
  - Cloudera Operational DB & Cloudera Data Flow
- Unlike other public cloud services, your data will always remain under your control in your VPC
- Control cloud costs by automatically spinning up workloads when needed and suspending their operation when complete



# CDP Public Cloud – Key Benefits

There are a number of common benefits that CDP delivers



Simplify Data  
Analytics



You Own  
Your Data



First Class  
Security



Hybrid  
Flexibility



Common  
Skill Set



Data  
Lifecycle



Easy and  
Portable

# WHAT IS CDP PRIVATE CLOUD?



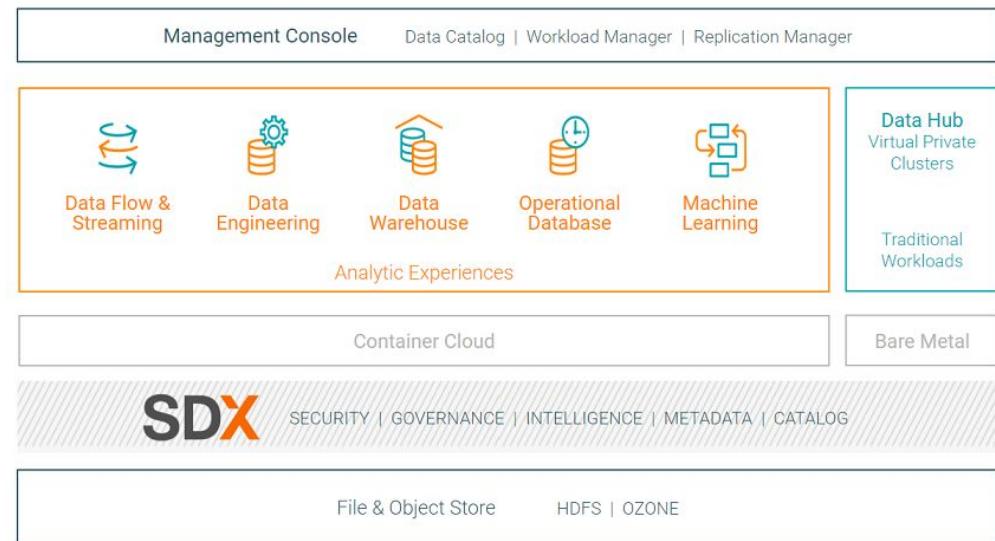
**New set of data analytics applications**  
Featuring use-case optimized interfaces



**Running on a container cloud**  
Fast provisioning & scaling, efficient, simple



**With access to a shared data lake**  
That is secured and governed

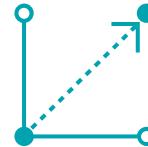


# ADVANTAGES OF CDP PRIVATE CLOUD



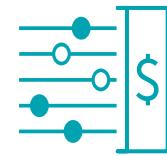
## Faster Time-to-Value

Simplified provisioning and onboarding of new use-cases



## Meet SLAs with Predictable Performance

Workload isolation and better managed multi-tenancy minimizes spikes impacting critical workloads



## Improve Cost Efficiency

Optimized resource utilization from disaggregated storage affords cost efficiencies across the cluster

---

# Why CDP?

# CLOUDERA BUSINESS VALUE

Empowering people to transform complex data into clear and actionable insights



FASTER TIME  
TO VALUE

**2X**

Increased  
user productivity



BETTER  
TCO/ROI

**60%**

Higher ROI  
versus competition

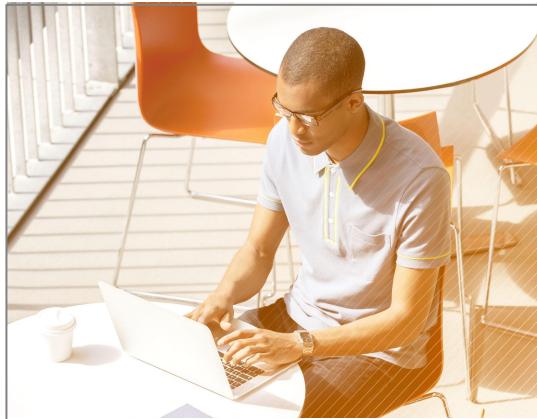


ONE PLATFORM  
UNLIMITED UTILITY

**1**

One Data Cloud  
for Business & IT

# CLOUDERA POWERING DATA-DRIVEN CUSTOMERS



**LinkedIn**<sup>®</sup>

At-scale migration from on-premises to public cloud for increased flexibility and scale



**REEF**

Enabling new growth opportunities via  
CDP Public Cloud and streaming analytics



**ExxonMobil**

Increasing ROI via a hybrid cloud for  
massively scalable, real-time data ingestion

# INDUSTRY ANALYST RECOGNITION

## Enterprise Data Cloud

*Enterprise Data Platform*



January 2021

*Cloud Data Ecosystems*



January 2020

*Enterprise Intelligence Platforms*



December 2019



May 2021

## Cloudera Data Platform (CDP)

*...To realize the full potency of hybrid cloud, organizations really need a holistic approach to the entire data lifecycle. Before CDP, they had to assemble the pieces themselves – a costly, time-consuming undertaking with potential gotchas lurking at every turn..."*



January 18, 2021



January 22, 2021

*...CDP is an enterprise data platform built on open-source software...that offers key data analytics and artificial intelligence functionality. CDP can leverage all data types, including structured and unstructured data, relational data and streaming data from any point in the data lifecycle...."*

# CLOUD DATA WAREHOUSE PERFORMANCE TESTING

## Cloudera Delivers Better Price Performance

Industry standard TPC Benchmark

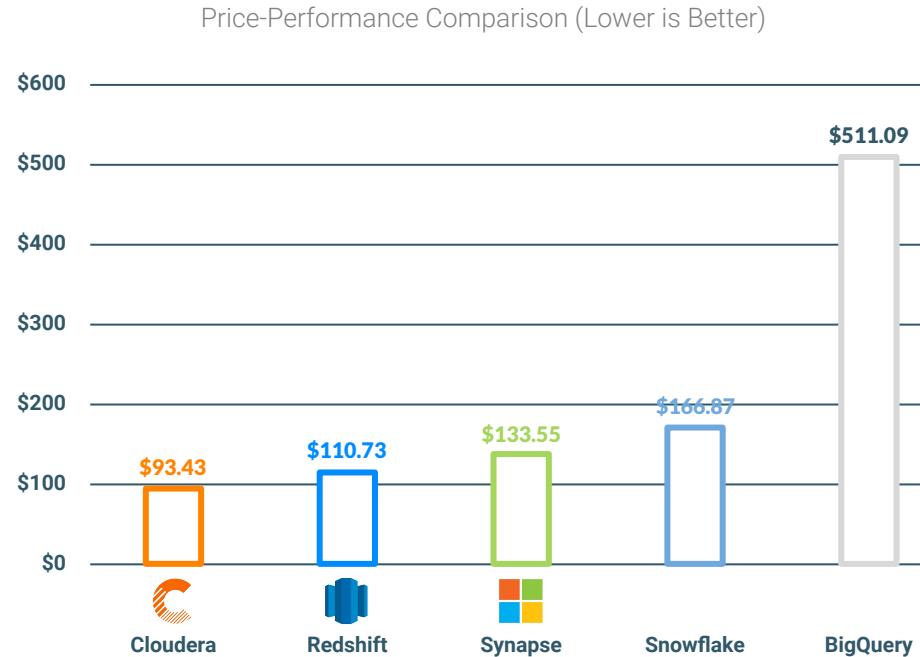
20% lower costs than  
Amazon Redshift

40% lower costs than  
Microsoft Synapse

80% lower costs than Snowflake

550% lower costs than Google

## Cloud Data Warehouse Performance Testing - February 2021

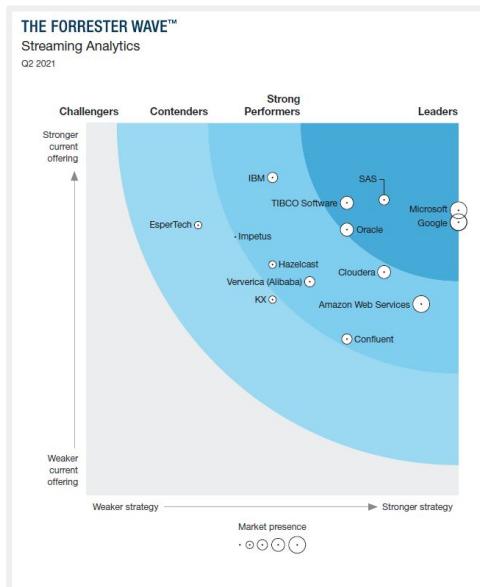


More can be learned about the TPC-DS benchmark at <http://www.tpc.org/tpcds/>.

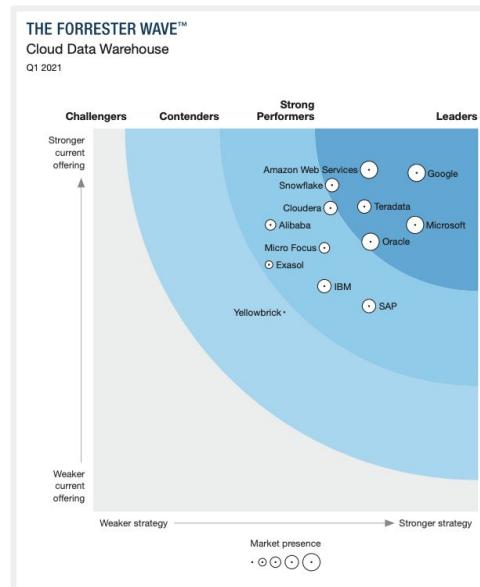
<https://gigaom.com/report/cloud-data-warehouse-performance-testing-cloudera/>

# CLOUDERA RECOGNIZED IN FORRESTER WAVES

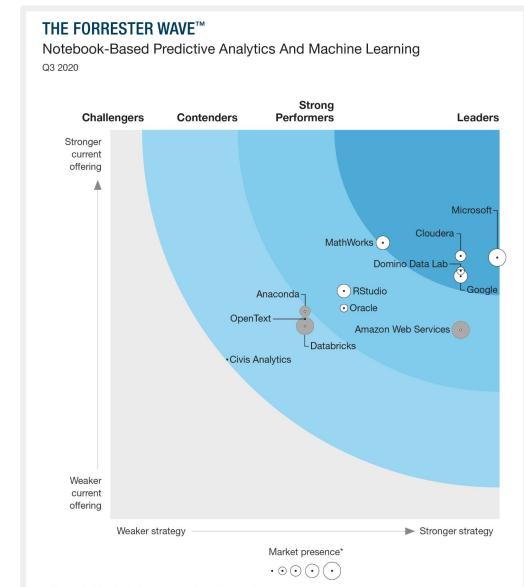
## Streaming Analytics



## Data Warehouse



## Machine Learning



The Forrester Wave™ is copyrighted by Forrester Research, Inc. Forrester and Forrester Wave™ are trademarks of Forrester Research, Inc. The Forrester Wave™ is a graphical representation of Forrester's call on a market and is plotted using a detailed spreadsheet with exposed scores, weightings, and comments. Forrester does not endorse any vendor, product, or service depicted in the Forrester Wave™. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change.

# CASE STUDY: FINANCIAL SERVICES CUSTOMER

Customer saves \$10M

## Before CDP



Amazon EMR

\$300K

\$12-16M

## After CDP



Amazon  
EC2



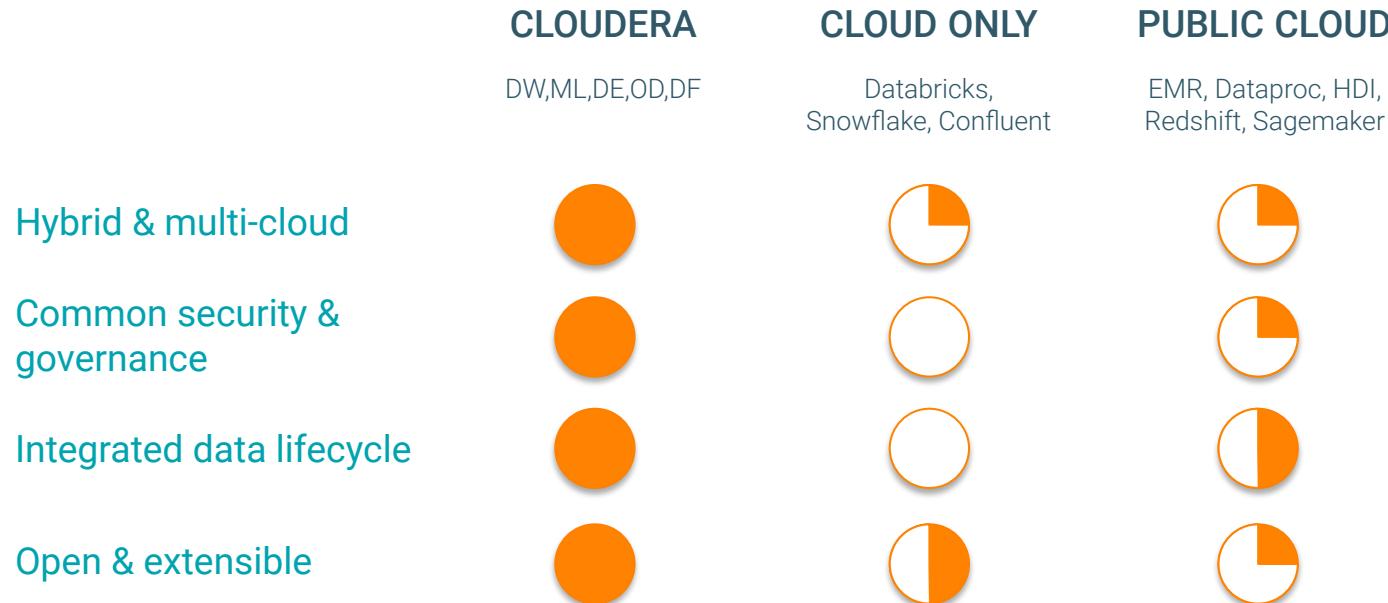
Amazon  
S3

\$850K

\$3M

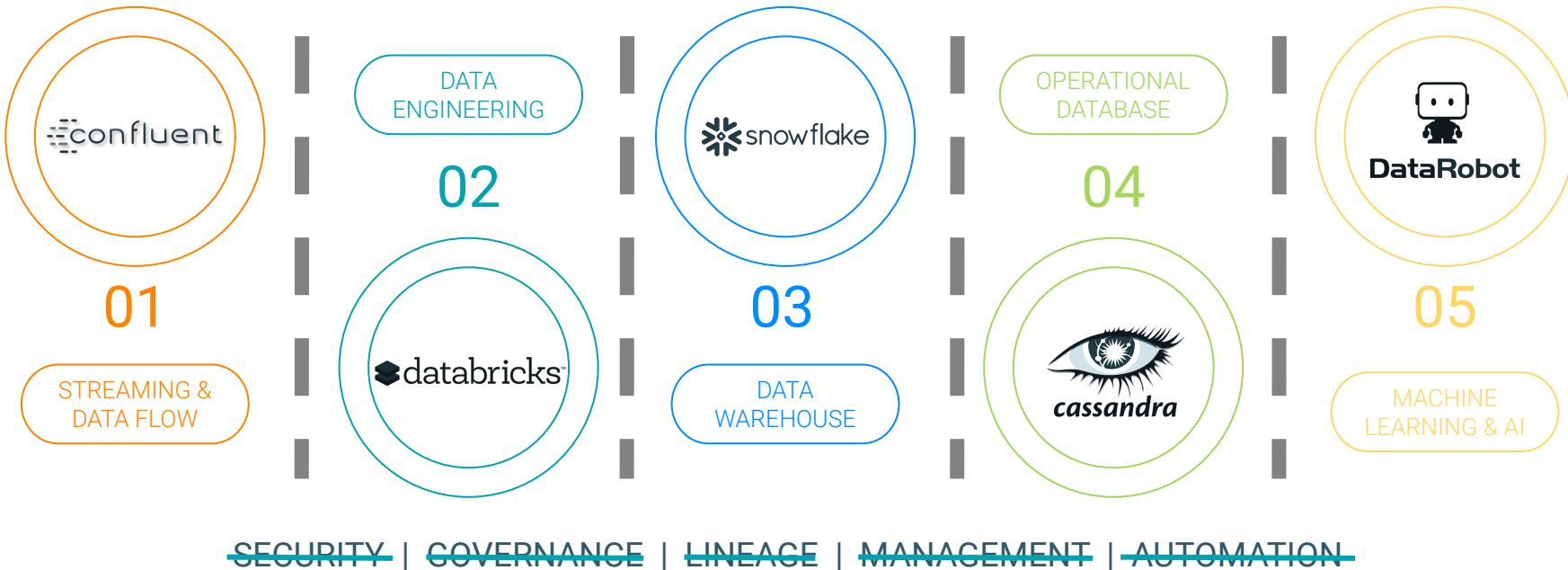
# COMPETITIVE LANDSCAPE

Cloudera stands out as a hybrid and multi cloud platform for data management and analytics with enterprise security and governance powered by open-source innovation

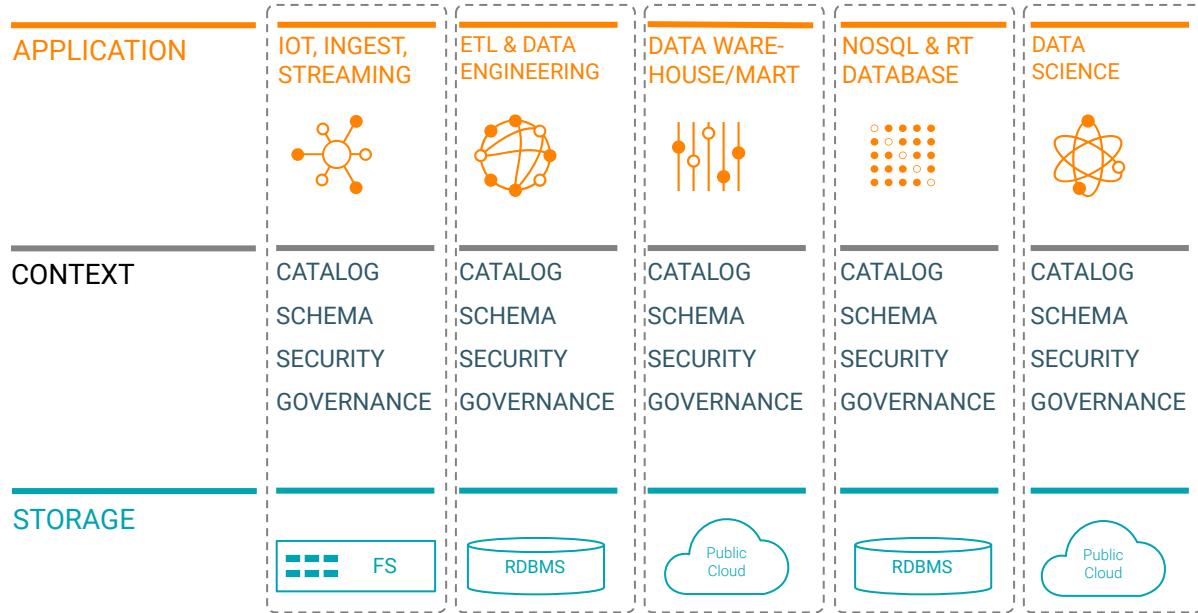


# POINT SOLUTIONS HAVE AN INTEGRATION TAX

Result: Higher Net OpEx, TCO ... Security and governance (GDPR, etc.) is an afterthought

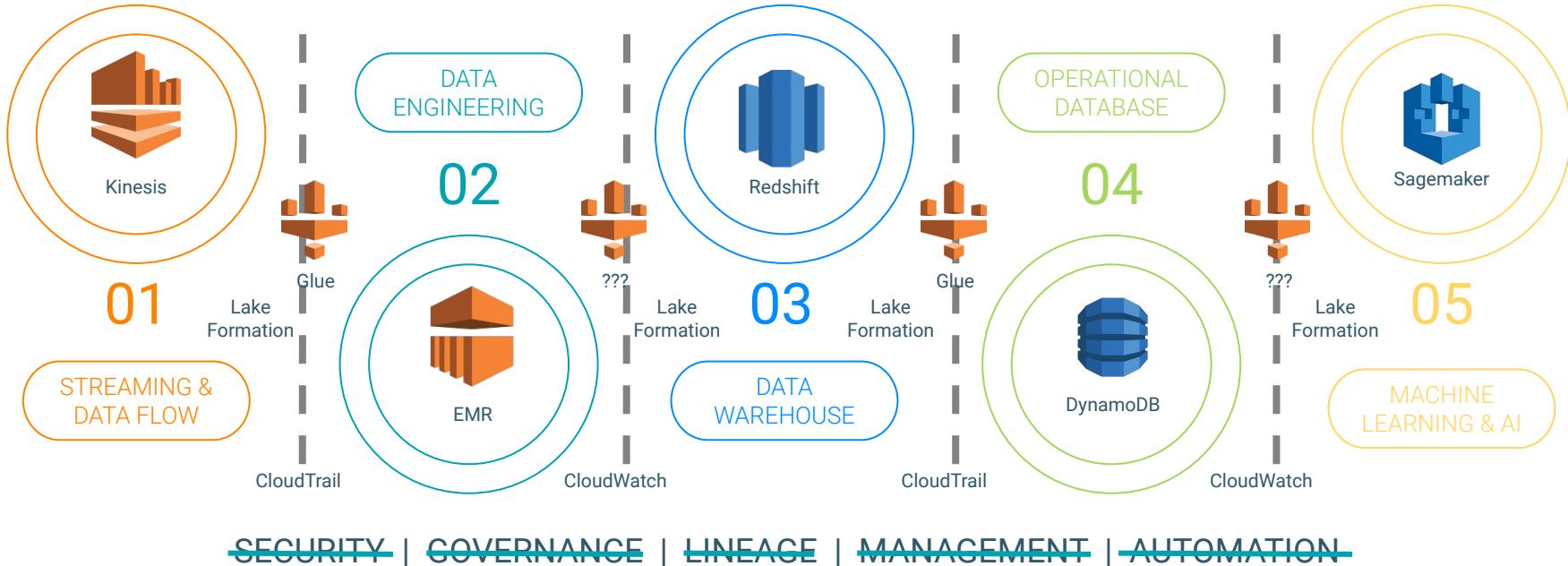


# DATA AND INSIGHT SILOS



# POINT SOLUTIONS FROM AWS

Easy to buy with basic security, limited integration and zero automation

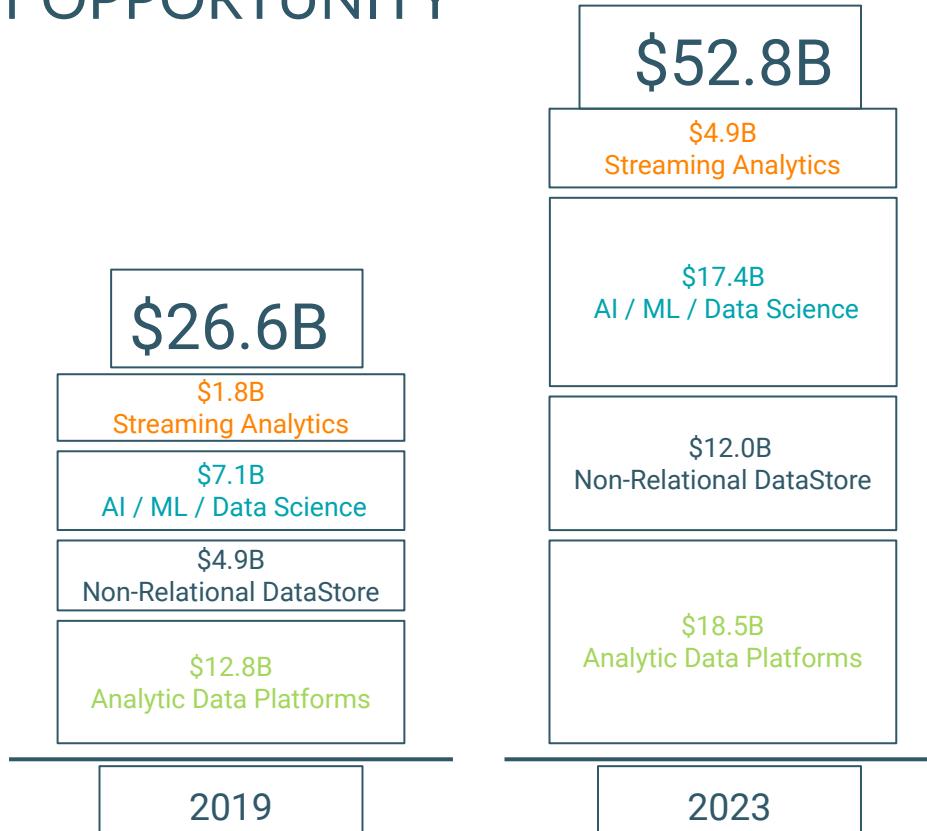


---



What's in it for you?

# MARKET OPPORTUNITY



## Outlook

Public cloud share will increase from 25% to 50% of market for Big Data & Analytics

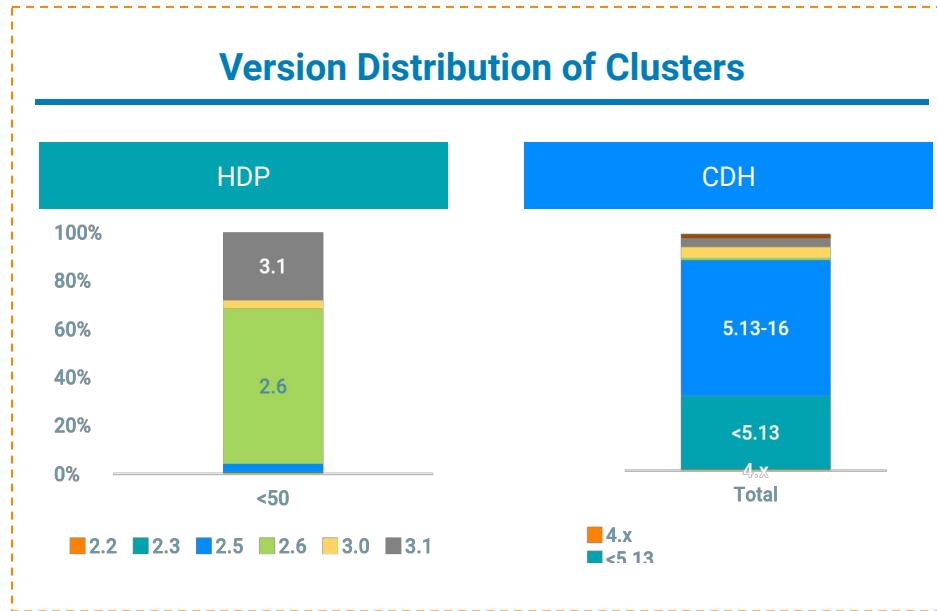
Cloud deployments expected to grow 35% CAGR

On-premises deployments expected to grow 9% CAGR

IoT-related data CAGR of 29% through 2025 not represented in market sizing

Source: IDC 2019

# ~70% OF CUSTOMERS ON HDP2 AND CDH5



### Current End of Support (EoS) Dates

The following table specifies the planned End of Support Schedule for Cloudera Enterprise. All future dates are provided for planning purposes only and are subject to change, but with the expectation that dates may move later but will not move earlier. In each case, the projected EoS Date can be considered to be the last day of the month specified in the table below.

Release	GA Date	EoS Date
Cloudera Enterprise 6.3	July 2019	March 2022
Cloudera Enterprise 6.2	March 2019	March 2022
Cloudera Enterprise 6.1	December 2018	December 2021
Cloudera Enterprise 6.0	August 2018	August 2021
Cloudera Enterprise 5.16	November 2018	December 2020
Cloudera Enterprise 5.15	June 2018	December 2020
Cloudera Enterprise 5.14	January 2018	December 2020

### End-of-Support Dates

#### Hortonworks Data Platform (HDP)

The end-of-support date is the last day of the month indicated.

Release	GA Date	End of Full Support
HDP 2.3	7/22/15	July 2018
HDP 2.4	3/1/16	March 2019
HDP 2.5	8/29/16	August 2019
HDP 2.6	4/3/17	December 2020
HDP 3.0	7/13/18	July 2021
HDP 3.1	12/11/18	December 2021

# FY22 SALES STRATEGY - CLOUDERA



New Logo  
Acquisition

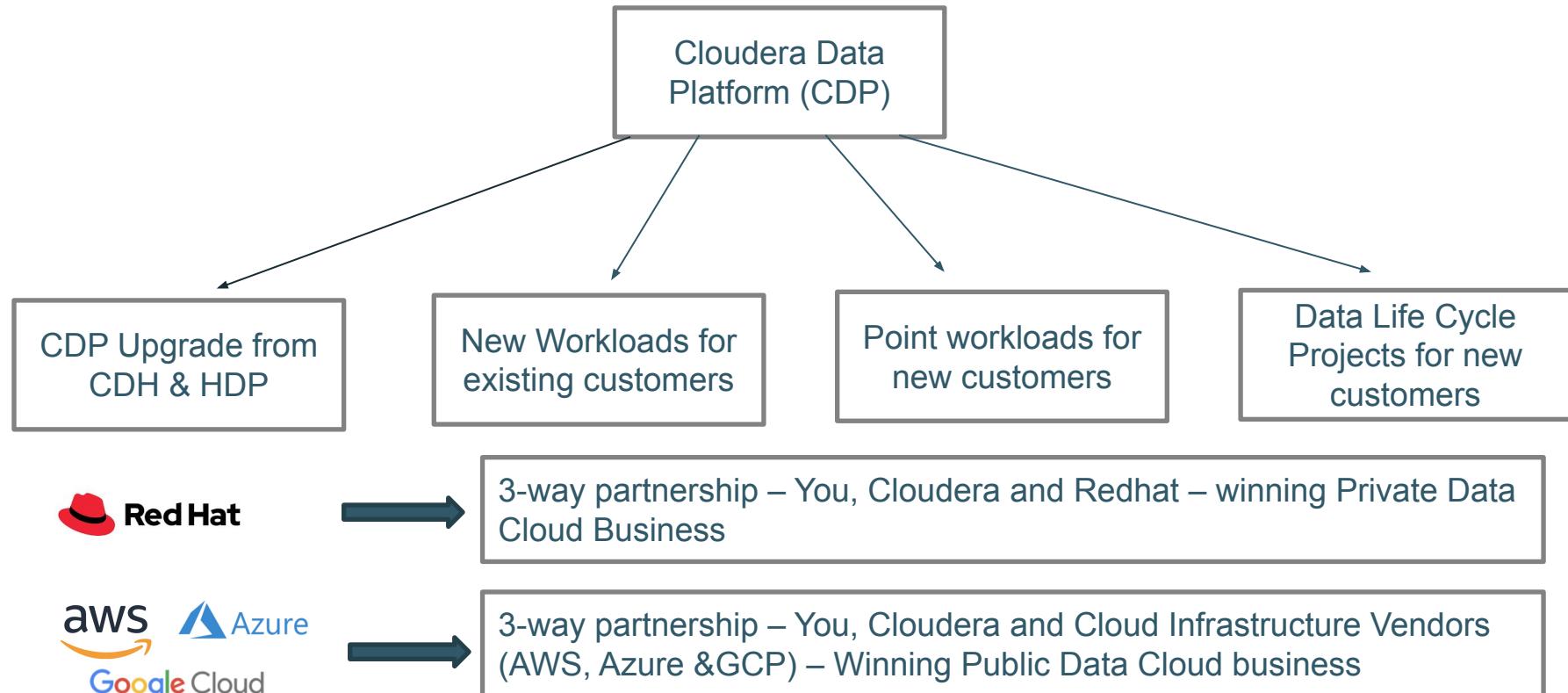


Pipeline  
Generation



Field  
Engagement

# Opportunities ahead of us



# Moving Customers to CDP - What's in it for you?

It's not just an upgrade - it's about re-architecting the data infrastructure to be truly Hybrid / Cloud-Native

Great opportunity for both high-value consulting as well as project delivery services

Delivers a cloud-native platform for the client that can handle many varied data workloads

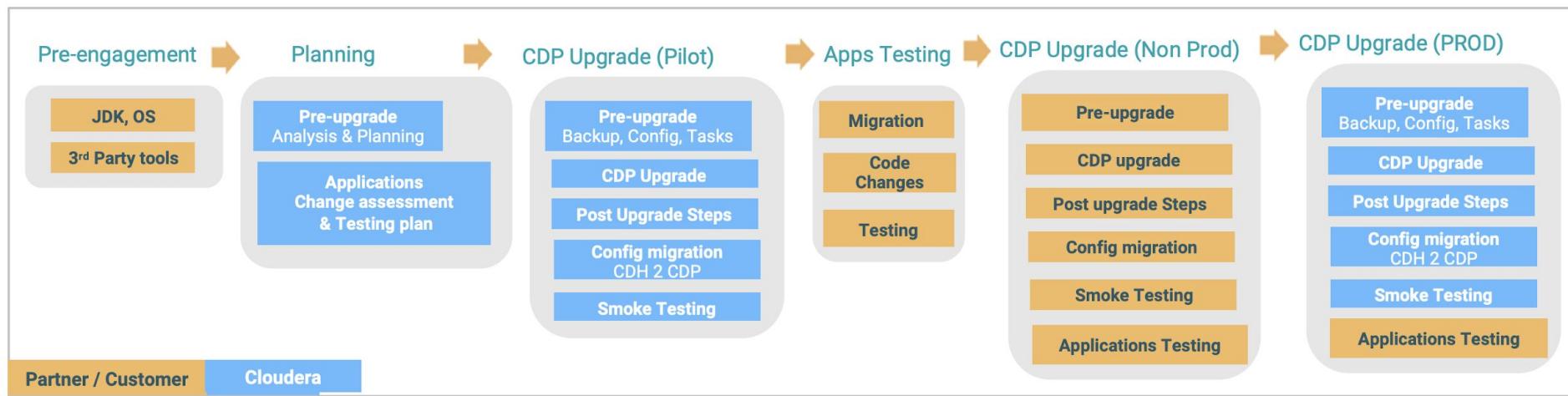
Deeper strategic client engagements with longer, multiple cycles

Deliver both cloud strategy and data strategy without compromise

Lower-risk integration of multiple data workloads in a single platform

Take advantage of AWS/Azure funding

# ACTIVITIES SUMMARY



- Analysis & Planning**
- Review CDH cluster deployment
  - Review App arch design
  - Review data pipeline
  - Review known issues
  - Review Sentry policies
  - Assess 3rd party tools compatibility risks
  - Applications changes assessment
  - Validate pre-requisites for upgrade
  - Create detailed plan

- Cluster Upgrade (Pilot)**
- Pre-upgrade prep & configuration
  - Upgrade to CDP
  - Post upgrade configuration
  - Post upgrade platform configuration migration
  - Smoke testing

# CLOUDERA DATA PLATFORM

## THE POWER OF “AND”



- Hybrid **and** multi-cloud
- Purpose-built **and** lifecycle analytics
- Simple to use **and** secure by design
- Open **and** extensible
- For data engineers **and** data scientists
- Faster time to value **and** better ROI/TCO
- The Data Cloud for Business **and** IT

---

# Q&A

# THANK YOU

CLOUDERA

# COMPETITIVE PRODUCT/ASSET ECOSYSTEM MAPPING

PRODUCT CATEGORY	CLOUDERA						
DATAFLOW AND STREAMING	CDF: NiFi, Minifi, EFM, Kafka, SMM, Kafka Streams, Storm, SAM	Kinesis, Kafka-as-a-Service	Stream Analytics, Data Factory, Events Hubs	Cloud Pub/Sub, Cloud Dataflow	Confluent Control Center Kafka-as-a-Service		
DATA WAREHOUSING	CDP: Hive LLAP, Impala	Redshift	HDI SQL data warehouse	BigQuery		Snowflake Connector	SQL Data warehouse Multi-cloud; cloud-only
OPERATIONAL DATABASE	CDP: Kudu, HBase, Phoenix	DynamoDB	MySQL, CosmoDB	Cloud BigTable, Spanner, Firestore			
DS, ML & AI	CML: CDSW, Spark, Zeppelin	SageMaker	AzureML, Databricks	ML Engine, Vision, Cloud AI		DBX Notebook Data Science Workbench	Databricks Connector
SECURITY & GOVERNANCE	CDP: SDX -Knox, Atlas, Ranger	IAM, Cloudformation	Security Center, Data Catalog, Active Directory	IAM, KMS, DLP	Role-based Access, Copy, Read/Write Controls	Role-based Access, Copy, Read/Write Controls	Role-based Access, Copy, Read/Write Controls
STORAGE & COMPUTE	CDP: HDFS, YARN	S3, Glacier, EC2, EKS	WASB, ADLS, Azure Compute, Kubernetes	Cloud Storage, Kubernetes		DBX Delta Lake Data Engineering	Native CSP
HADOOP DISTRIBUTION	CDP, HDP, CDH	EMR	Azure HDInsight	Dataproc			
PRIVATE CLOUD / HYBRID	CDP Private Cloud	Outpost	Azure Stack	Anthos			