Transforming Cloudera analytics agility in a fast-moving world

Brent Compton, Sr Director Tech Marketing, Red Hat
Idan Levi, formerly at GoI, now in Office of CTO, Red Hat

May 2019

The Story of Unlocking

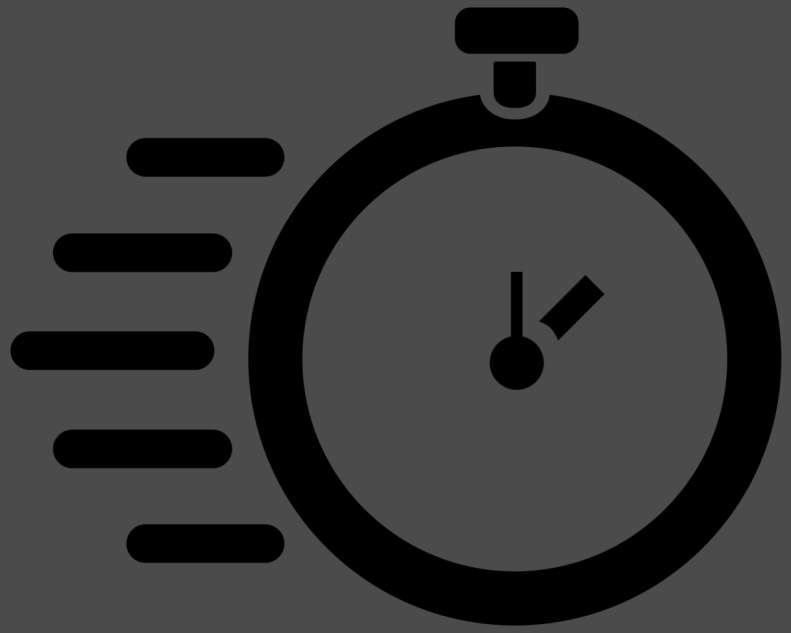
Data

Compute

Innovation

It’s All About Time to Market

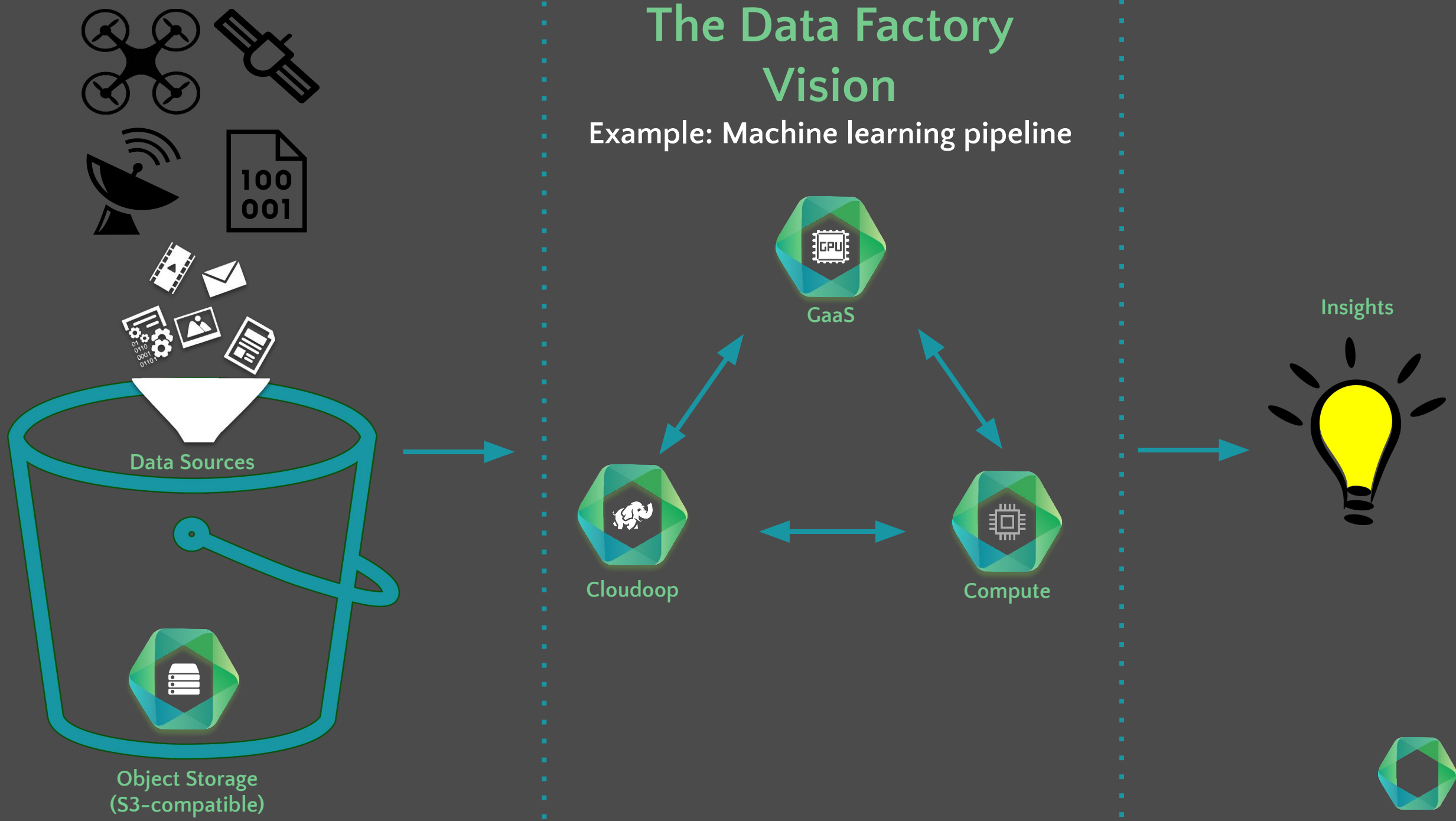
~ 4–5 Months



~ 8–14 Minutes

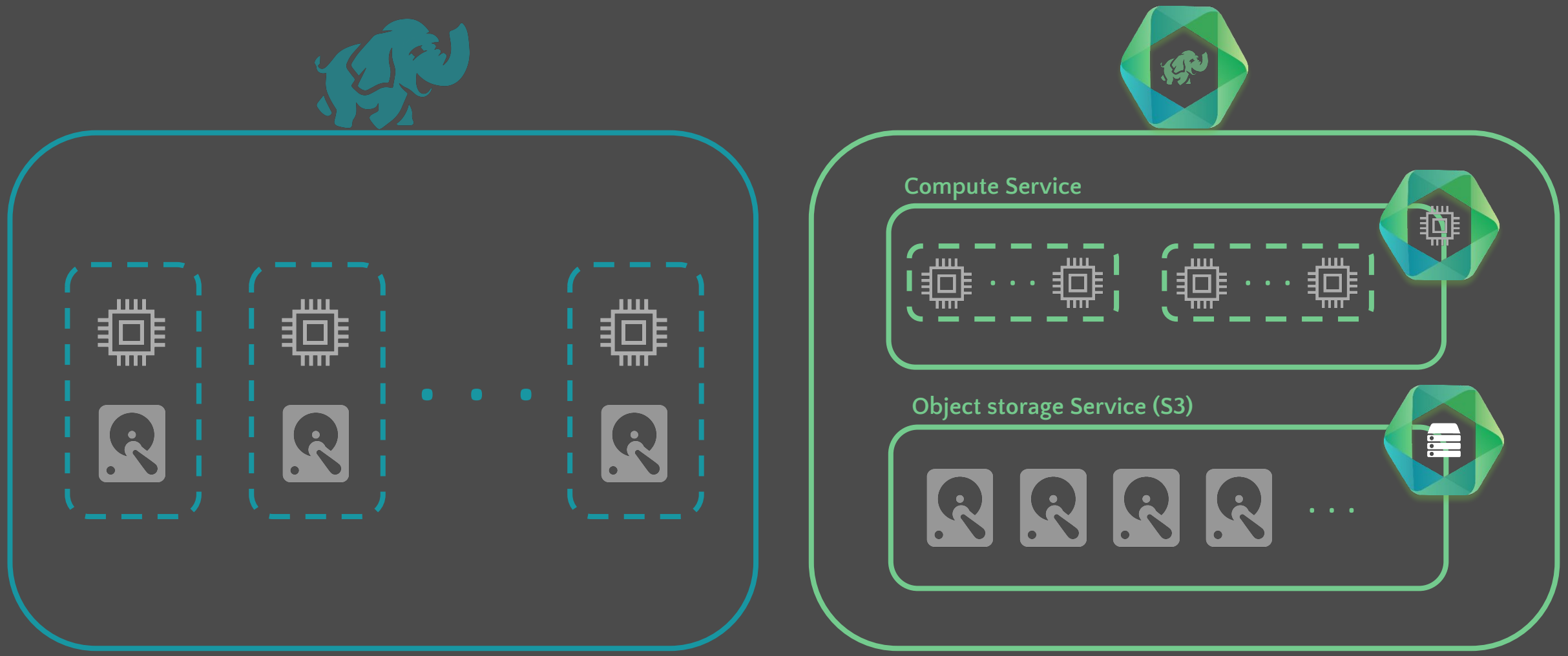
The Data Factory Vision

Example: Machine learning pipeline



Cloudoop

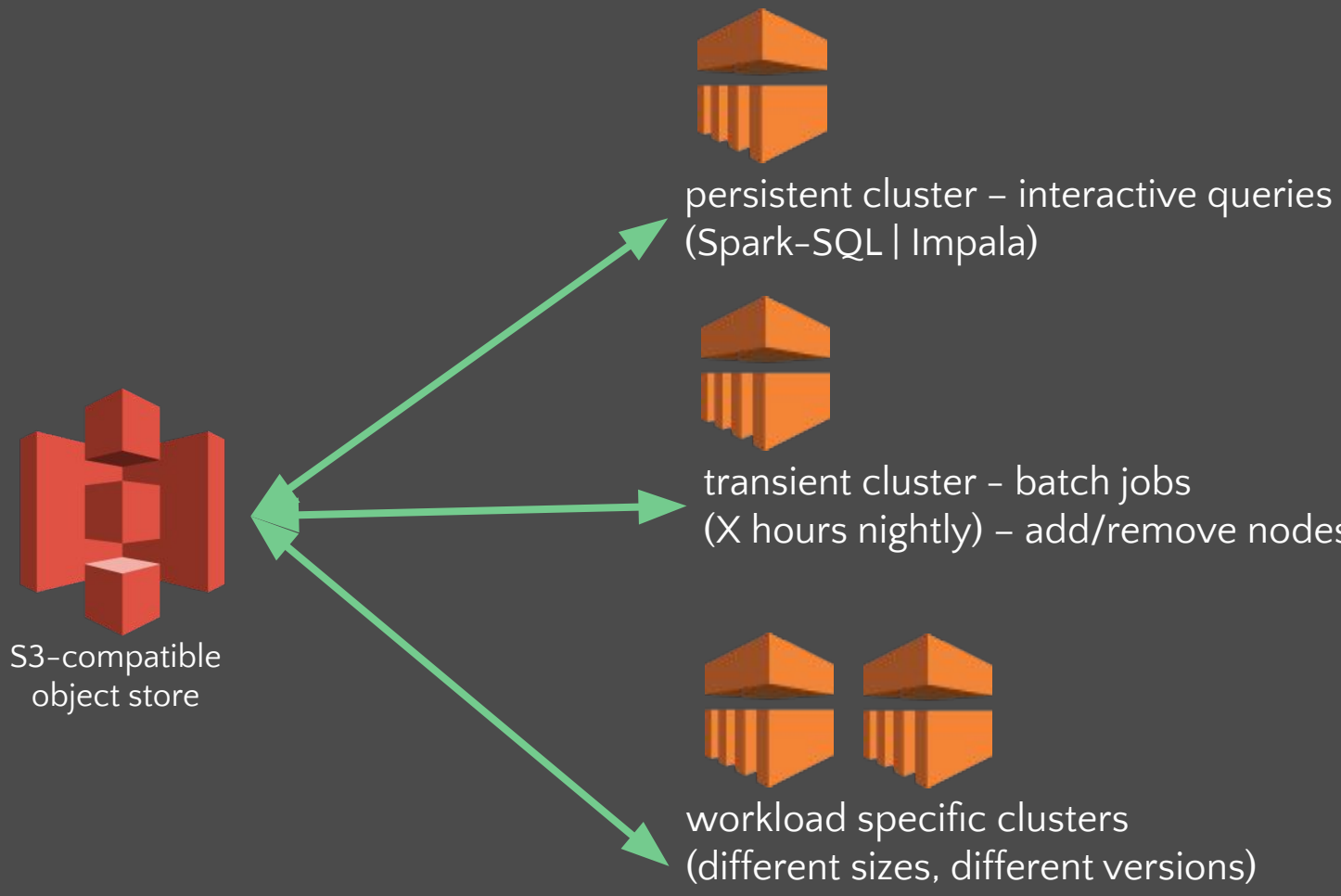
Native Hadoop vs. Cloudoop



Cloud Approach - Unlocking

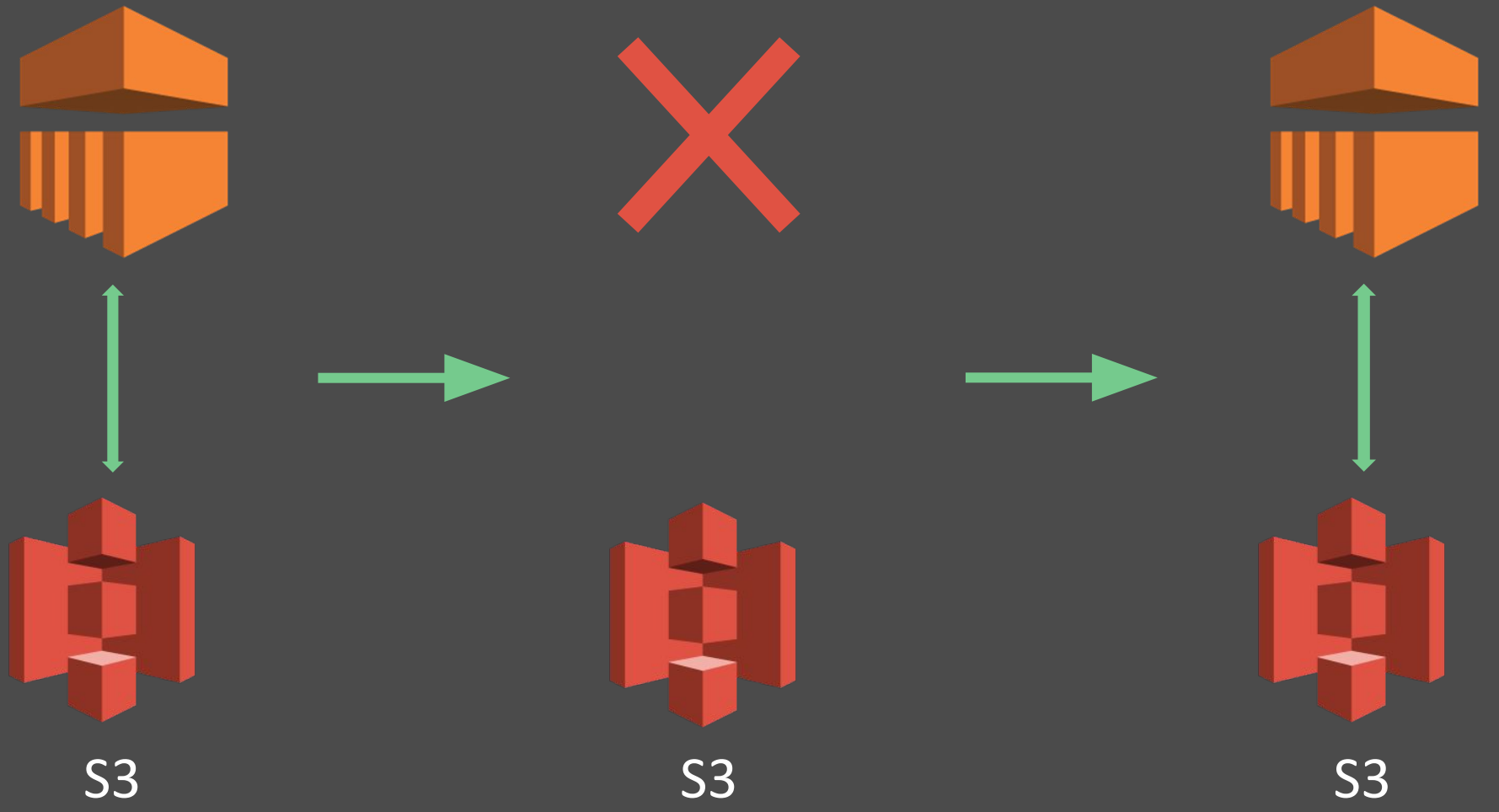
- Elasticity
- Compute-data decoupling
- Scalability
- Dynamic
- Multi tenancy
- SLA

Cloud Approach – AWS experience on-prem



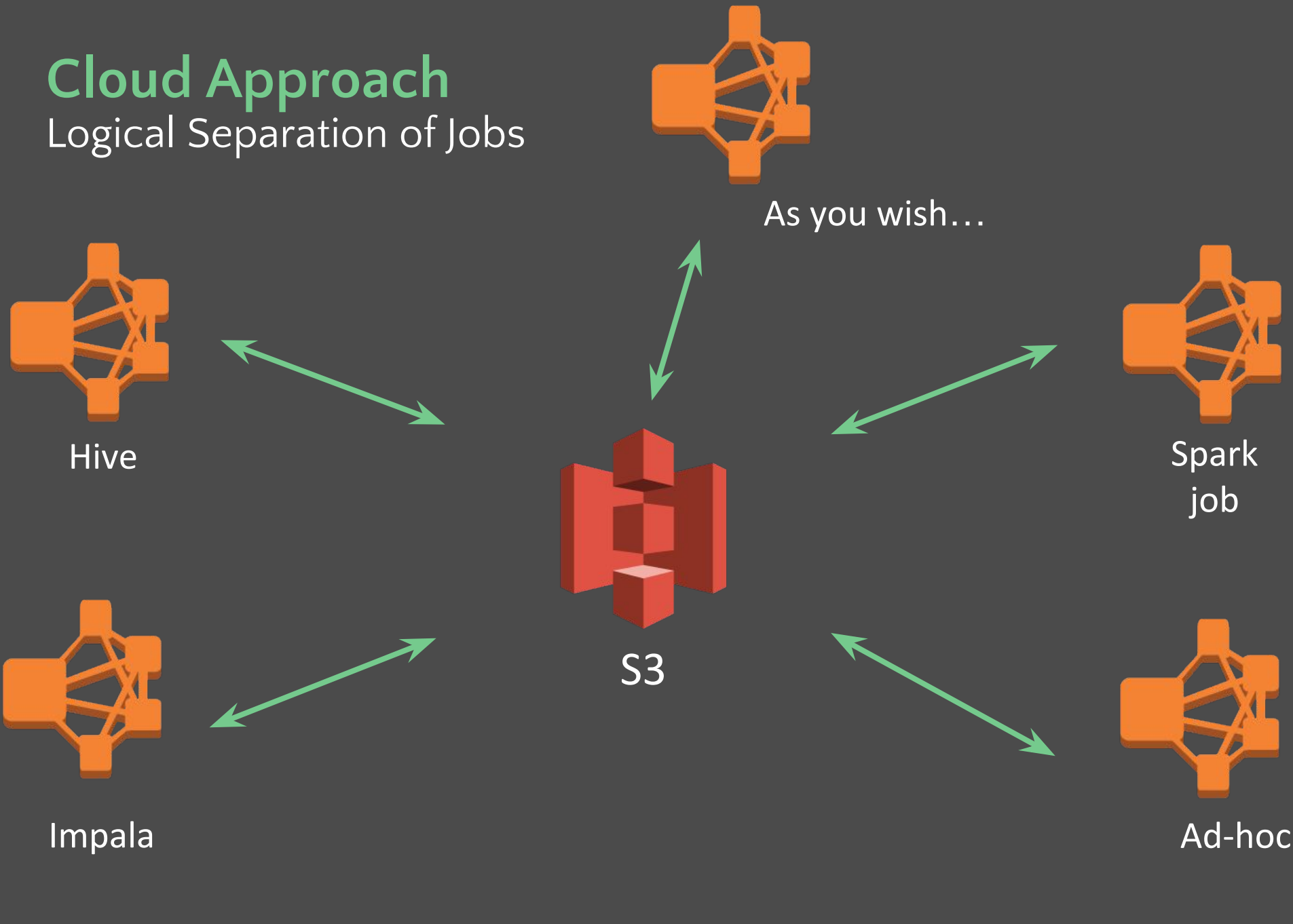
Cloud Approach

Switch Off Clusters



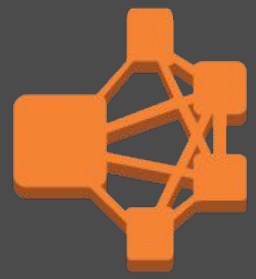
Cloud Approach

Logical Separation of Jobs

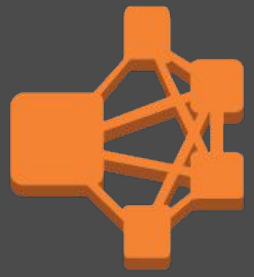


Cloud Approach

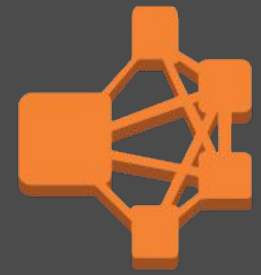
Logical Separation of Jobs



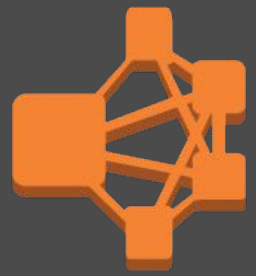
As you wish…



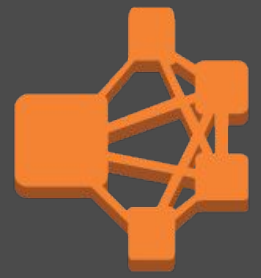
Hive



Spark Job

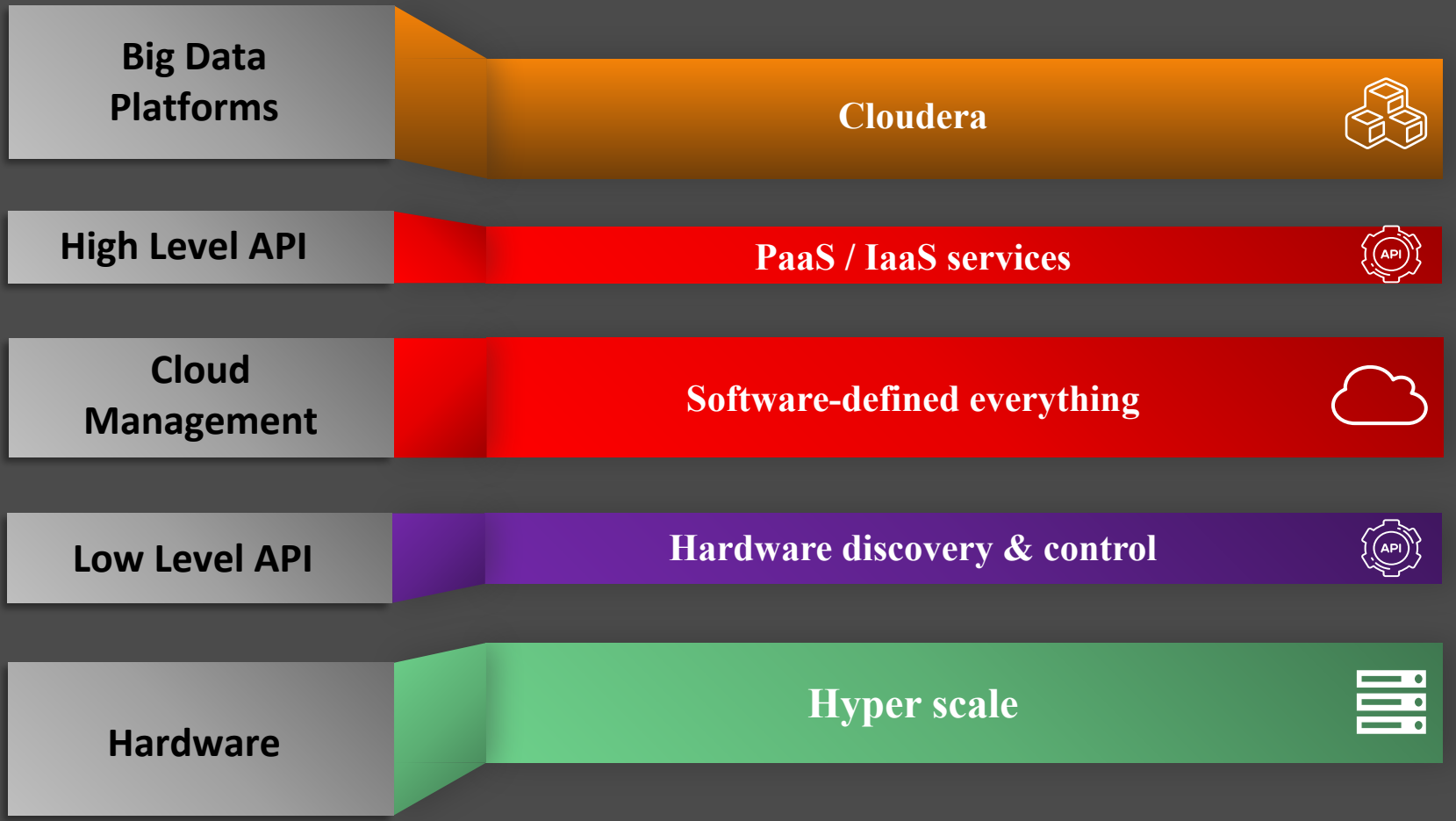


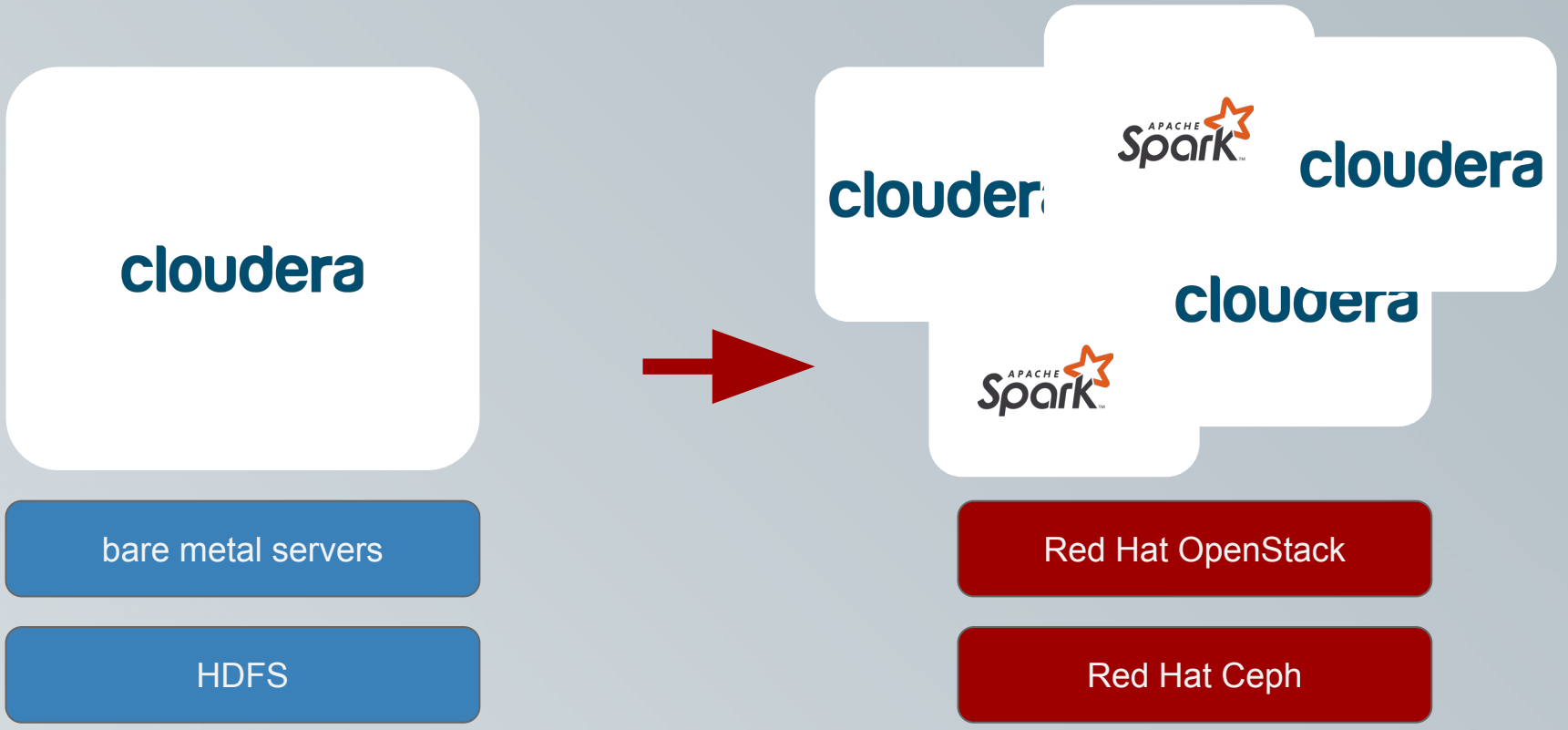
Impala



Ad-Hoc

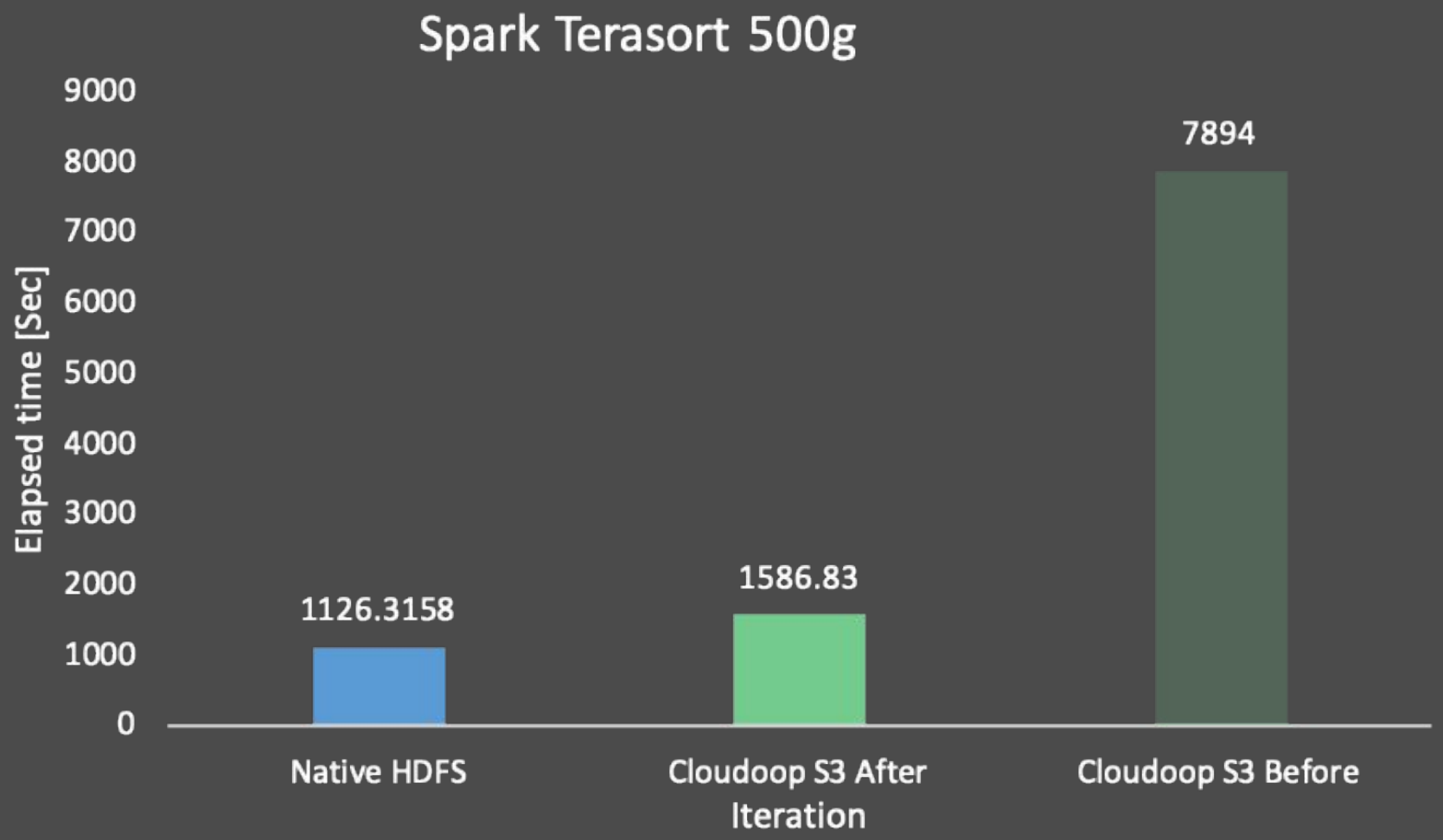
High Level Architecture




cloudera
bare metal servers
HDFS
cloudera
Spark
cloudera
cloudera
Spark
Red Hat OpenStack
Red Hat Ceph

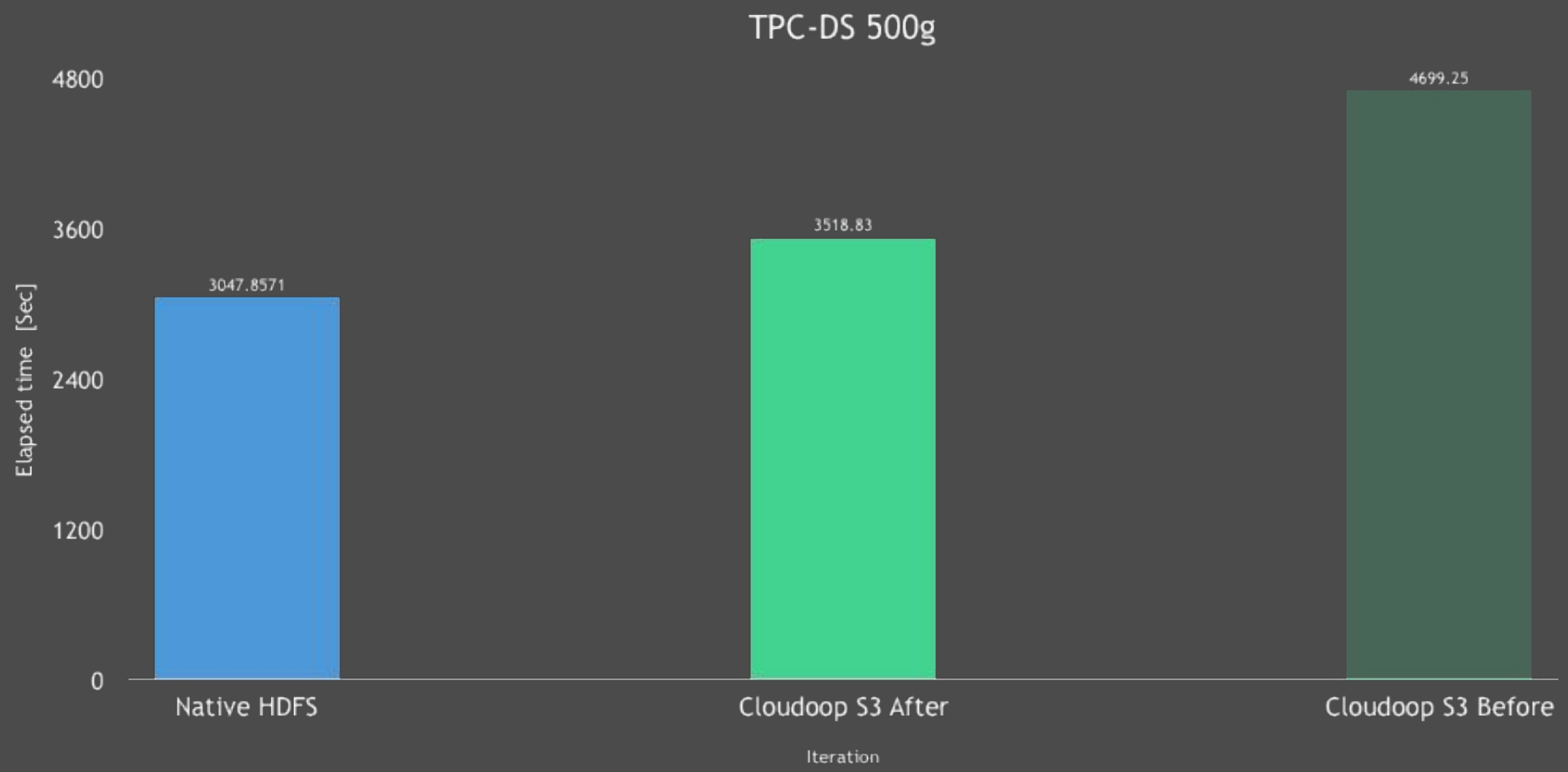
Proof Of Concept

Fine tune – Spark Terasort Before & After



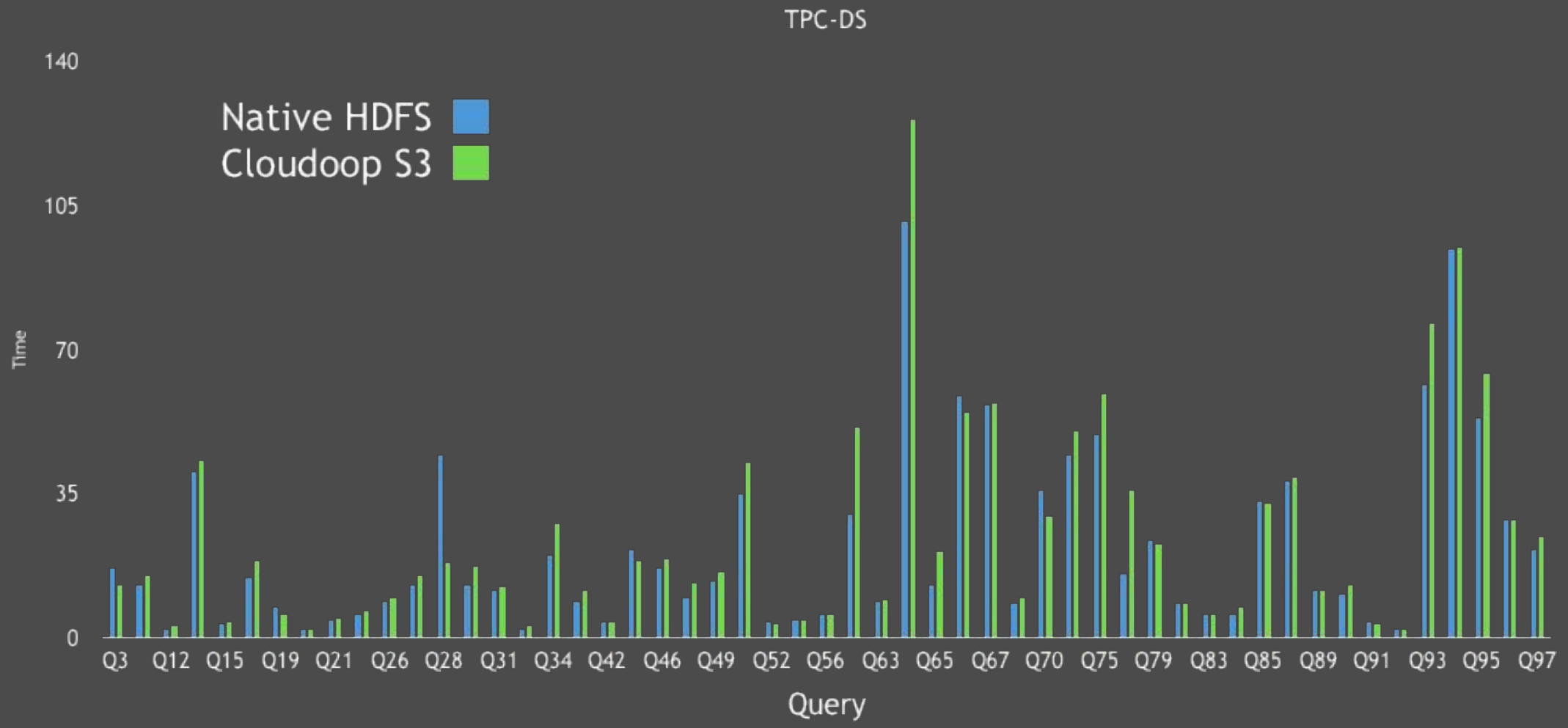
* Lower is better

Fine tune – TPC-DS Before & After



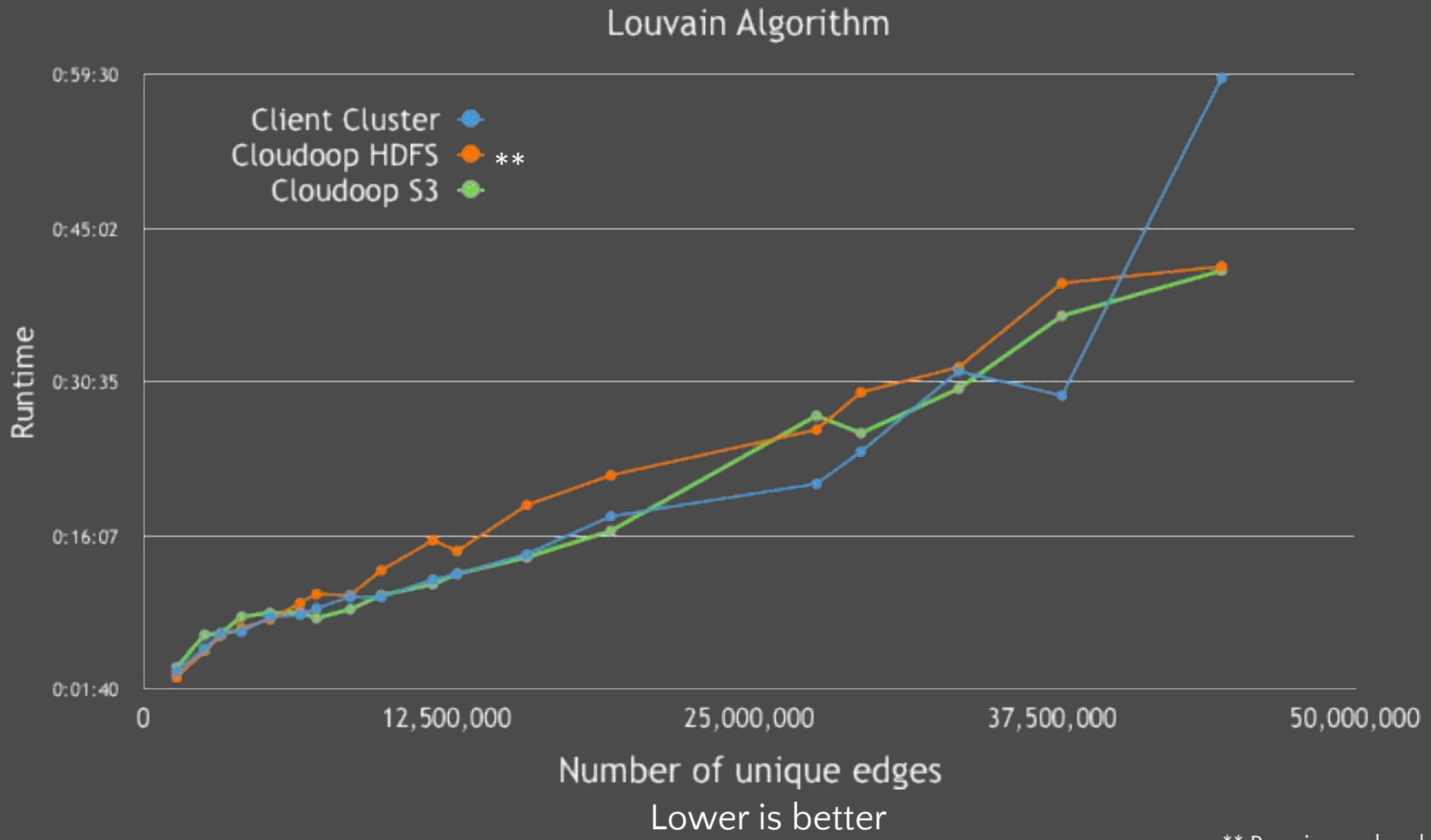
* Lower is better

TPC-DS Benchmarks



Lower is better *

Testimonials



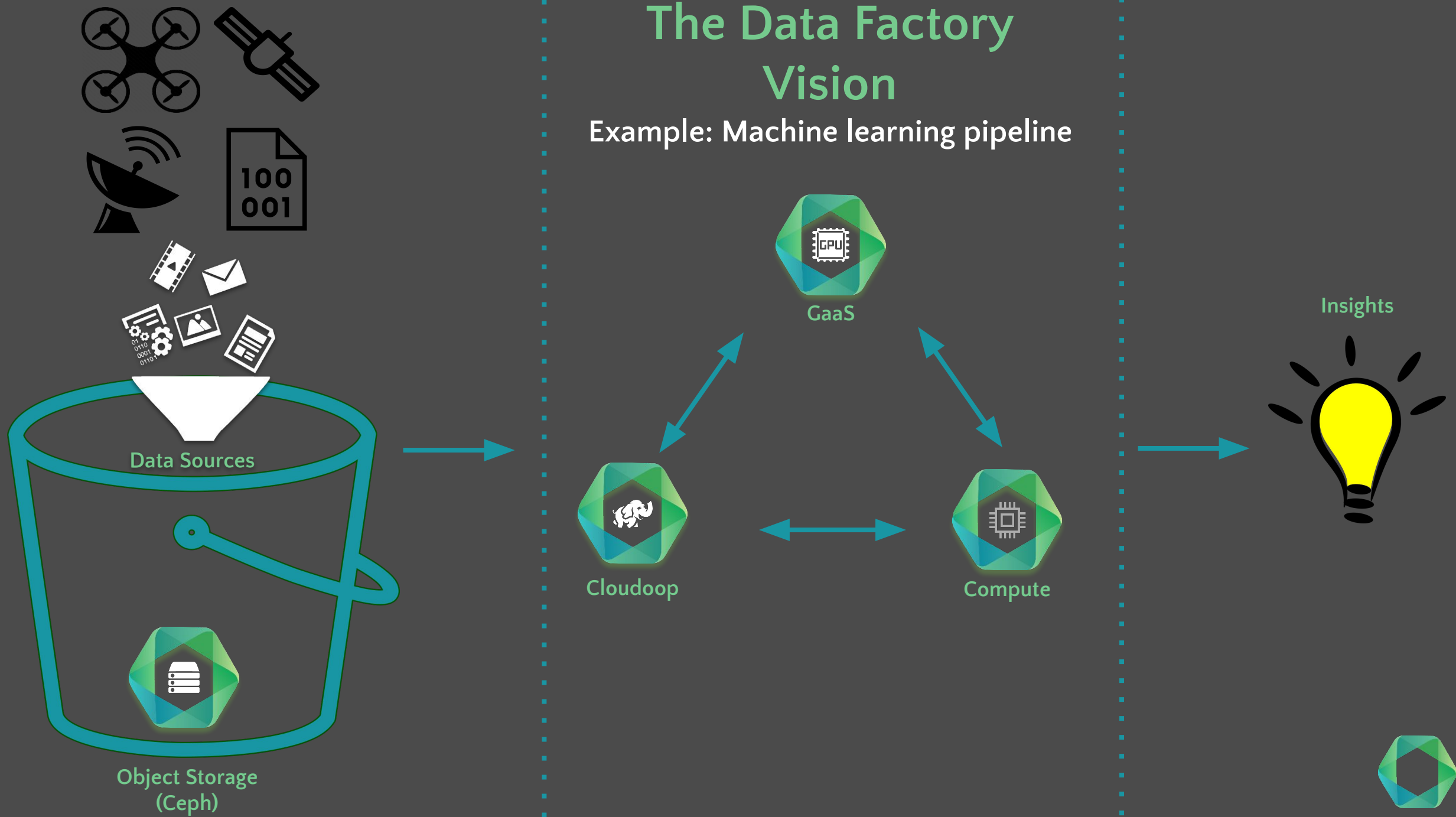
** Running on local SSDs

Cloudoop – Phase 1

- Versions
  - CDH 5.13
  - Spark 2.2
- APIs
  - Create cluster
  - Delete cluster
  - List clusters
  - Describe cluster
- Fine tuned cluster with full management access

The Data Factory Vision

Example: Machine learning pipeline



Strategic partnership

FIND US AT RED HAT SUMMIT

redhat.com/storage

@redhatstorage

redhatstorage.redhat.com

- At the Storage lockers
- At the Red Hat booth
- At one of Storage dedicated sessions (red.ht/storageatsummit)
- At the Community Happy Hour (Tues 6:30, Harpoon Brewery)
- At the Hybrid Cloud Party (Wed, 7:30, “Committee” restaurant)

Red Hat OpenShift Container Storage
red.ht/videos-RHOCS

Red Hat data analytics infrastructure solution
red.ht/videos-RHDAIS

Red Hat Hyperconverged Infrastructure
red.ht/videos-RHHI

RED HAT SUMMIT

THANK YOU

linkedin.com/company/Red-Hat

facebook.com/RedHatinc

youtube.com/user/RedHatVideos

twitter.com/RedHat