cloudera

Hadoop 3 is coming – what's new and what's next?

About Wei-Chiu

Apache Hadoop committer
Software Engineer, Cloudera

cloudera®

Agenda

The Problem
What is Hadoop
Major Hadoop 3 Features
What's Next?

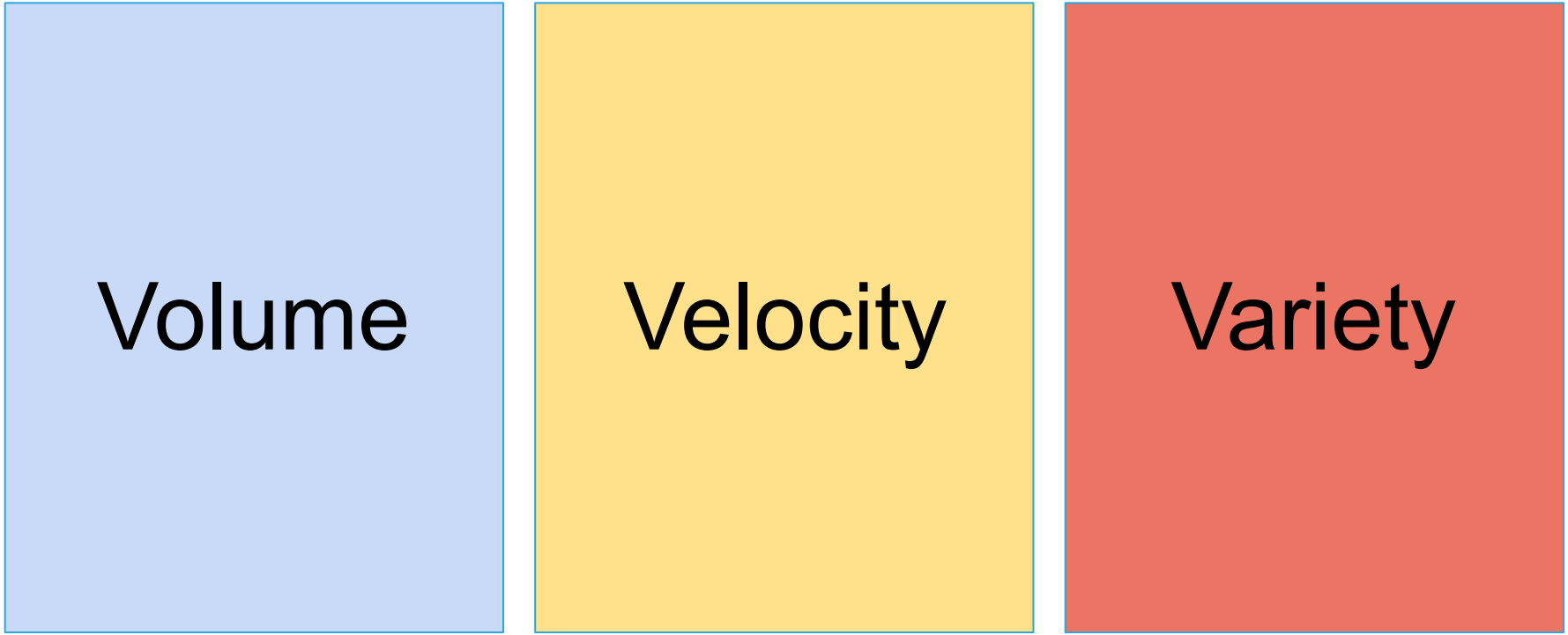
cloudera

“Data helps solve problems”

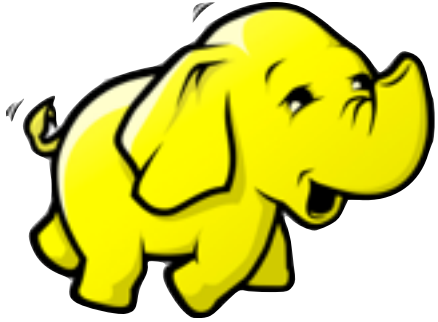
- Anne Wojcicki

cloudera

Big Data - 3Vs



Apache Hadoop



The de facto Big Data Analytics platform

A distributed framework to support large scale computation on commodity hardware

- Petabyte+ storage, 1,000+ compute nodes
- Inspired by Google
- Originally developed by Yahoo!, donated to Apache Software Foundation.
- Open source :)
- 183 committers, thousands contributors

Apache Hadoop

cloudera

Cloudera

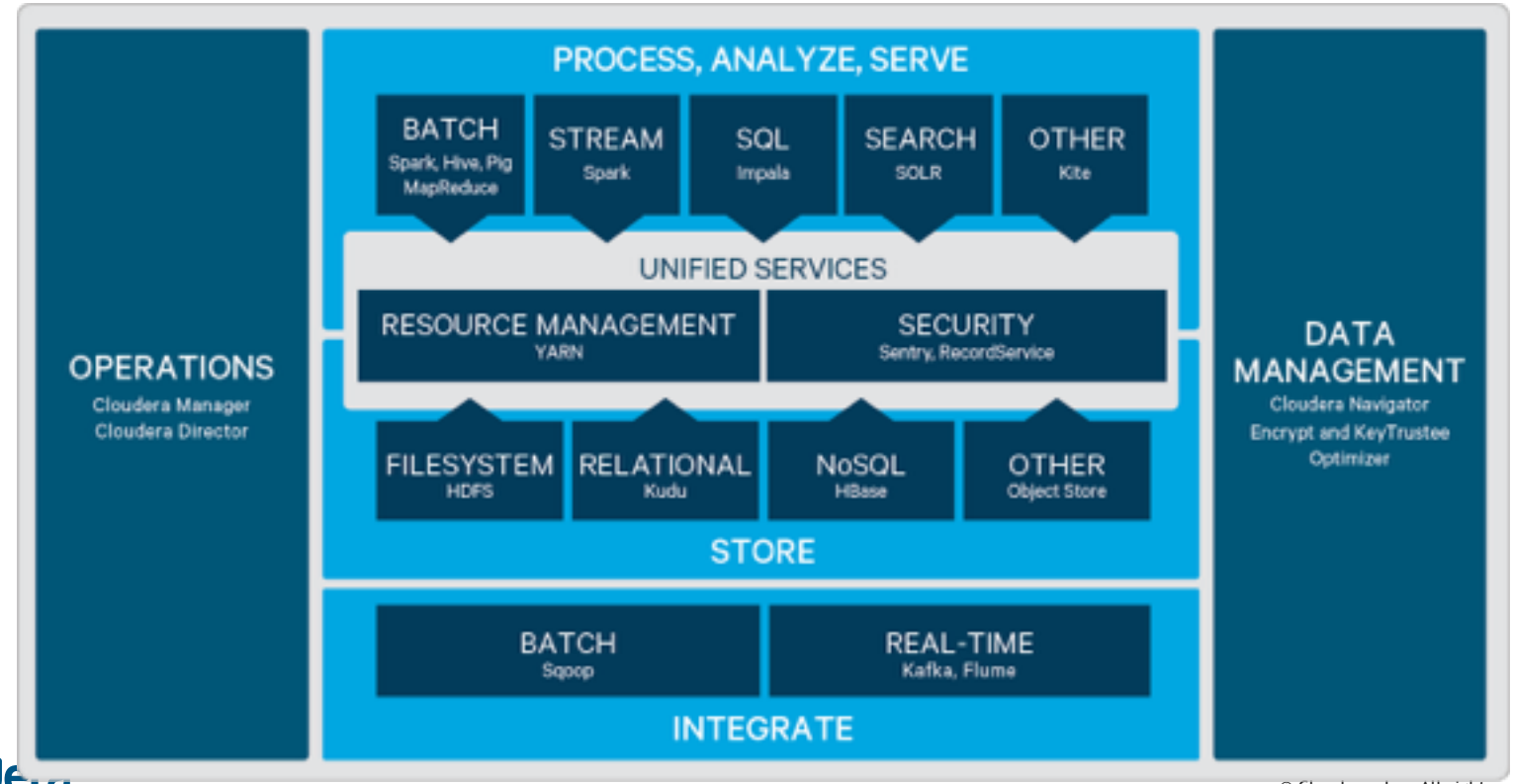
Commercializes Hadoop* technology
Open source, open culture
CDH - Cloudera's Distribution for Hadoop
- Platform. Open source

Cloudera Manager (CM), Cloudera Navigator, Key Trustee
- Cluster management, monitoring. Proprietary

*Hadoop and its associated projects

cloudera

Hadoop Ecosystem



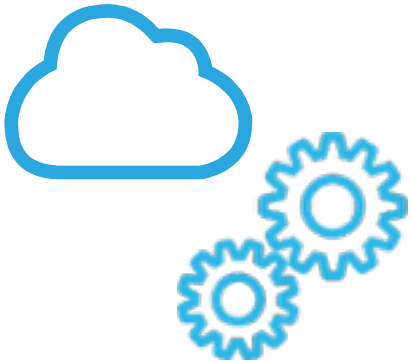
Well, data itself is a problem ...

Clusters becoming larger

Storage: reduce storage cost
Compute: much larger cluster

More enterprise adoption

High availability
High performance
Seamless experience



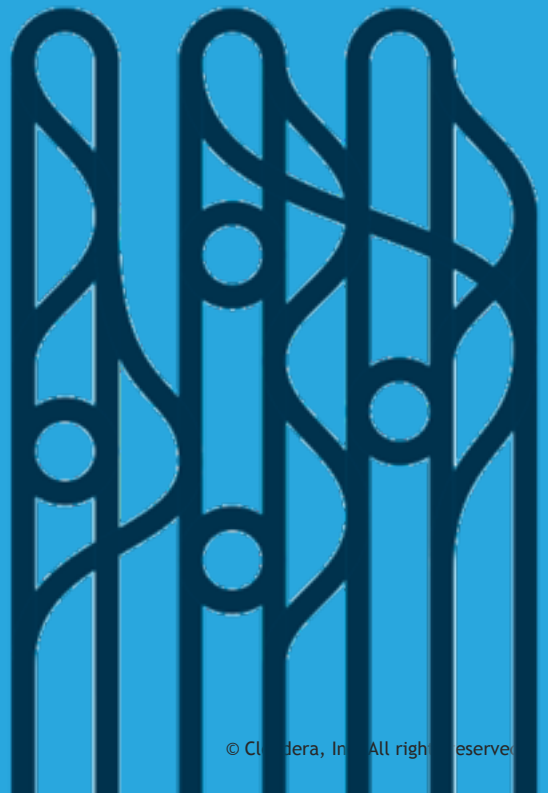
New applications

More cloud usage
Ease of development

cloudera

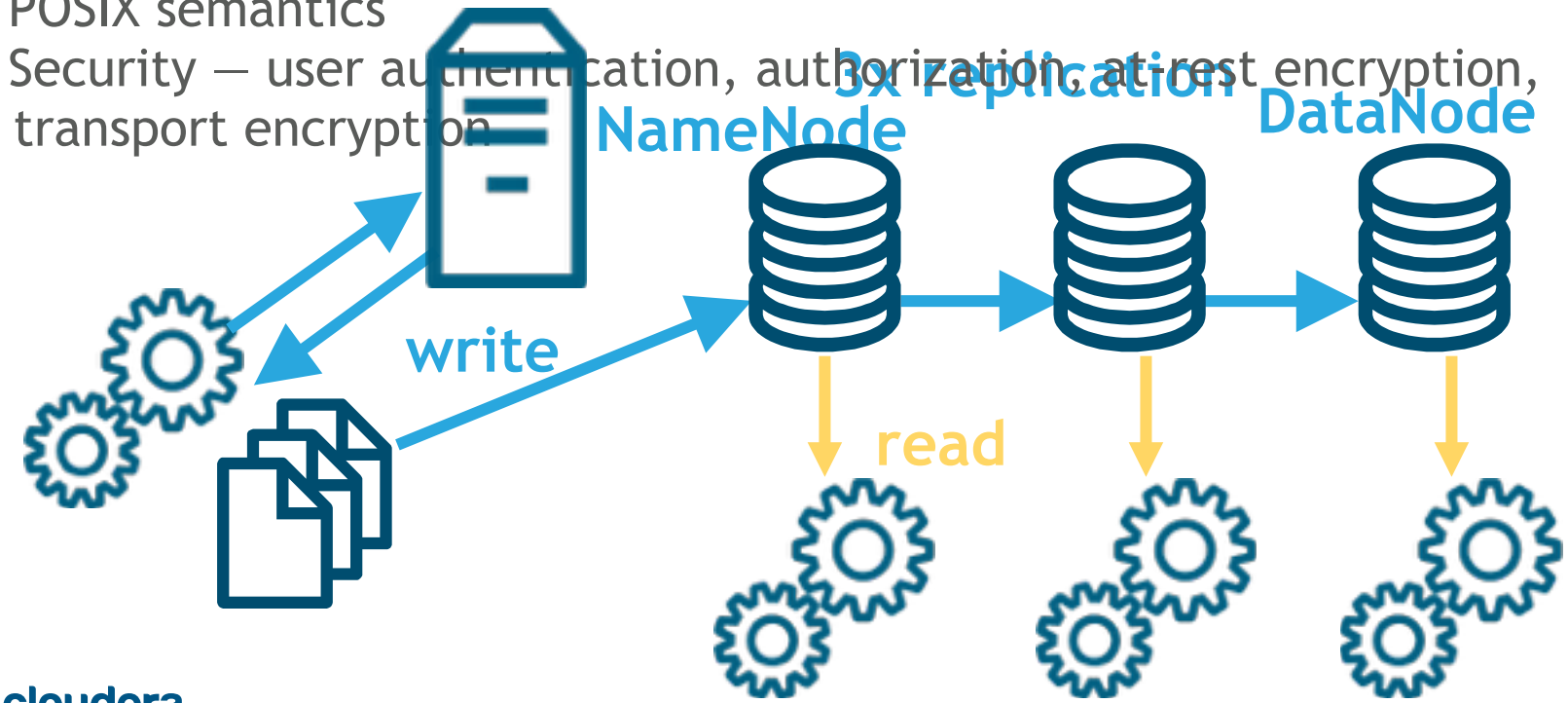
cloudera

HDFS Erasure Coding



Hadoop Distributed File System

A fault tolerant, highly scalable storage system
POSIX semantics
Security — user authentication, authorization, at-rest encryption, transport encryption



Hadoop Distributed File System

Advantage

- Failure tolerant

But

- 3x storage cost
- 3x datacenter space
- 3x power consumption

How to reduce storage overhead?

Erasure Coding 101

- Parity bit
  - XOR
  - If X is lost, X can be reconstructed using Y and X ^ Y
  - 50% overhead ((3-2)/2)
  - Can tolerate one failure
- Reed-Solomon
  - RS(k,m) tolerates m failures in k data cells.
  - XOR = RS(2,1)

X	Y	X ⊕ Y
0	0	0
0	1	1
1	0	1
1	1	0

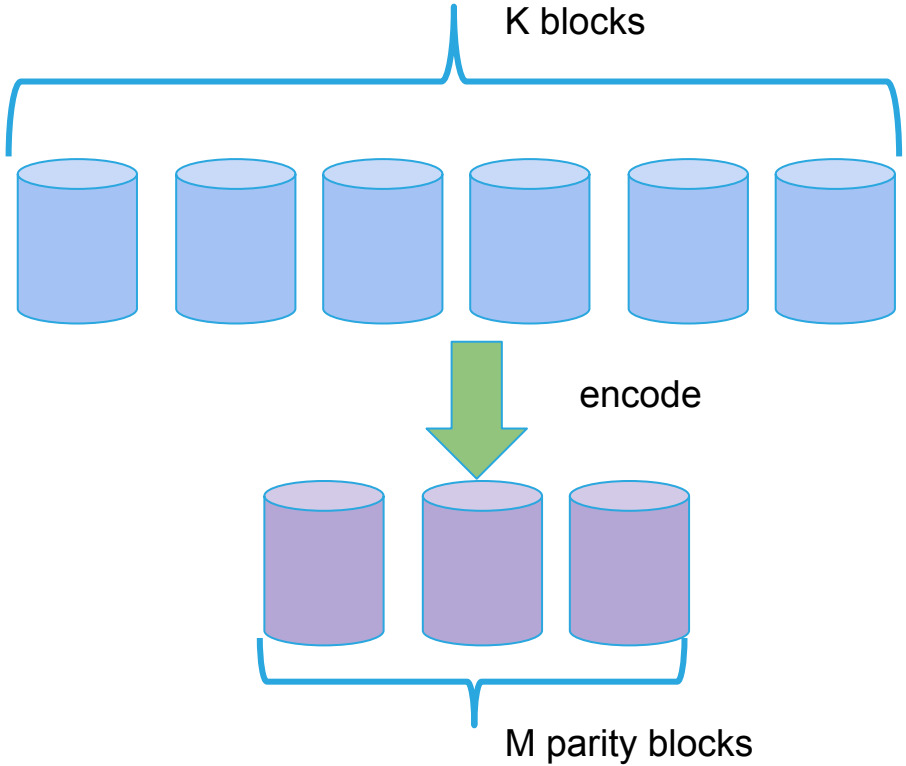
Erasure Coding 101

Reed-Solomon

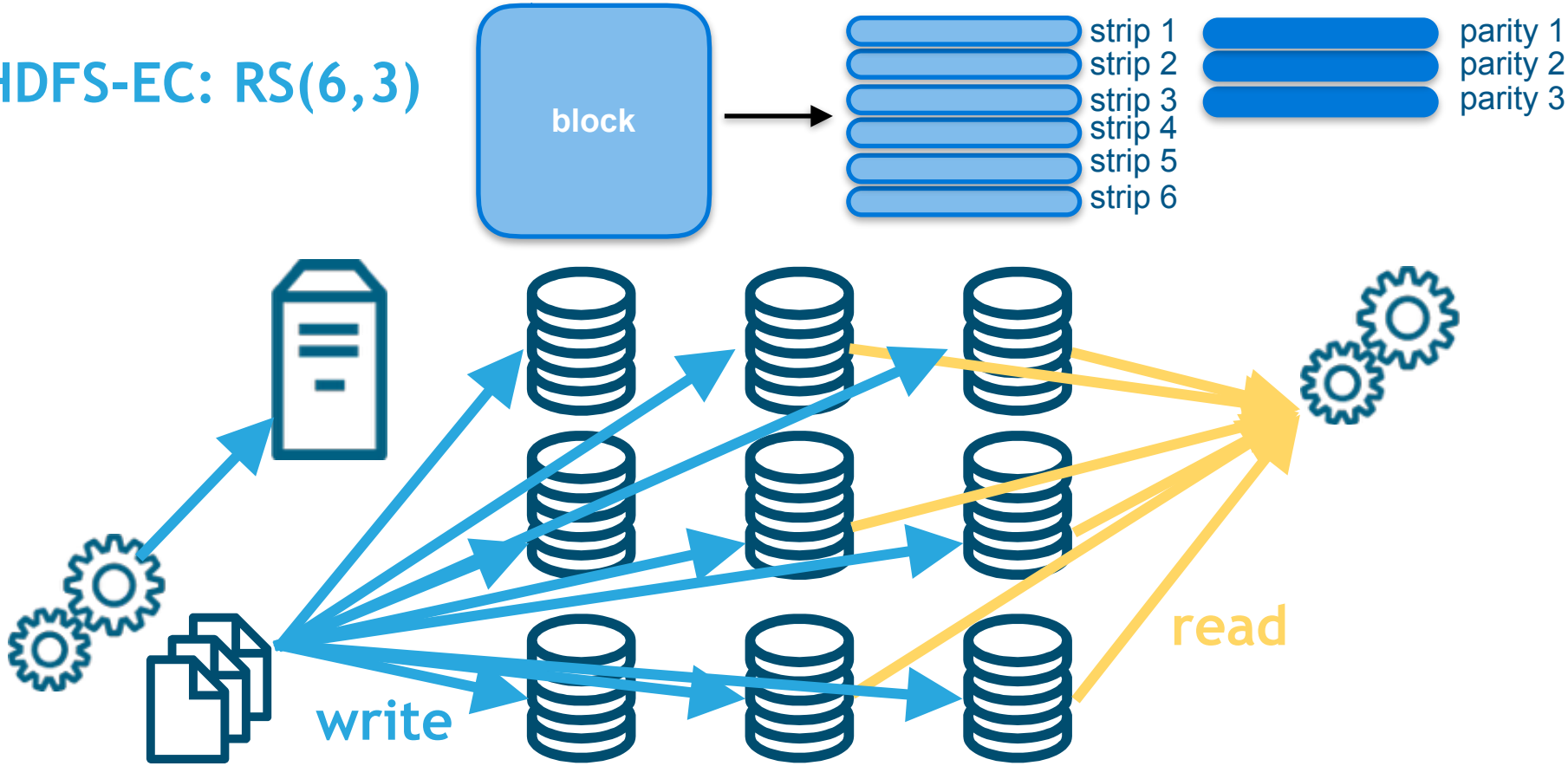
- Compute parity bits for redundancy
- Blocks can be reconstructed after failures
- Configurable durability v.s. storage overhead

RS(6,3)

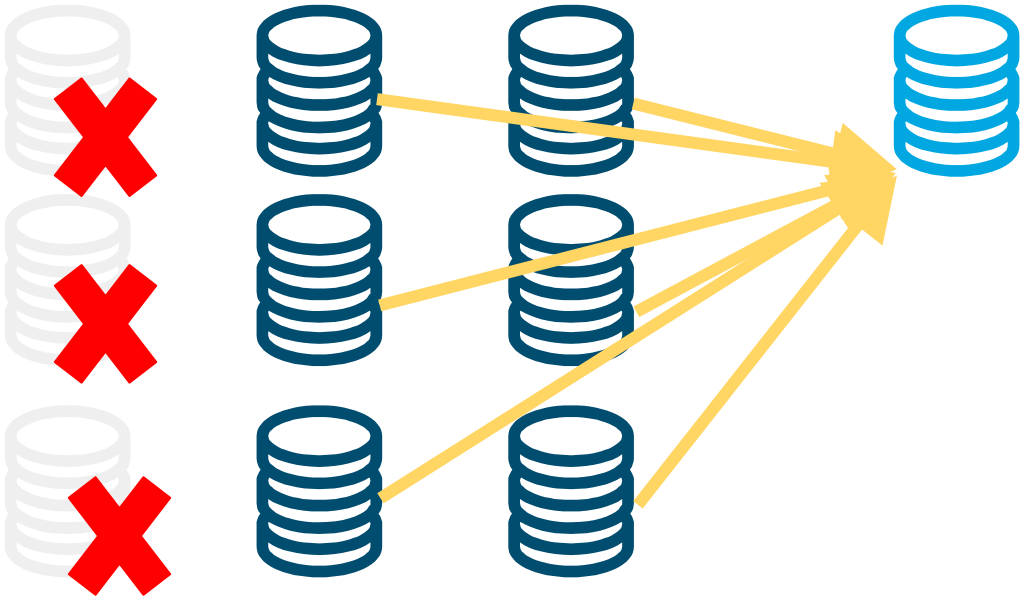
- = 50% storage overhead
- (9-6)/6

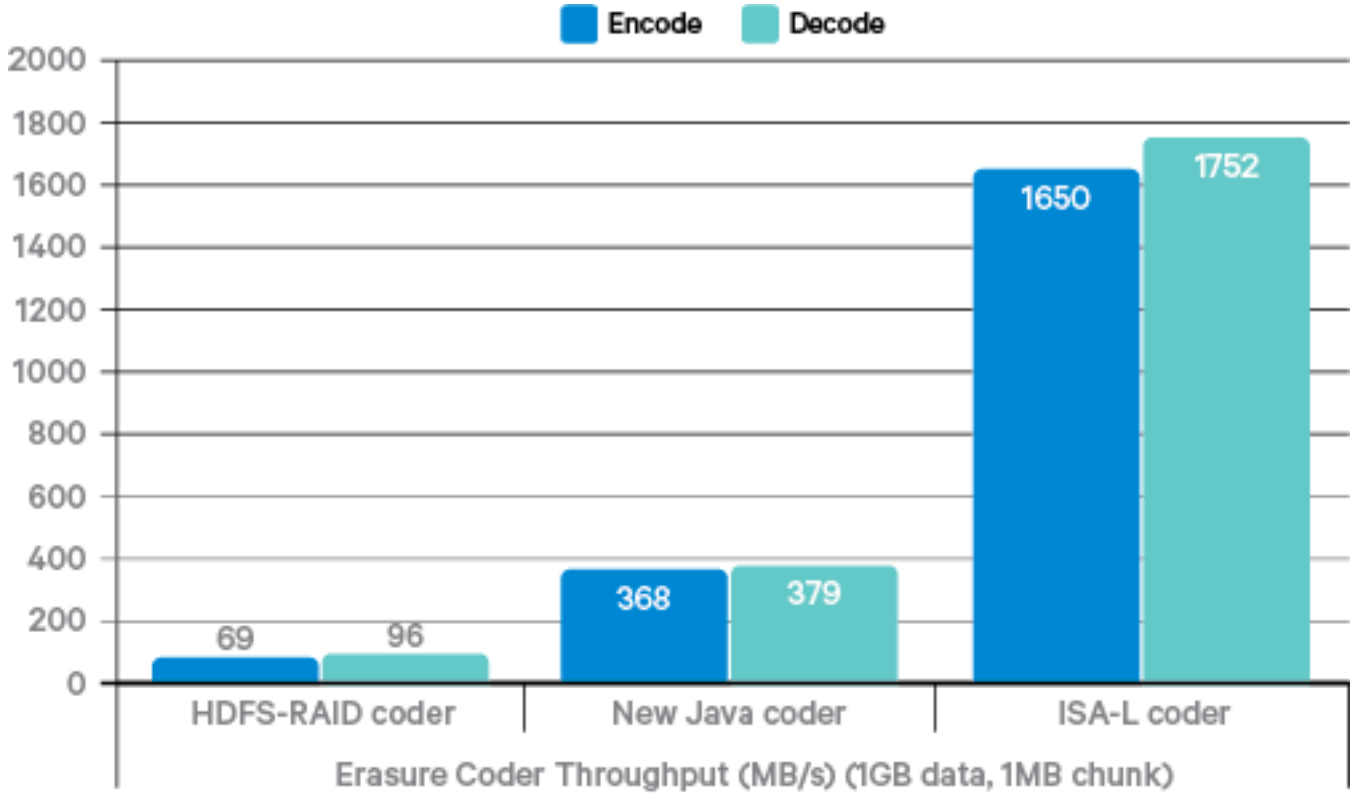


HDFS-EC: RS(6,3)



HDFS-EC: Failure Handling

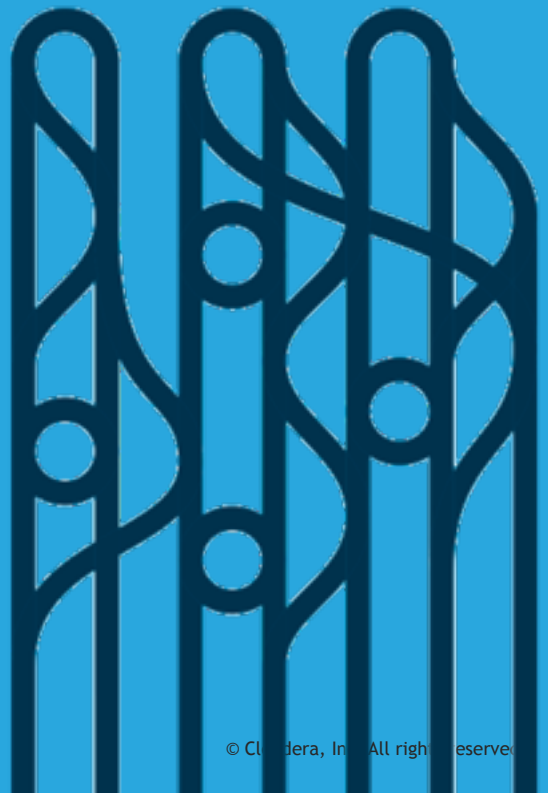




cloudera

cloudera

YARN Federation

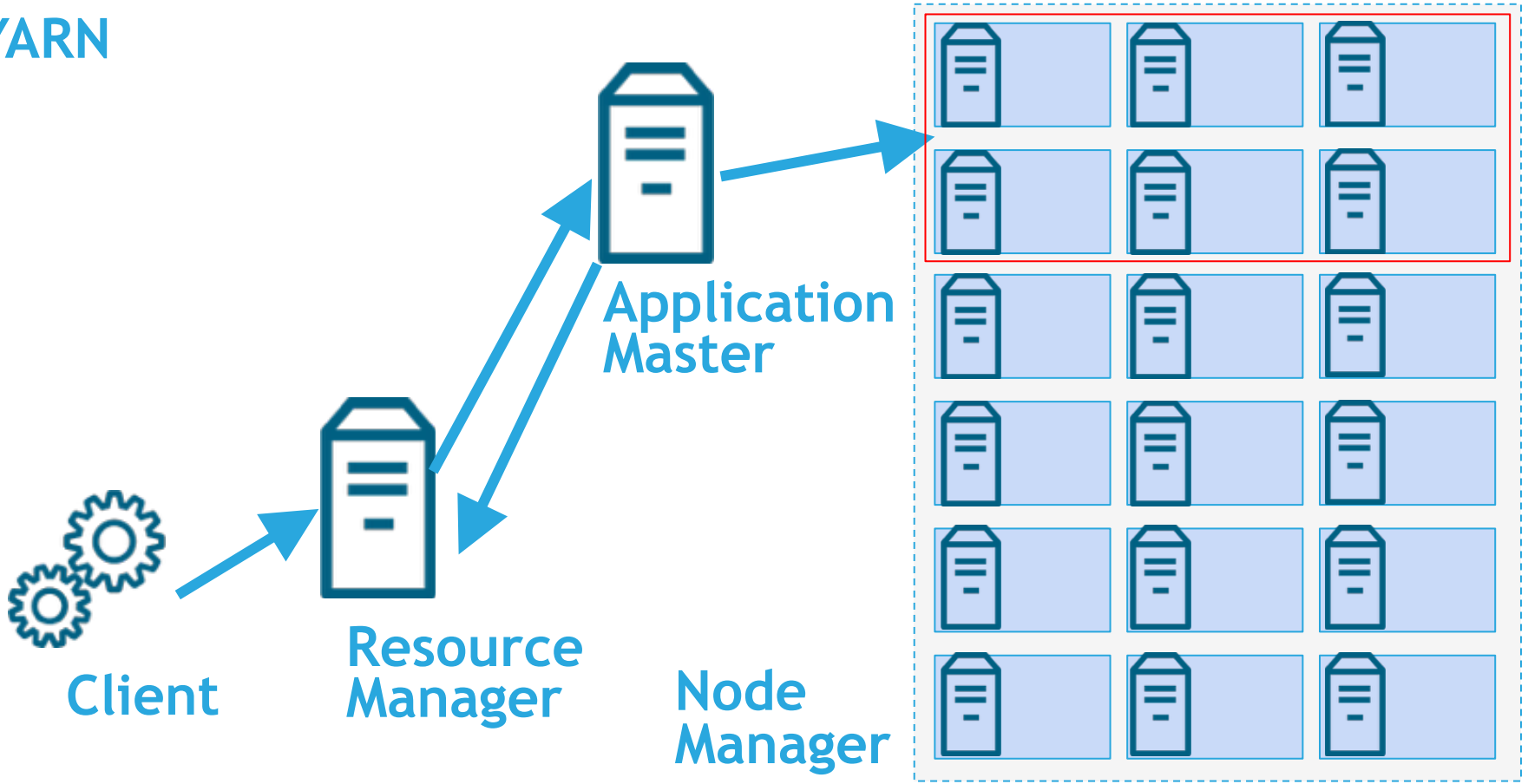


YARN

A resource management framework for Hadoop clusters

- Highly scalable, 4000 - 8000 nodes in production
- Hive, Oozie, Spark, …
- HBase

YARN

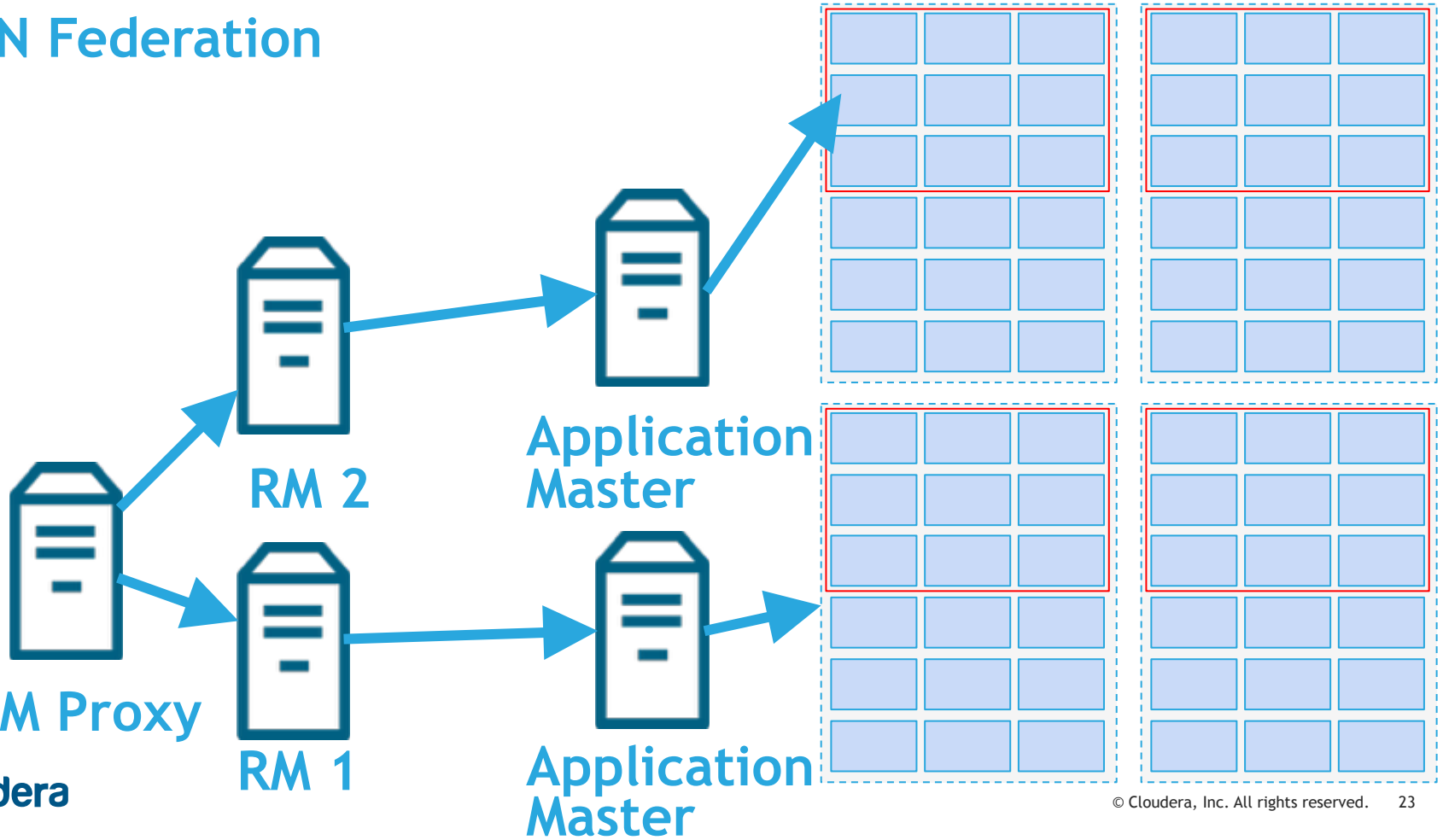


YARN Federation

Developed by Microsoft
Extreme scale

- 100,000 compute nodes
- Resource Manager becomes the bottleneck

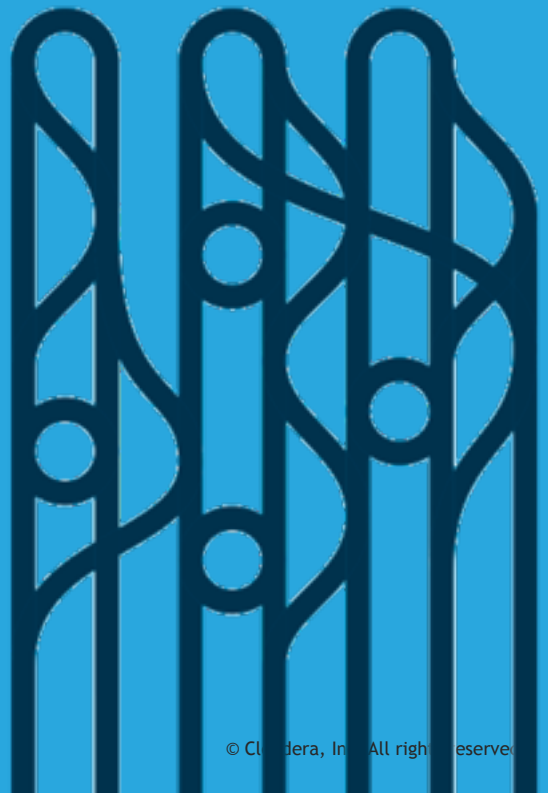
YARN Federation



cloudera

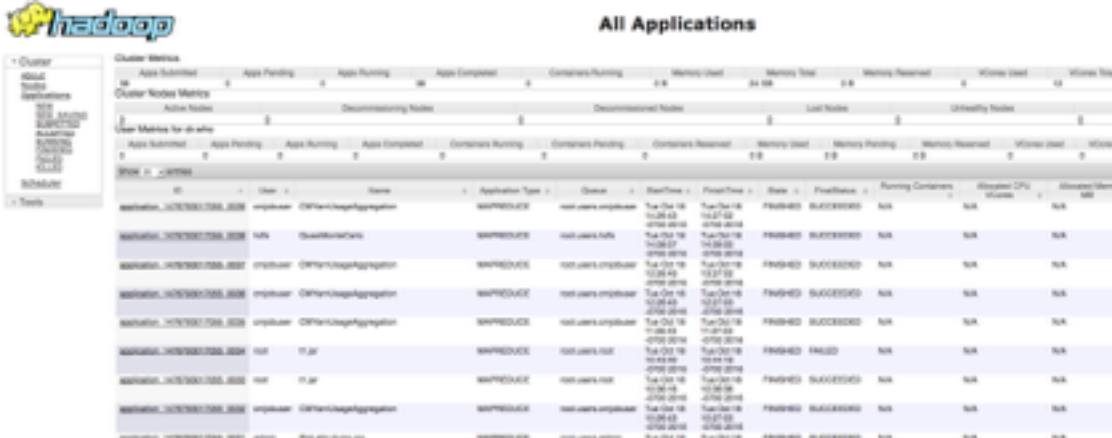
cloudera

YARN Timeline Service v2



Job History Server

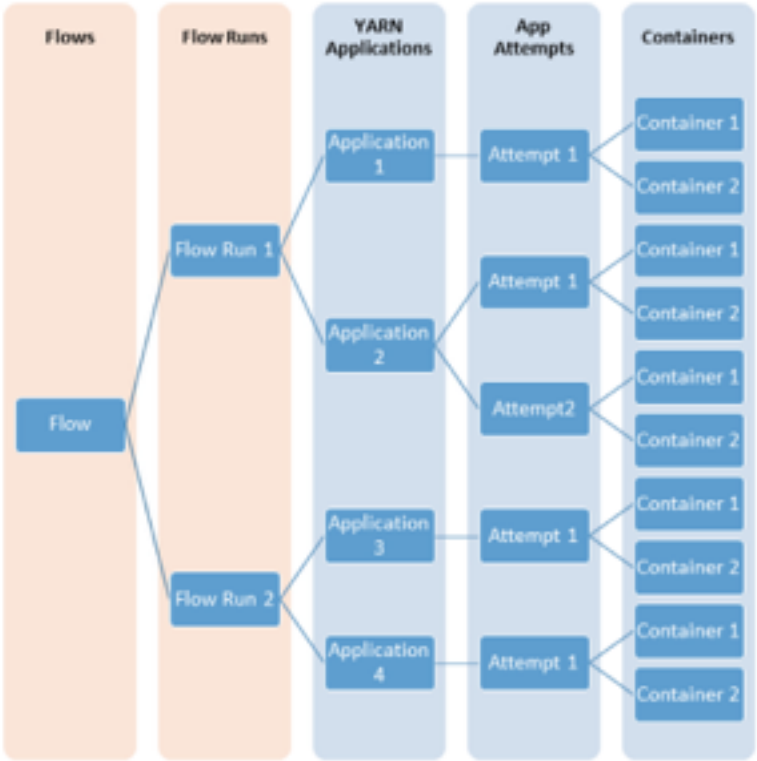
- Keeps track of job progress
  - Collect or retrieve information of MapReduce jobs
- Extensibility
  - MR only
- Usability
  - No YARN level events
  - Metrics can only be retrieved after job terminates



cloudera

Application Timeline server v2

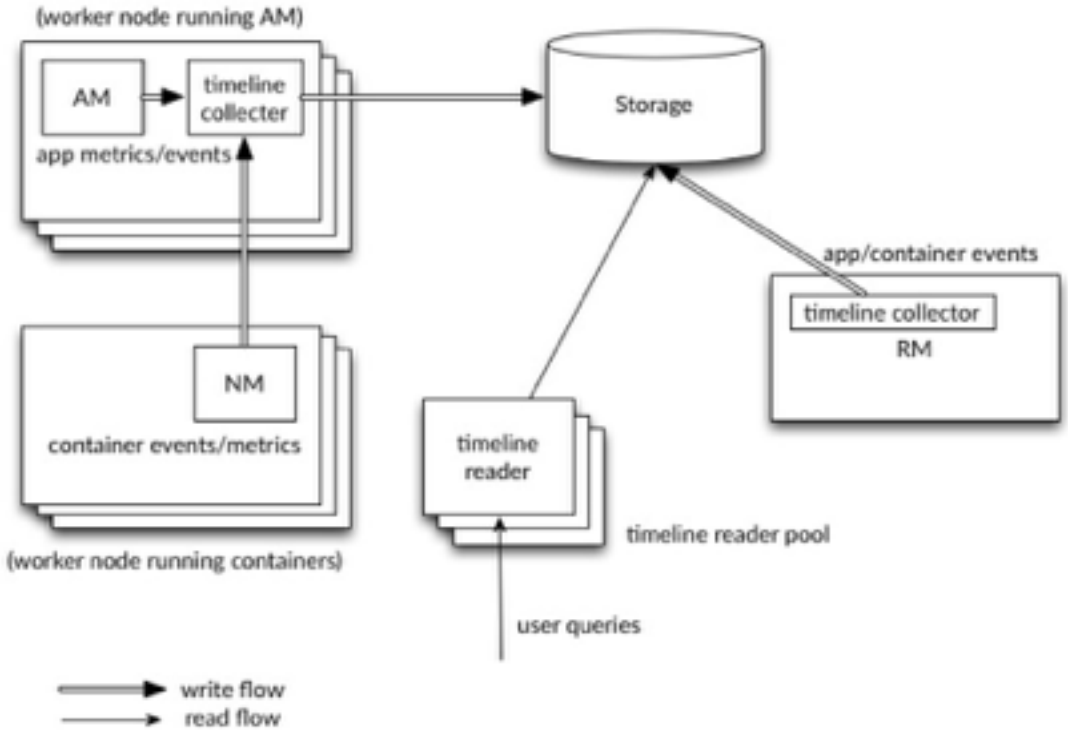
- Development led by Twitter
- Usability
  - Flow: logical group of applications
- Scalability
  - HBase
- Use cases
  - Analyze application performance.
  - Cluster capacity planning.



ASTv2 Architecture

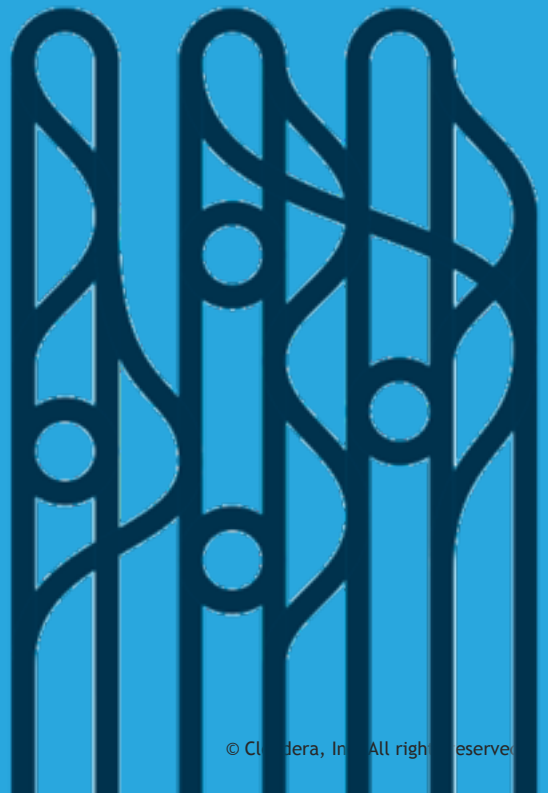
Use HBase for storage
Use cases:

- Analyze application performance.
- Cluster capacity planning.

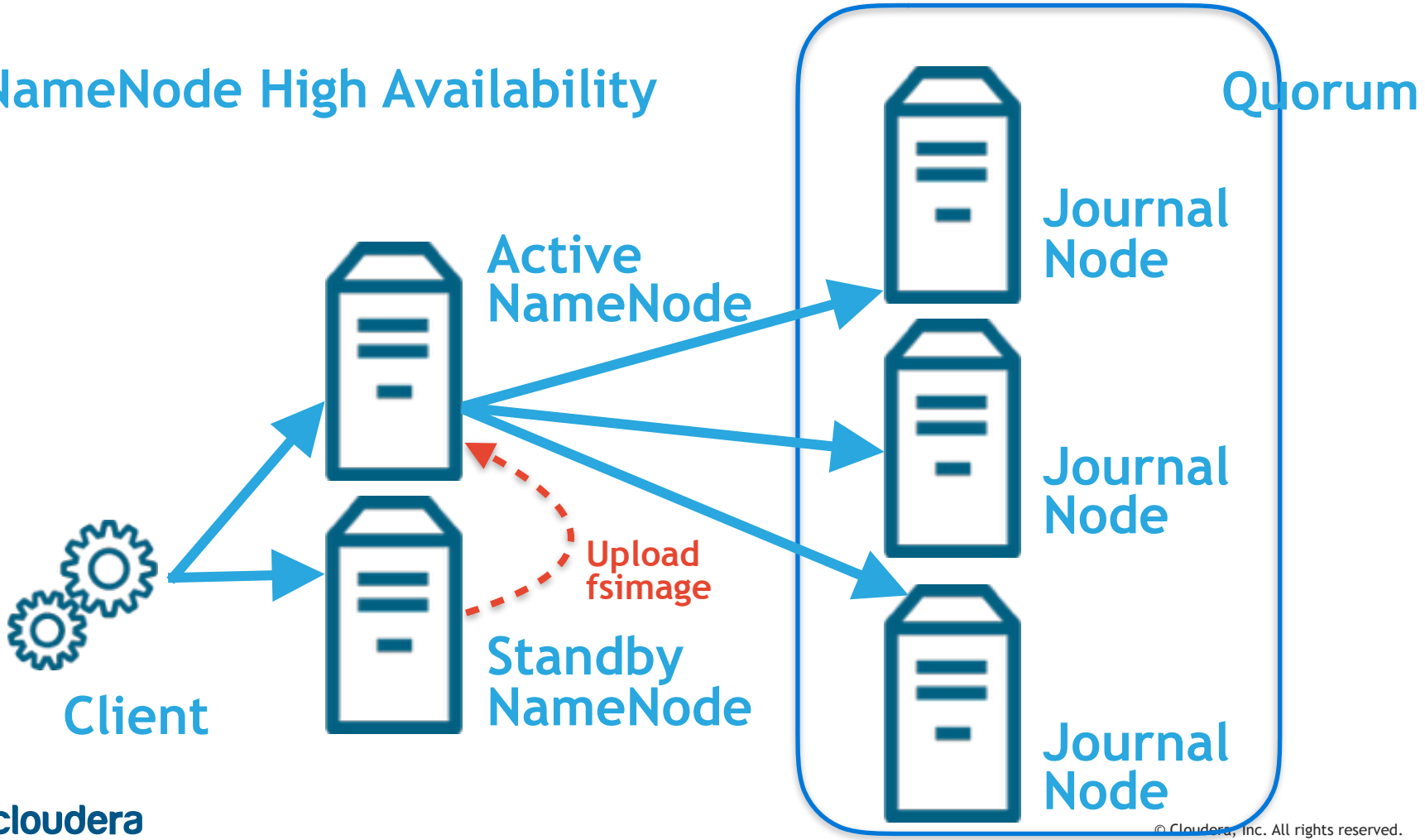


cloudera

HDFS Multi Standby NameNodes

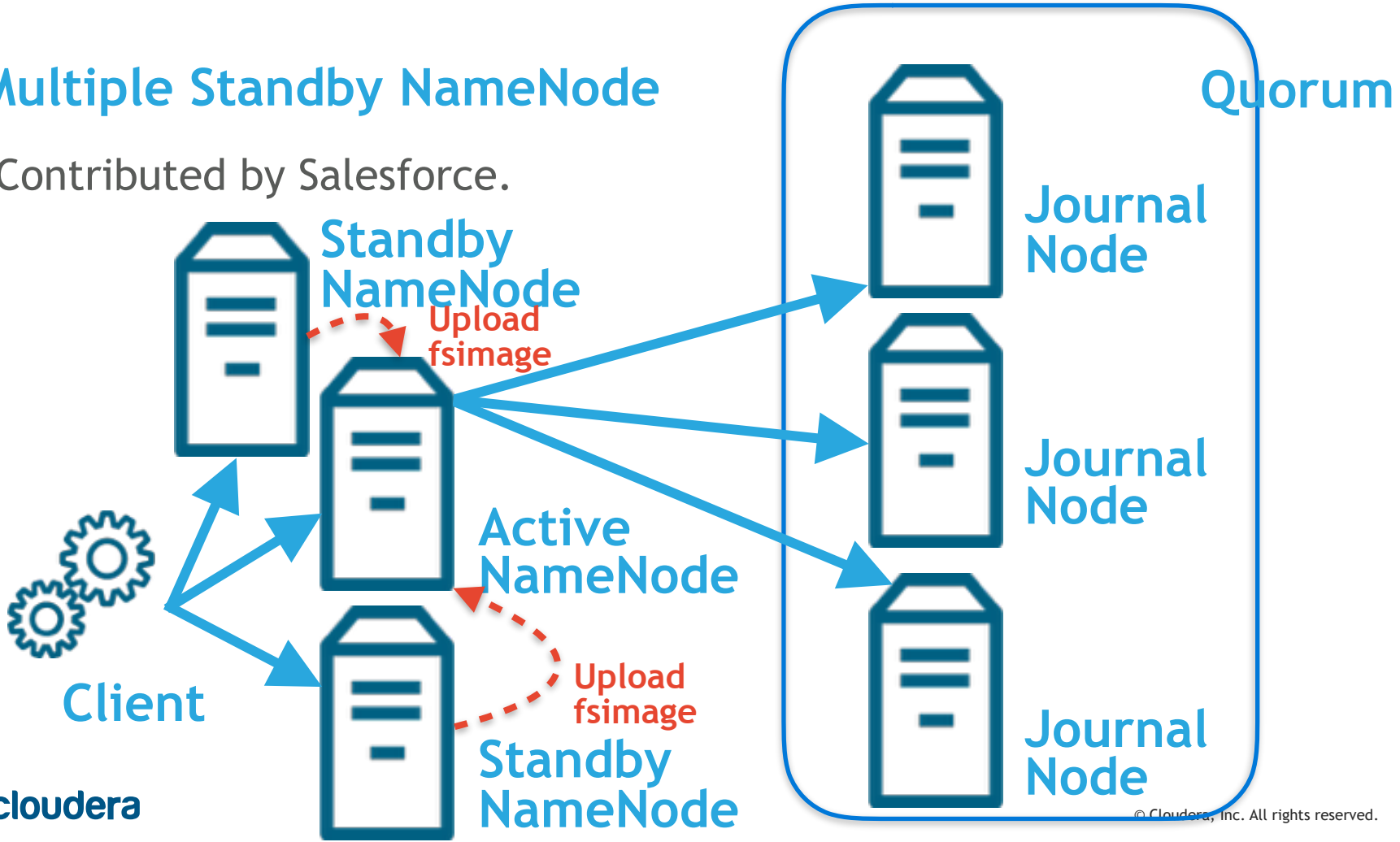


NameNode High Availability



Multiple Standby NameNode

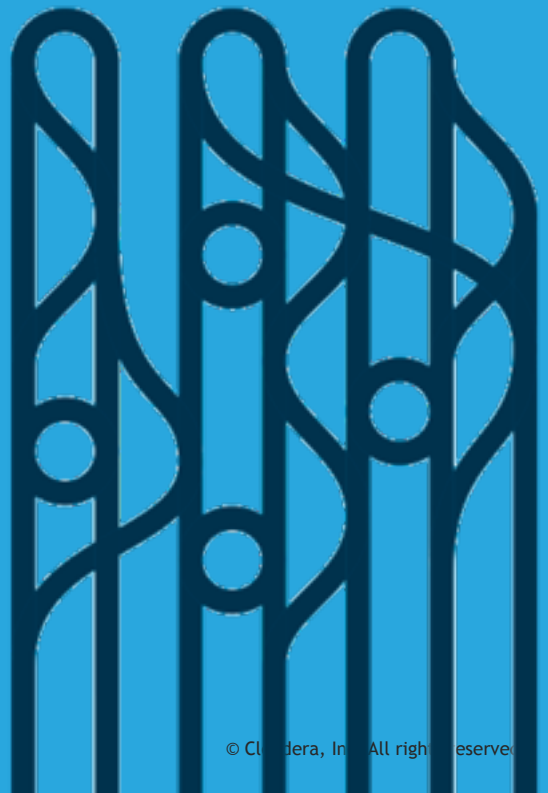
Contributed by Salesforce.



cloudera

cloudera

Classpath Isolation



Dependency Hell

cloudera

Dependency Hell

Hadoop was not initially designed as foundation of many applications.

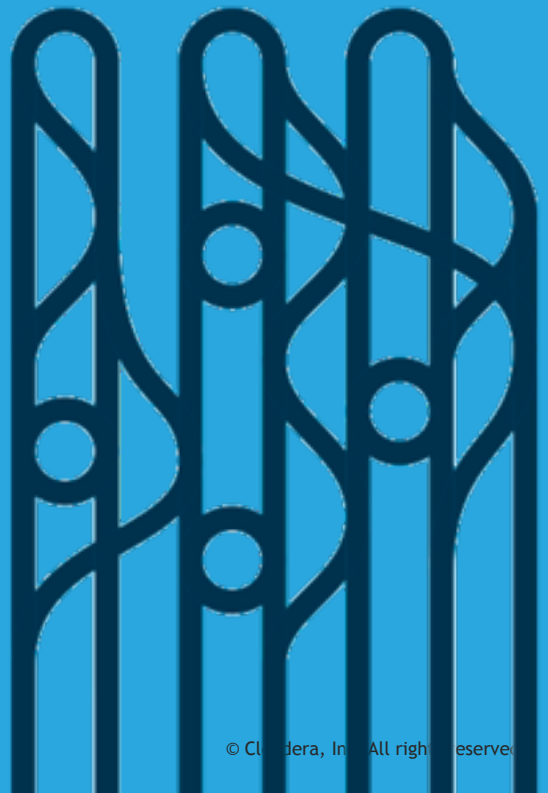
- More applications depending on Hadoop
- harder for Hadoop to upgrade dependency libraries.
- Potential risk to break existing applications
- Increase exposure to security vulnerabilities

Classpath Isolation

- Separate client-side classpath from server-side

cloudera

Cloud

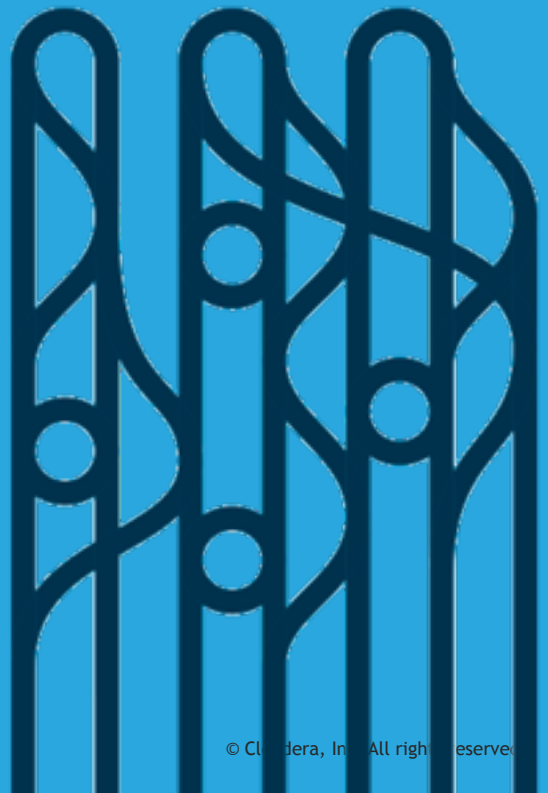


Other features

- Cloud connectors
  - Microsoft Azure Data Lake filesystem
  - Aliyun Object Storage Service

cloudera

Misc.

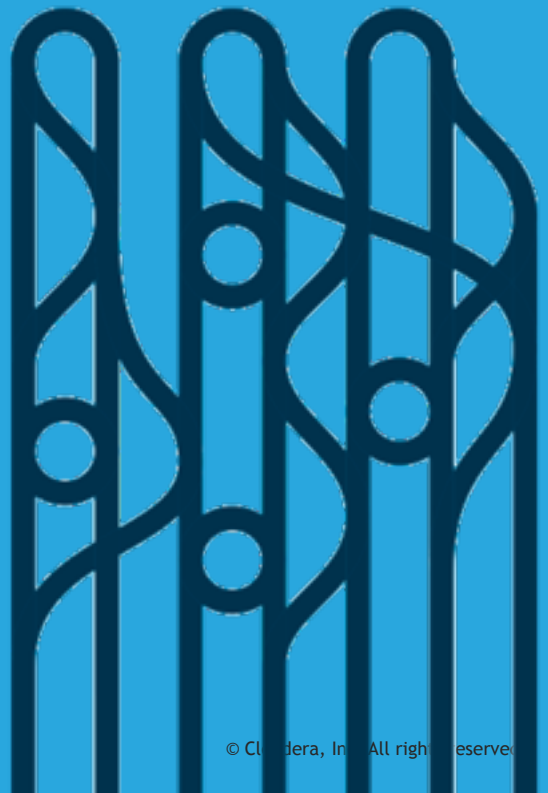


Other features and incompatibility

- Shell script rewrite
- Requires Java 8
- Server ports
- Remove legacy features
  - S3 file system → S3A (recommended) or S3N
  - Hftp → webhdfs/httpfs
  - Bookkeeper Journal Manager → Quorum Journal Manager

cloudera

What's next?

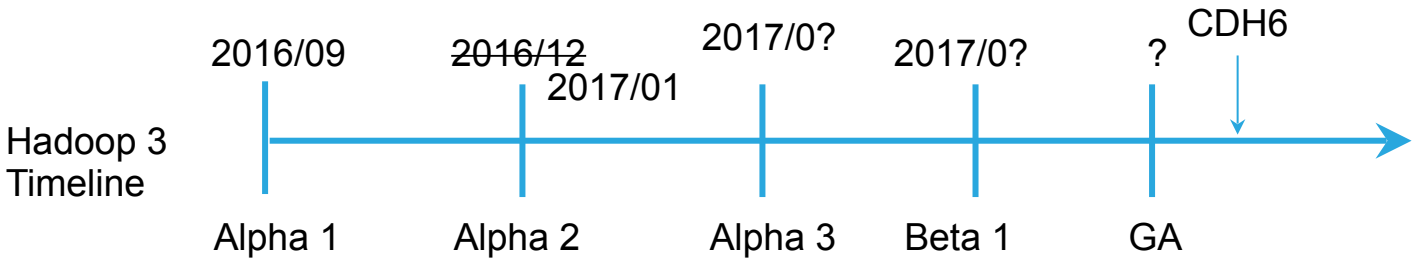


Now what?

Developers
- Use it early, test it early and file bug reports.

Administrators
- Test upgradability

Users
- Expect better user experience.



Future? Hadoop 4?

- We don’t know yet.
- Ozone (HDFS-7240)
  - Object store for HDFS
- HDFS over cloud (HDFS-9806)
- Emerging applications and use cases
  - Docker
  - Deep learning
- Hardware Trend
  - Cloud storage
  - Faster ethernet (40GBps), high density (> 100TB) storage node
  - Memory technology
  - Locality will not be a deciding factor.

Ozone (HDFS-7240)

Status quo

- NameNode is becoming a bottleneck
- A general file system may not suit the specific need of an application

Solution

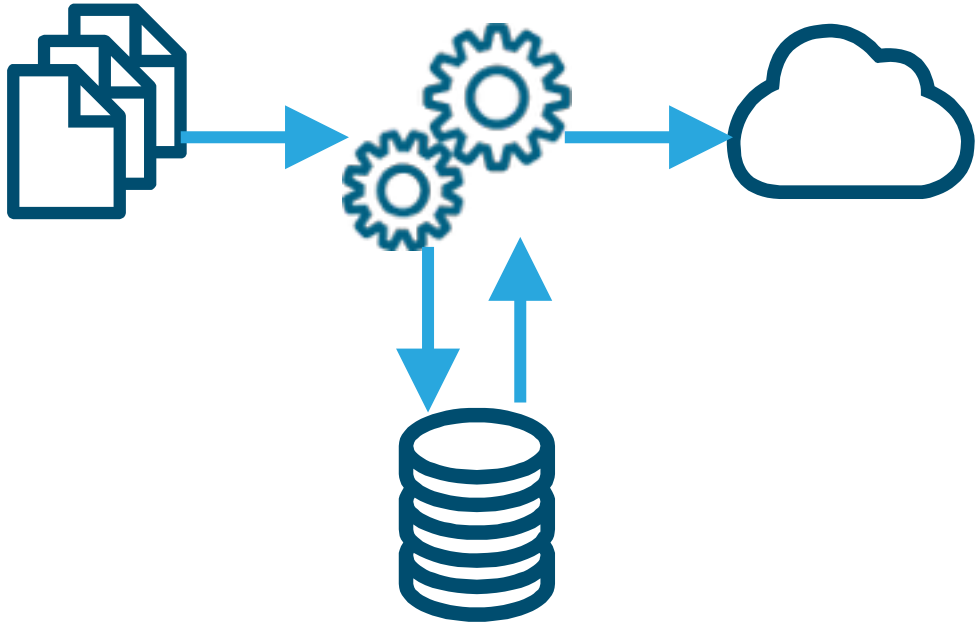
- Split HDFS namespace into blob stores

HDFS over Cloud (HDFS-9806)

Use case
- Use HDFS for temporary data
- Use cloud for permanent storage

The problem
- Data management
- Consistency

Solution
- HDFS as metastore and cache
- Cloud as backend data store



cloudera

cloudera

Ask Bigger Questions

References

- Introduction to HDFS Erasure Coding in Apache Hadoop
- Enable YARN RM scale out via federation using multiple RM's
- Application Timeline Server - Past, Present and Future
- HDFS-6440 Support more than 2 NameNodes
- How-to: Use the New HDFS Intra-DataNode Disk Balancer in Apache Hadoop