



The **LiveData** Company

WANdisco FUSION & A LIVE DATA PLATFORM

Technical Product Overview

2018

WANdisco Confidential and Proprietary

© WANdisco 2018 | wandisco.com



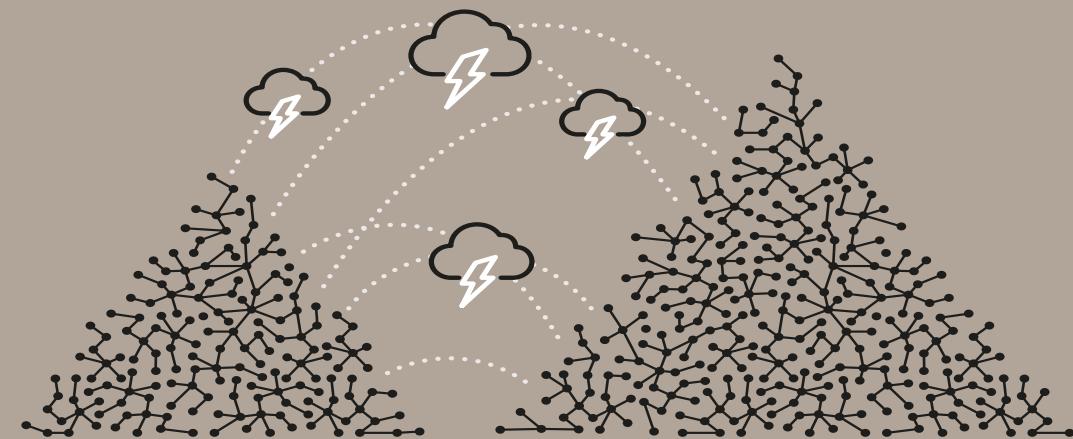
THE DATA REPLICATION PROBLEM

Keeping data replicated and consistent across these different solutions is hard

Small (gigabyte) scale data set
= “acceptable” outage (15 minutes)



Large (terabyte, petabyte, exabyte) scale data set
= “unacceptable” outage (days/weeks/months)



HYBRID CLOUD REQUIRES NEW REPLICATION TECHNOLOGY



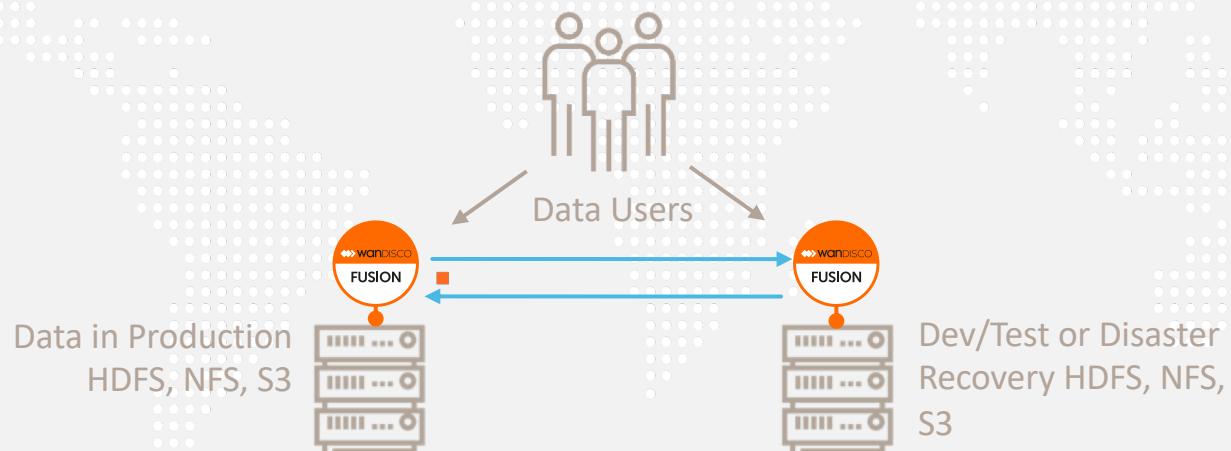
WANdisco FUSION

WANdisco Fusion is **unparalleled, omni directional replication technology** that gives you LiveData - consistent data everywhere, spanning platforms and locations, even for changing data at petabyte scale.

LiveData
CRITICAL DATA ALWAYS AVAILABLE



LiveData all the time, everywhere

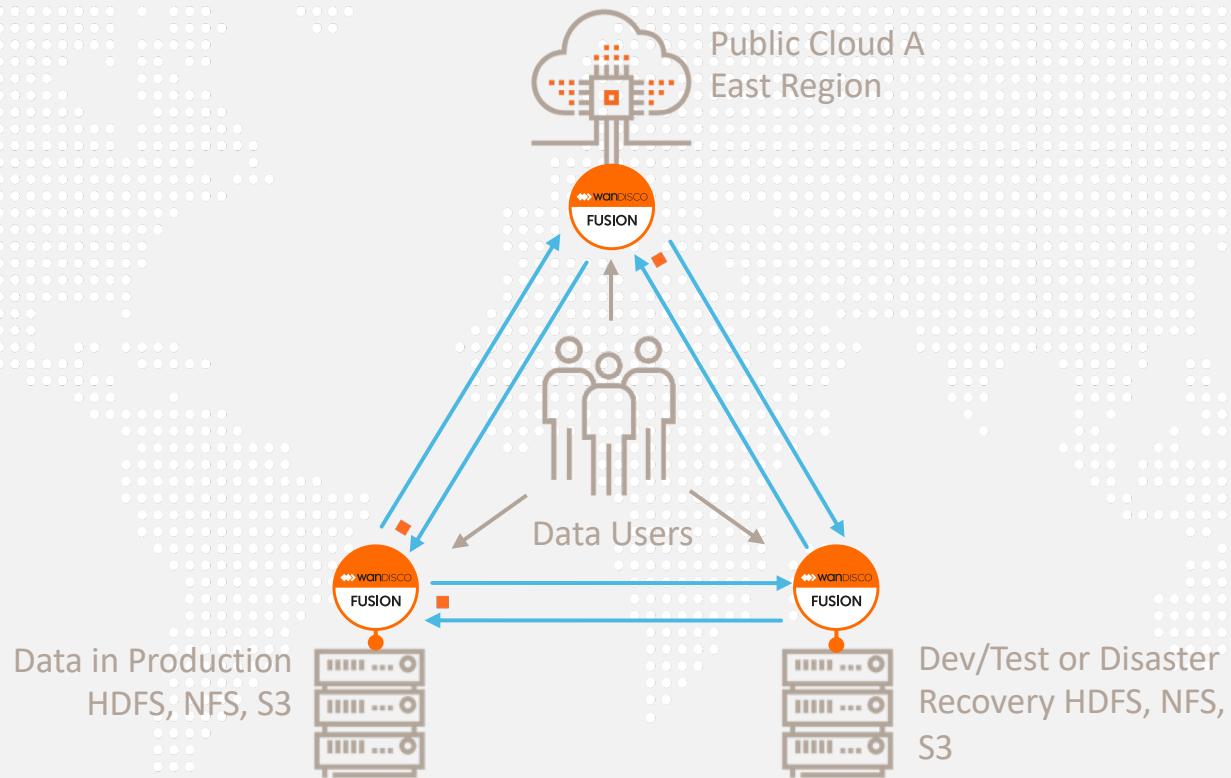


WANdisco Confidential and Proprietary

© WANdisco 2018 | wandisco.com



LiveData all the time, everywhere

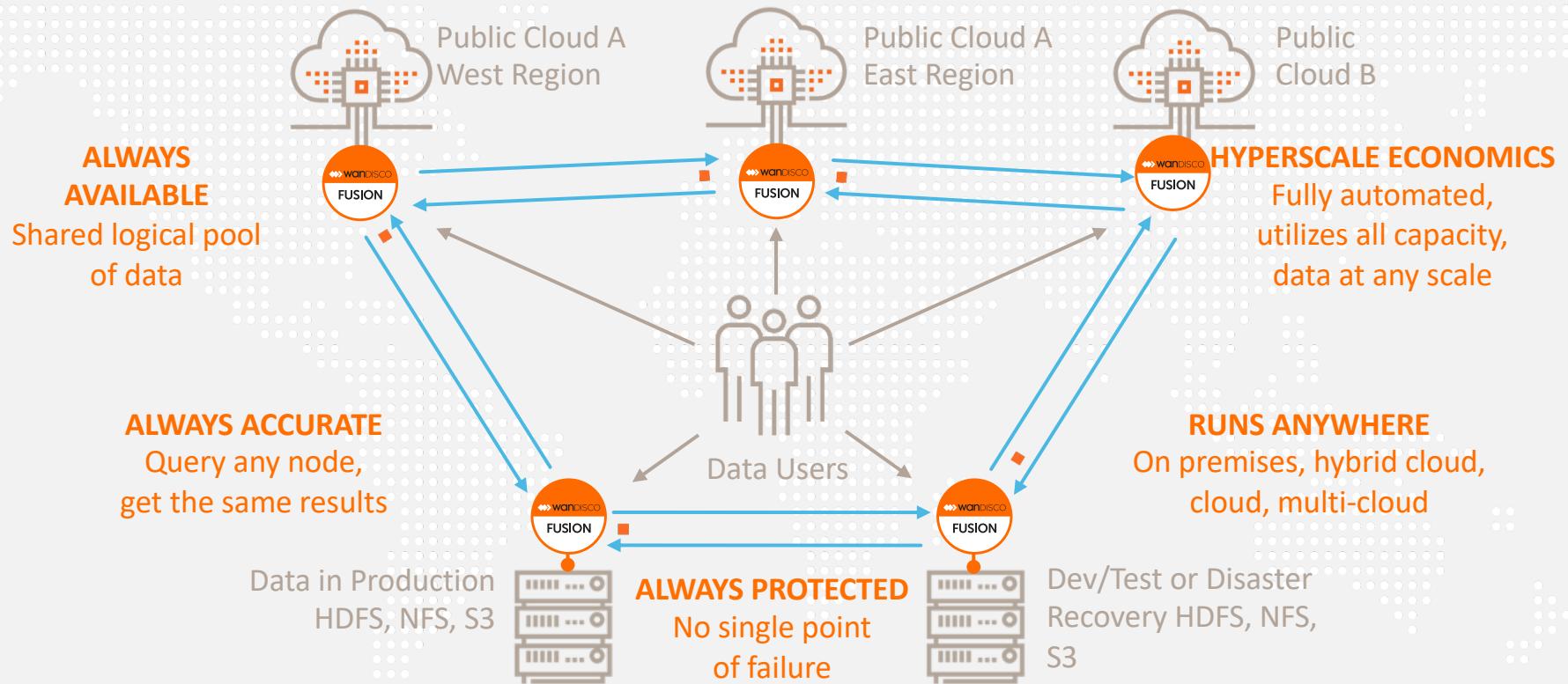


WANdisco Confidential and Proprietary

© WANdisco 2018 | wandisco.com

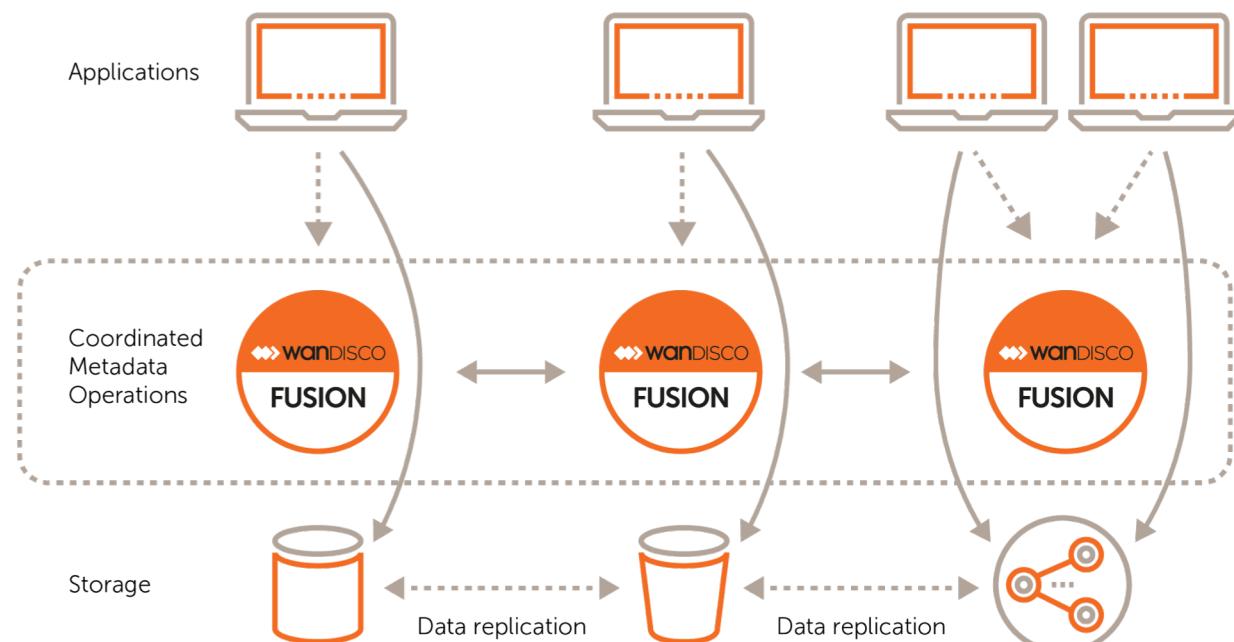


LiveData all the time, everywhere





HOW WANDISCO FUSION WORKS



WANdisco Confidential and Proprietary

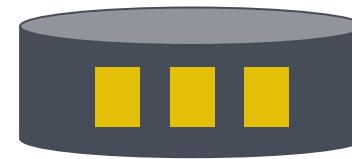
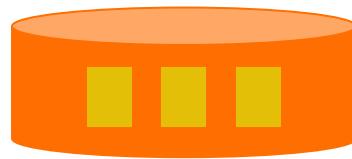
© WANdisco 2018 | wandisco.com





WANDISCO FUSION DATA FLOW

Create/Append



WANdisco Confidential and Proprietary

© WANdisco 2018 | wandisco.com

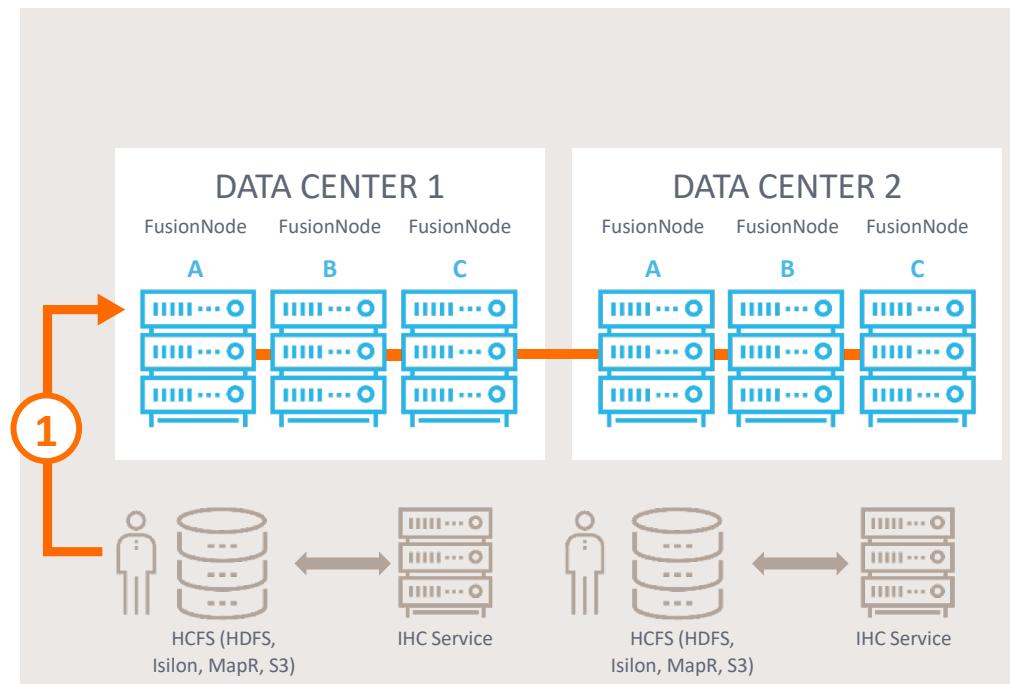
WANdisco FUSION WORKFLOW



HDFS  HDFS

Fusion workflow

1. User makes a request to create / change a file



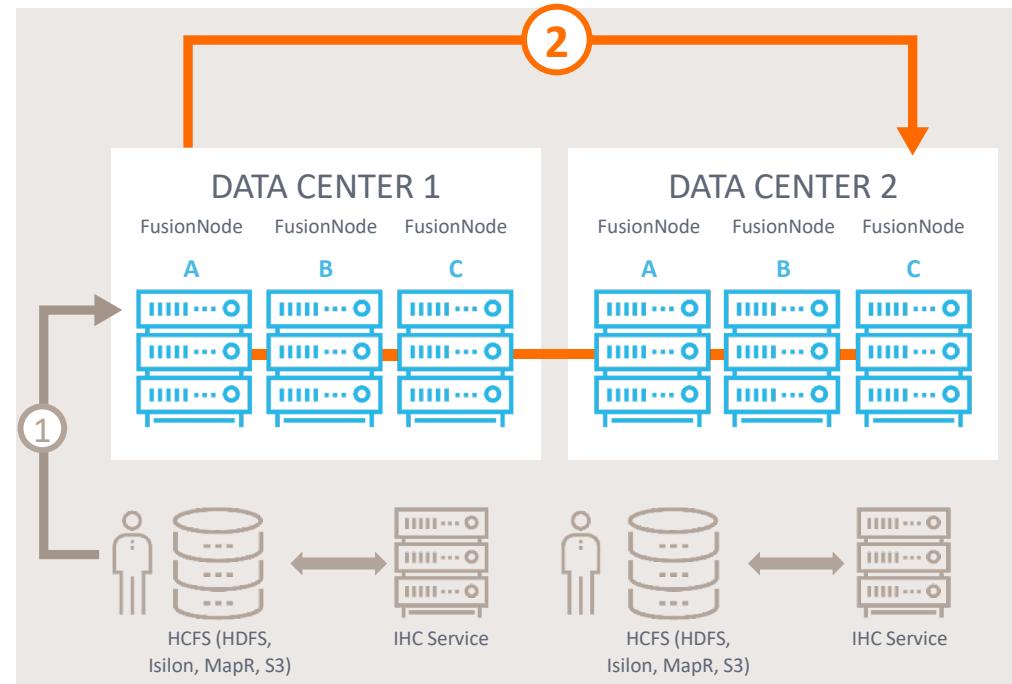
WANdisco FUSION WORKFLOW



HDFS HDFS

Fusion workflow

1. User makes a request to create / change a file
2. Fusion coordinates File Open to other clusters involved (membership)



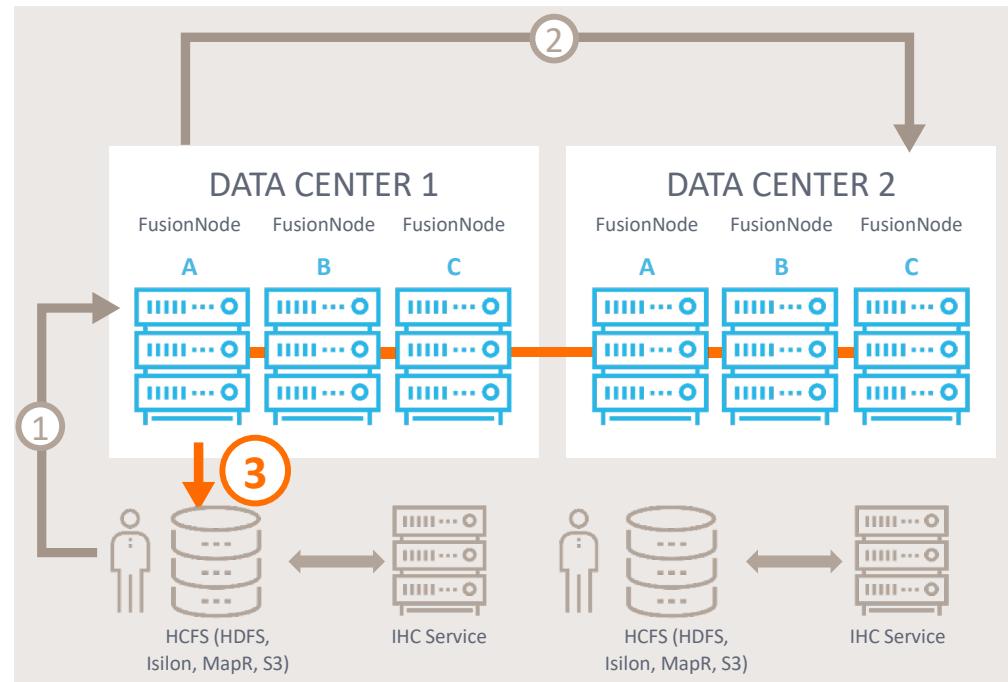
WANdisco FUSION WORKFLOW



HDFS  HDFS

Fusion workflow

1. User makes a request to create / change a file
2. Fusion coordinates File Open to other clusters involved (membership)
3. File is added to underlying storage



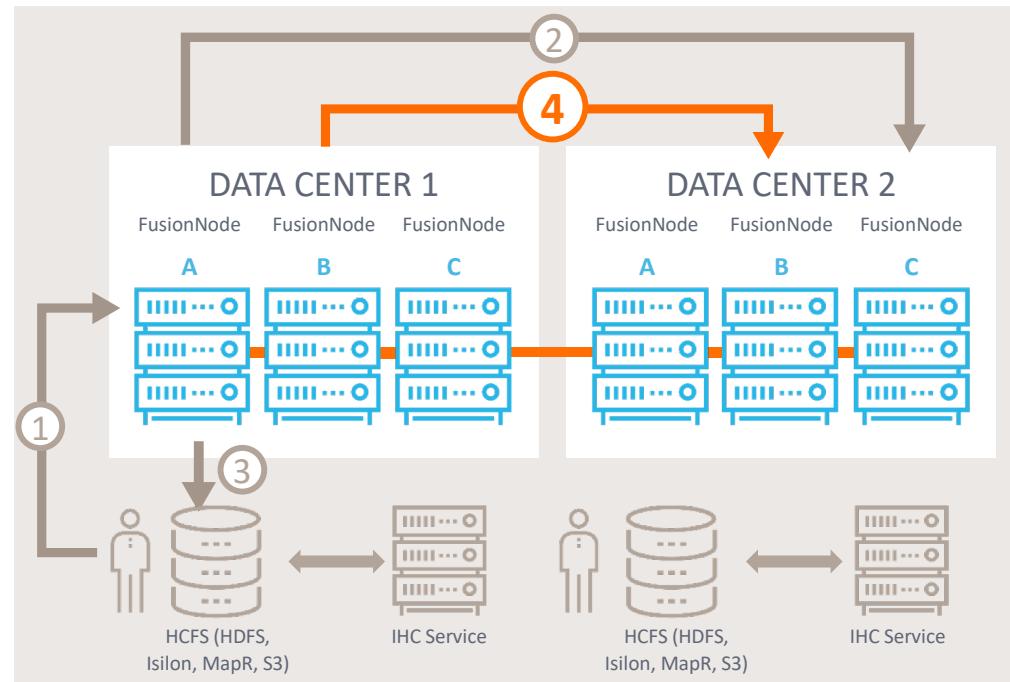
WANdisco FUSION WORKFLOW



HDFS  HDFS

Fusion workflow

1. User makes a request to create / change a file
2. Fusion coordinates File Open to other clusters involved (membership)
3. File is added to underlying storage
4. Fusion coordinates at configurable write increments and File Close with other clusters (membership)



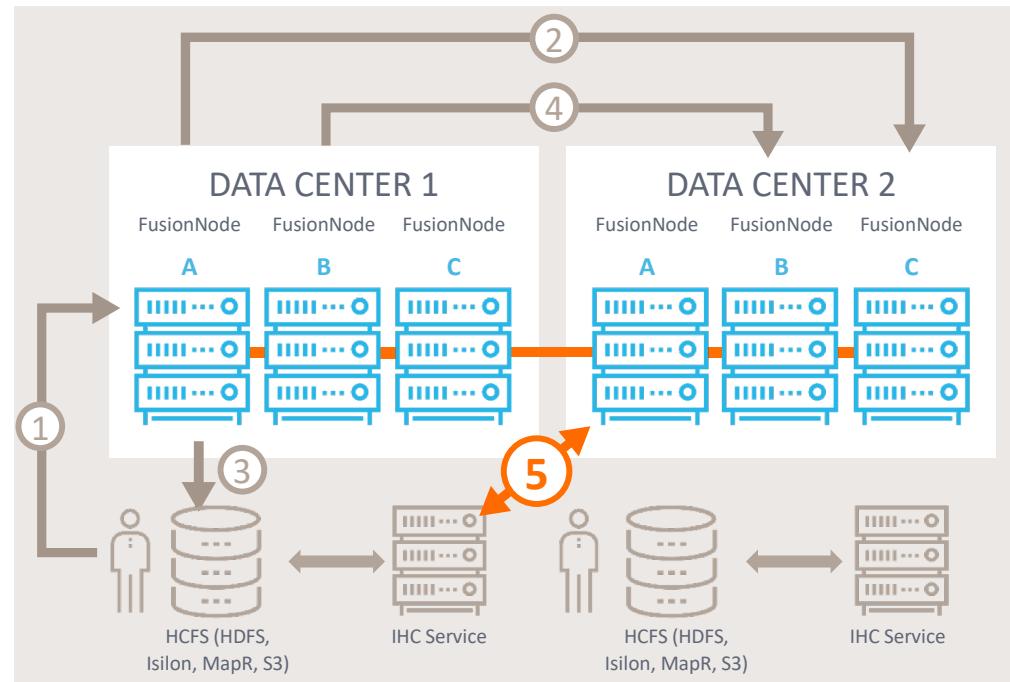
WANdisco FUSION WORKFLOW



HDFS HDFS

Fusion workflow

1. User makes a request to create / change a file
2. Fusion coordinates File Open to other clusters involved (membership)
3. File is added to underlying storage
4. Fusion coordinates at configurable write increments and File Close with other clusters (membership)
5. Fusion server at remote cluster pulls data from IHC server on source cluster



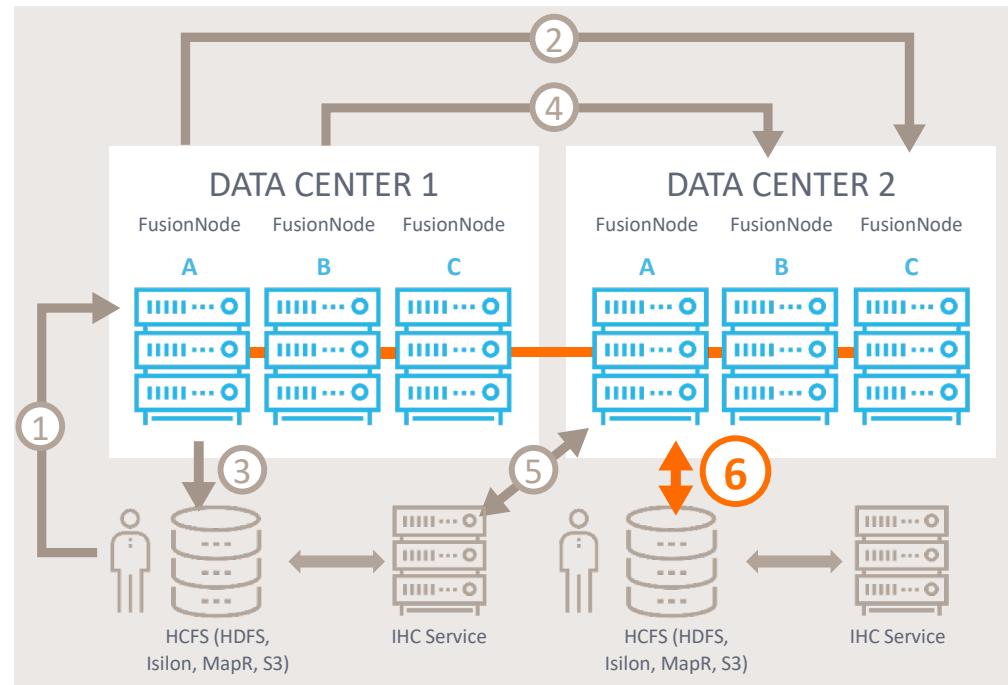
WANdisco FUSION WORKFLOW



HDFS HDFS

Fusion workflow

1. User makes a request to create / change a file
2. Fusion coordinates File Open to other clusters involved (membership)
3. File is added to underlying storage
4. Fusion coordinates at configurable write increments and File Close with other clusters (membership)
5. Fusion server at remote cluster pulls data from IHC server on source cluster
6. Fusion server at remote site writes data to its local cluster





ARCHITECTURE PRINCIPLES

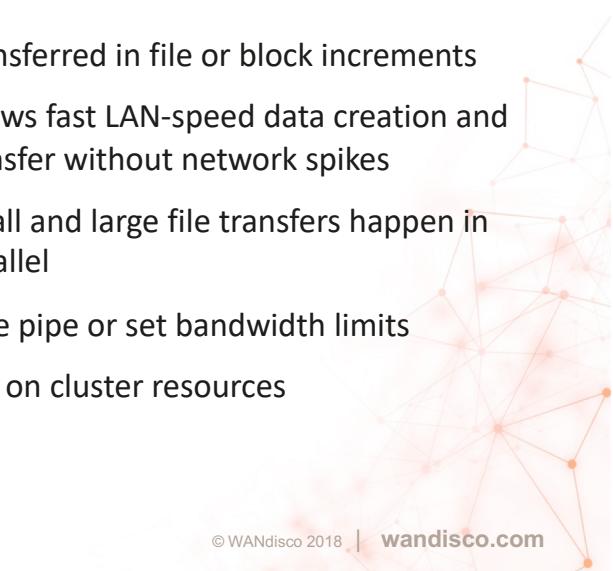
Strict consistency of metadata with fast data ingest and transfer

Synchronous replication of metadata between clusters / data centers

- Uses patented IP to coordinate updates to namespace
- Provides strict single copy consistency

Asynchronous replication of data over the WAN

- Content read from source and replicated to all zones in the background
- Data can be transferred in file or block increments
 - Allows fast LAN-speed data creation and transfer without network spikes
 - Small and large file transfers happen in parallel
- Utilizes available pipe or set bandwidth limits
- Minimal impact on cluster resources

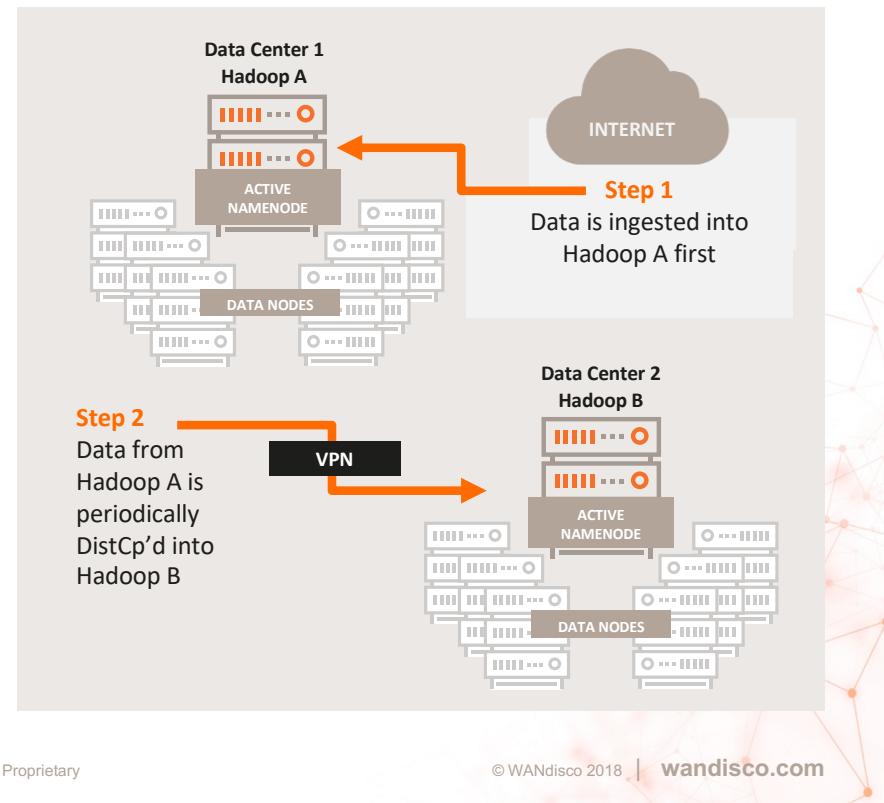




THE DATA REPLICATION PROBLEM APPROACHES IN USE TODAY: PERIODIC SYNCHRONIZATION

DISTCP

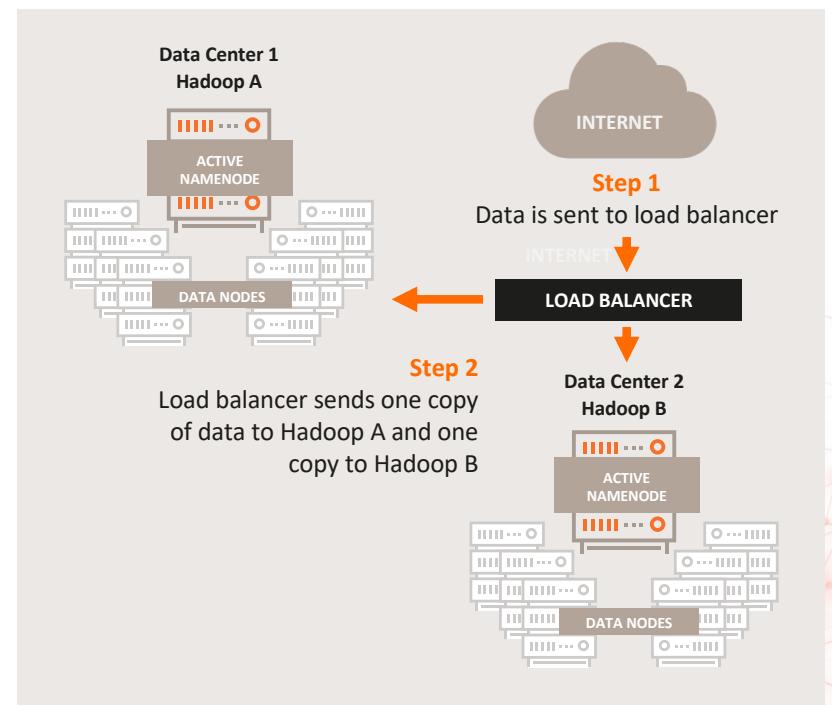
- Runs as MapReduce
- DR Data Center is typically read only
- Over time, Hadoop clusters become inconsistent
- Manual and labor intensive process to manage and reconcile differences
- Inefficient use of the network
- Only works between the same Hadoop distribution





THE DATA REPLICATION PROBLEM APPROACHES IN USE TODAY: PARALLEL (DUAL) INGEST

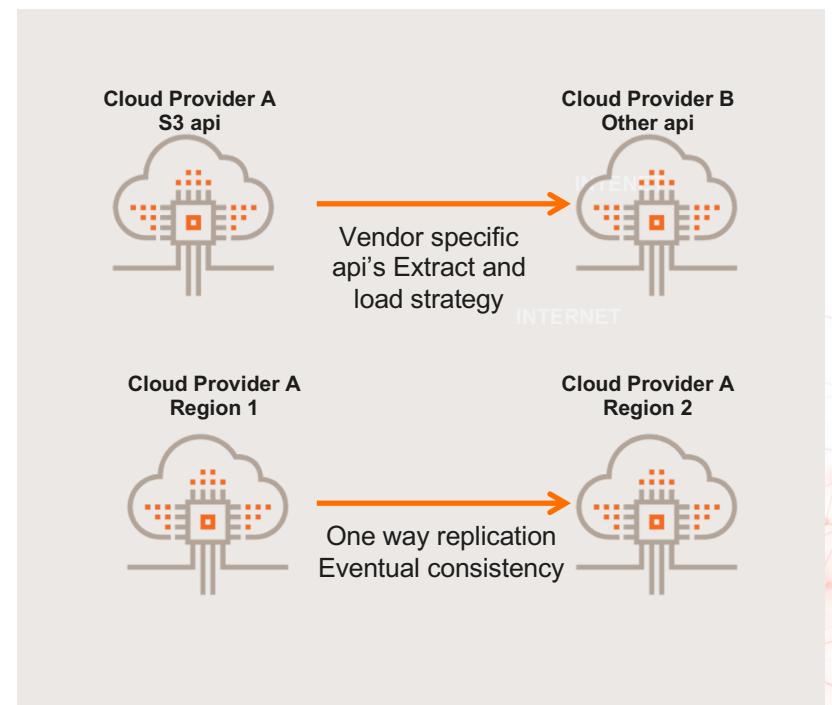
- Disruptions in either of the Hadoop clusters cause the two file systems to diverge
- Potential to run out of buffer when WAN is down
- Requires constant attention and sysadmin hours to keep running
- Data updated on one cluster is not replicated



THE DATA REPLICATION PROBLEM APPROACHES IN USE TODAY: CLOUD



- Global consistency not possible in a multi-cloud environment (vendor specific only)
- Uni-directional
- In some environments, replication is eventually consistent, but not guaranteed to be immediately available





THE LIVEDATA DIFFERENCE

SOLUTION PROVIDER	TECHNOLOGY	WHY THE LIVEDATA DIFFERENCE MATTERS
Some Big Data Vendors and Data Integration Vendors	Copy/Backup DistCP, Dual load, Master/replica, CDC, Snapshot	<ul style="list-style-type: none">• Point-in-time• One-way• Manual batch process• Lacking sync from downstream use• Vendor-specific (isolation)
Some Cloud Vendors	Cloud object storage	<ul style="list-style-type: none">• Vendor-specific (isolation)• Limited capability and access• Eventually consistent
WANDISCO	Fusion Kernel LiveData	<ul style="list-style-type: none">• Full r/w access• Real-time• Guaranteed consistency at petabyte-scale



FUNDED INITIATIVES: HOW CUSTOMERS LEVERAGE WANDISCO LIVEDATA

DISASTER RECOVERY

Automatic recovery with guaranteed data consistency, zero disruption, cost savings

MIGRATION TO CLOUD

Migrate transactional production system data to the cloud with no downtime or data loss

HYBRID CLOUD

On-prem and cloud environments stay in sync; every data change captured wherever it occurs

ANALYTICS INFRASTRUCTURE

Gain immediate, always-accurate insights from data generated across multiple environments

MULTI-CLOUD

Operate seamlessly between public clouds with 100% continuous data consistency

INTERNET OF THINGS

Ingest and analyze data at edge locations simultaneously; real-time accuracy everywhere

SECURITY AND COMPLIANCE

Complete control of data consistency and location to satisfy regulatory requirements



The **LiveData** Company

Visit us at www.wandisco.com

WANDisco Confidential and Proprietary

© WANDisco 2018 | wandisco.com