Data movement for globally deployed Big Data Hadoop architectures

Scott Rudenstein
VP Technical Services

November 2015
WANdisco Background

• WANdisco: Wide Area Network Distributed Computing
  ▶ Enterprise ready, high availability software solutions that enable globally distributed organizations to meet today’s data challenges of secure storage, scalability and availability

• Leader in tools for software engineers – Subversion
  ▶ Apache Software Foundation sponsor

• Highly successful IPO, London Stock Exchange, June 2012 (LSE:WAND)

• US patented active-active replication technology granted, November 2012
Scott Rudenstein has worked in commercial software sales for 18 years in Application Lifecycle Management and High Performance Computing. Throughout his career in the US and UK, he specialized in replication, where data and environments need high availability, disaster recovery and backup capabilities. As WANdisco’s VP of Technical Services, Scott works with partners, prospects and customers to help them understand and evolve the requirements for mission critical/enterprise-ready Hadoop.

Phil is the WANdisco Central Region Director, and brings over 20 years of experience working with clients on enhancing their Big Data, business intelligence, PLM, ALM, search, analytics and visualization environments. He has worked for JP Morgan, CSC, Kodak, Dassault Systemes, and Attivio with the last eight (8) years focused on Big Data and is an active DAMA member.
How Much Data We Generate in 60 Seconds

- 72 MILLION Google Searches
- 14 MILLION New Songs Added
- 15 THOUSAND New WordPress Posts
- 70 NewRegistered Domains
- 347 New Email Transactions
- 204 MILLION Emails sent
- 20 THOUSAND New Twitter Photos
- 350GB of data uploaded
- 41 THOUSAND Facebook posts
- 1.4 MILLION Twitter Tweets
- 278 Photos Shared
- 20 MILLION Photos viewed
- $83,000 Amazon Sales
- 11 THOUSAND Active LinkedIn Users
- 11 THOUSAND Pinterest followers
- 278 Transactions
- 571 New Websites
- 104 THOUSAND Photos shared
- 17 THOUSAND New Walmart Photos
Even More Challenging with the Internet of Things

<table>
<thead>
<tr>
<th>Year</th>
<th>Population</th>
<th>Devices</th>
<th>Devices per Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>6.3 Billion</td>
<td>500 Million</td>
<td>0.08</td>
</tr>
<tr>
<td>2010</td>
<td>6.8 Billion</td>
<td>12.5 Billion</td>
<td>1.84</td>
</tr>
<tr>
<td>2015</td>
<td>7.2 Billion</td>
<td>25 Billion</td>
<td>3.47</td>
</tr>
<tr>
<td>2020</td>
<td>7.6 Billion</td>
<td>50 Billion</td>
<td>6.58</td>
</tr>
</tbody>
</table>
The Result: Hadoop is Under Pressure

Data-at-rest

**Reporting:**
Trends, Search indexes, KPIs, Matrices and BI

**Exploration:**
Data science, Large data sets analysis

Data-in-motion

**Streaming analysis:**
Real time summary and aggregation

**Transactional processing:**
Per event analytics, stream analysis
Before we move data

We’ve brought data into the cluster from multiple data sources, but before we move it we need to consider the following:

• We need to understand where the ingested data is going and what is the data’s purpose.
  ‣ Backup
  ‣ Disaster recovery
  ‣ Compute restraints
  ‣ High availability
  ‣ Enterprise Data Warehouse

• And we need to understand the data’s relevance or timeliness required for the data to be actionable.

• Network Capacity

• Security

• Cluster Utilization
Enterprise Ready Hadoop

Characteristics of mission critical Hadoop clusters sharing data

Require Data Availability
- SLA’s, Regulatory Compliance
- Regional datacenter failure
- Data Relevance

Require Hadoop Deployed Globally
- Share Data Between Data Centers
- Data is Consistent and Not Eventual

Ease Administrative Burden
- Reduce Operational Complexity
- Simplify Disaster Recovery
- Lower RTO/RPO

Allow Maximum Utilization of Resource
- Within the Data Center
- Across Data Centers
Key Issue For **Sharing** Data Across Clusters

LAN / WAN

\[
\begin{align*}
\text{Availability} & \quad + \quad \text{Consistency} & \quad = \quad \text{Complexity}
\end{align*}
\]
Considerations

Questions before transfers

• What is the value of the data being copied?
  › Customer facing / Business critical
  › Internal, medium priority
  › Disaster recovery / backup

• How important is it that the data is verified?
  › Compliance – must be
  › Recovery – near as possible
  › Don’t care – can get/regenerate the data again

• How many unique folders need to be replicated site to site?
  › What is the complexity of the transfer pattern?
  › Does the receiving site need to transfer data back to the originating site?
  › How does your organization troubleshoot a failure in replication

• Is the Disaster Recovery Site for?
  › Running critical applications when the production site is lost
  › Recovering data only
  › Both – must run critical applications and provide recovery data

• How is the Disaster Recovery Site sized?
  › Equal to total production clusters
  › Smaller footprint
  › Larger footprint

• What are the SLA’s that must be met?

• Is some data (folders) higher priority than other data?

• What are you using to control the copy process (i.e. distcp, Falcon, BDR, home grown)?

• How many tasks are there for production to copy the data (or will be in production if not currently)?
The complexity issue
The complexity issue

C ‘1’

repl1

repl2

C ‘2’

repl1

repl2
The complexity issue

HBase

<table>
<thead>
<tr>
<th>repl1</th>
<th>repl2</th>
<th>repl3</th>
</tr>
</thead>
</table>

Hive

<table>
<thead>
<tr>
<th>repl1</th>
<th>repl2</th>
<th>repl3</th>
</tr>
</thead>
</table>

HDFS

<table>
<thead>
<tr>
<th>repl1</th>
<th>repl2</th>
<th>repl3</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>repl1</th>
<th>repl2</th>
<th>repl3</th>
<th>repl4</th>
</tr>
</thead>
</table>
The complexity issue
Analysis of common replication techniques
Multi Data Center Hadoop Today

What's wrong with the status quo

Periodic Synchronization - DistCp

1. Data is ingested into Hadoop A first
2. Data from Hadoop A is periodically DistCp'ed into Hadoop B

Parallel Data Ingest - Load Balancer, Streaming

1. Data is sent to load balancer
2. Load balancer sends one copy of data to Hadoop A and one copy to Hadoop B
Multi Data Center Hadoop Today

Hacks currently in use

- Runs as Map reduce
- DR Data Center is read only
- Over time, Hadoop clusters become inconsistent
- Manual and labor intensive process to reconcile differences
- Inefficient use of the network
- N to M datanode communication

Periodic Synchronization - DistCp
Multi Data Center Hadoop Today

Hacks currently in use

- Hiccups in either of the Hadoop cluster causes the two file systems to diverge
- Potential to run out of buffer when WAN is down
- Requires constant attention and sys-admin hours to keep running
- Data created on the cluster is not replicated
- Use of streaming technologies (like flume) for data redirection are only for streaming

Parallel Data Ingest - Load Balancer, Streaming

1. Data is sent to load balancer
2. Load balancer sends one copy of data to Hadoop A and one copy to Hadoop B
What we want in our Data Centers

What’s in a Data Center

Standby Datacenter

• Idle Resource
  ▶ Single Data Center Ingest
  ▶ Disaster Recovery Only
• One way synchronization
  ▶ DistCp
• Error Prone
  ▶ Clusters can diverge over time
• Difficult to scale > 2 Data Centers
  ▶ Complexity of sharing data increases

Active Datacenter

• DR Resource Available
  ▶ Ingest at all Data Centers
  ▶ Run Jobs in both Data Centers
• Replication is Multi-Directional
  ▶ Hive
  ▶ HBase/Other
• Absolute Consistency
• ‘N’ Data Center support
  ▶ Global Hadoop shared only appropriate data
Scaling Hadoop Across Data Centers

Continuous Availability and Disaster Recovery over the WAN

- The system should appear, act, operate as a **single cluster**
  - Continuous replication of data and metadata
  - RPO is as low as possible due to continuous replication as opposed to periodic
- Parts of the cluster on different data centers should have **equal roles**
  - Data could be ingested or accessed through any of the centers
- Data creation and access should be at **LAN speed**
  - Running time of a job executed on one data center as if there are no other centers
- **Failure scenarios**: the system should provide service and remain consistent
  - WAN Partitioning does not cause a data center outage
Use Cases
Disaster Recovery

- Data is as current as possible (no periodic synchs)
  - Low RPO
- Virtually zero downtime to recover from regional data center failure
  - Active/Active means your already running
- Meets or exceeds strict regulatory compliance around disaster recovery
  - Auditability
Multi Data-Center

Ingest and multi-tenant workloads

- Ingest and analyze anywhere
- Analyze Everywhere
  - Fraud Detection
  - Equity Trading Information
  - New Business
  - Etc…
- All Datacenter(s) can be used for work
  - No idle resource
Cluster Zones

• Share data, not processing
  - Isolate lower priority (dev/test) work
  - Share data not resource
• Maximize Resource Utilization
  - No idle standby
• Mixed Hardware Profiles
  - Memory, Disk, CPU
  - Isolate memory-hungry processing (Storm/Spark) from regular jobs
• Mixed Vendor
  - CDH, PHD, HDP, MAPR
Migration

• Simplifies and automates migration seamlessly between competitive distributions or legacy versions.
  › MAPR, CDH – HDFS, CDH – Isilon, HDP – HDFS, HDP Isilon, PHD

• With easily configurable, real-time active-active replication, data stays secure and maintains its integrity during the migration process
  › Eliminates the manual labor and pain on IT

• Facilitates and expedites migrations and upgrades to distributions, reducing costly downtime and minimizing data loss.

• During migration process, enterprises can access their data continuously.
  › Both sites become operational
  › Data Liveliness

• Avoids vendor lock in
Regulatory Compliance

- Basel III
  - Consistency of Data

- Data Privacy Directive
  - Data Sovereignty
    - data doesn’t leave country of origin
Security Between Data Centers

• Hadoop clusters do not require direct communication with each other.
  
  ▶ No n x m communication among datanodes across datacenters
  ▶ Reduced firewall / socks complexities

• Reduced Attack Surface

• Fast network protocols can keep up with demanding network replication
What is Wandisco Fusion?

Active-active replication for Hadoop Compatible File Systems

- Active-active coordination of updates to HCFS data
- Replication of data between similar or mixed HCFS clusters
- Management UI backed by REST API

- HDP, CDH, MapR, Apache Hadoop
- HDFS, EMC Isilon, Amazon S3
- On-premise, cloud
WANdisco Fusion Overview

Powered by WANdisco’s patented DConE to enable:

- Active-active replication with guaranteed data consistency across clusters over LAN or WAN

Unifies clusters running on CDH, HDP, PHD, MapR, EMC Isilon, NetApp, EMC ECS, Amazon S3, MS Azure, OpenStack Swift, Google Cloud Storage and others

- Provides a single virtual namespace across clusters any distance apart

Totally non-invasive:

- Proxy server(s) deployed with each cluster
- No modification to underlying storage

An “SDK” (Software Development Kit) is available for developers to write plug-ins for additional data sources
The complexity issue

- **HBase**
  - repl1
  - repl2
  - repl3

- **Hive**
  - repl3

- **HDFS**
  - repl1
  - repl2
  - repl3
  - repl4
The complexity issue

DC '1'

rep1
rep2
rep3
rep4

DC '2'

rep1
rep2
rep3
rep4

DC '3'

rep1
rep2
rep3
rep4

DC '4'

rep1
rep4
QUESTIONS

Please feel free to submit your questions