

# Backup and Recovery for Hadoop: Plans, Survey and User Inputs

Prasanth Kothuri, CERN

Hadoop User Forum  
18th May 2016

# Agenda

- Need for Backup
- Backup Design Philosophy
- Questionnaire & User Inputs
- Implementation Challenges
- Size and Complexity



# What can go wrong?

- Hardware Failures
  - Disk Corruption
  - Node Failure
  - Rack Failure
- Human / Application Error
  - Logical corruption
  - Accidental deletion of data
- Site Failure

# What Needs Protection?

- Data sets
  - Data &
  - Meta data (NameNode, Hive, HBASE etc)
- Applications
  - System and user applications
- Configuration
  - Config of various Hadoop components

# Spotlight is on..

## Data sets

For many of the applications / analysis this is the **only copy of the data**, i.e the data sets cannot be regenerated

And the **applications** and **configuration** is protected by the way we work (git, puppet and service documentation)

# Why it's not already there?

- No out of the box point in time recovery solution
- Complex and not straightforward with several data stores (HDFS, HBASE)
- Hence we work with the **users** to understand their usage of Hadoop and come up with **possible implementations**

# Backup Design Philosophy

- Determine whether the data stored in HDFS are mission critical
- Determine the datasets to be backed up?
- Priority of the datasets to be backed up
- What's the best way to back up data?

# Backup Design Philosophy

- Changes since the last backup
  - Tracking the changes and the count of files added/changed
- The time interval between backups (reducing the window of possible data loss)
- The rate of new data arrival
  - Determine the size of new data
  - Time factor involved between two backups (data loss)



# Questionnaire – Gathering User Inputs

## Hadoop Service: Backup and Recovery Questionnaire

Name:

Application:

Hadoop Cluster: Choose an item.

Question	Response
How critical is your data on Hadoop clusters? If yes, do you have another copy of the data outside of Hadoop?	
What type of information does the data contain? (e.g. server logs)	
Describe your data organization / partition schemes? (e.g. HDFS folder structure)	
How dynamic is your data? (e.g. do you update the data which may have been archived)	
What kind of data volume growth can we expect?	
How quickly would you need to recover the data? (RTO)	
Approximately what is the maximum acceptable amount of data loss? (RPO)	
Could you envisage to trigger a backup yourself, by sending a list of hdfs files (or directories) to a backup tool?	
What is the required Backup Retention period?	
Should the restore performance favour bulk-restore or detail-restore, i.e. would you favour individual file picking over full recovery	

# Questionnaire – User Inputs

- How critical is the data and do you have another copy outside of hadoop
  - Semi-critical to very critical
  - Only one user responded as ‘not critical’
  - For about 50% of the applications data in hadoop is the only copy
- What type of data is stored in Hadoop
  - Server logs, security logs, system logs, job logs and metrics
  - Event metadata (ATLAS)
  - Accelerator devices logs

# Questionnaire – User Inputs

- Data organization
  - Most of the users responded with precise HDFS tree structure
  - Example:-
    - /project/itmon/archive/{provider}/{hostgroup}/YYYY-MM
    - /user/WDTMON/{provider}/YYYY/MM/DD
  - Data in HBASE
- How dynamic is the data
  - Data once written is not changed for most applications

# Questionnaire – User Inputs

- Daily volume growth
  - 2 TB/day to 150 GB/day
- Acceptable amount of data loss
  - All the users replied 'NO DATA LOSS'
- How quickly we need to recover
  - All of them replied as soon as possible

# Questionnaire – User Inputs

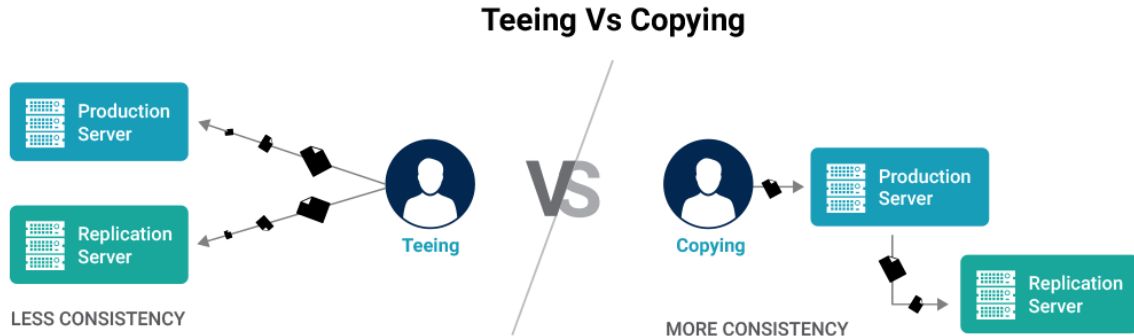
- Could you envisage to trigger the backup yourself?
  - Except one all of the users want central Hadoop service to run backups
- What is the required backup retention
  - Same as matching what's stored on HDFS
- Should the restore performance favour bulk or detail restore
  - User responses are a mixture between bulk and detail restore

# Backup Ideas

## Possible generic Hadoop backup solutions

### Copying vs Teeing

- Copying to another storage system like EOS/CASTOR
  - distcp is run to copy data from master storage to slave storage
- Teeing



# Backup Ideas

## HDFS Snapshots

- Snapshots of the entire file system or subtree of the file system can be taken for data backup, protection against user errors and disaster recovery.
- Snapshot differential report can be used for incremental backups

# Backup Ideas

## Possible backup solutions for HBASE

Aspects	Snapshot	Replication	Export	Copy Table	HBase API	Offline/ Manual
Performance impact	Less	Less	High	High	Medium	Depending on code
Data footprint	Smallest	Big	Big	Big	Big	Big
Cluster downtime	Less	None	None	None	None	High
Incremental backup	Not possible	Basic	Yes	Yes	Yes	Not possible
Implementation complexity	Easy	Medium	Easy	Easy	Complex	Medium



# Size and Complexity

Application	Current Size	Daily Growth
IT Monitoring	320.5 TB	140 GB
WLCG Monitoring	40.0 TB	60 GB
IT Security	156.8 TB	2048 GB
Accelerator Logging	500.0 TB	500 GB
ATLAS Rucio	5.9 TB	
CMS WMArchive	1.0 TB	50 GB
AWG	62.2 TB	
CASTOR Logs	163.1 TB	
WinCC OA	10.0 TB	25 GB
ATLAS EventIndex	220.0 TB	200 GB*
<b>Total</b>	<b>1.5 PB</b>	<b>4 TB</b>

# Challenges and Complexity

- Performing the backups without impacting performance
- Ability to determine the changes for the incremental backup
- Catering to various Hadoop processing frameworks (HDFS, HBASE, Impala etc)
- Consistency of the backups

# A word on Recovery

- Multiple scenarios – from total site failure to lost files
- Involves manual actions, can be time consuming and collaboration with the application users
- Some identified scenarios should be tested and rehearsed

# Conclusion

- Thanks to the **user response** to the questionnaire we now have **requirements** for the backup and recovery on Hadoop
- There are **several possible implementations**
- We will do the **PoC** to determine a solution that **covers all/most of the user requirements**

# Discussion / Feedback

Q & A