Migrating your clusters and workloads from Hadoop 2 to Hadoop 3

Suma Shivaprasad - Staff Engineer

Rohith Sharma K S - Senior Software Engineer

Speaker Info

Suma Shivaprasad

- Apache Hadoop Contributor
- Apache Atlas PMC
- Staff Engineer @ Hortonworks

Rohith Sharma K S

- Apache Hadoop PMC
- Sr.Software Engineer @ Hortonworks

HORTONWORKS

Agenda

- Why upgrade to Apache Hadoop 3.x?
- Things to consider before upgrade
- Upgrade process
- Workload migration
- Other aspects
- Summary

HORTONWORKS

Why upgrade to Apache Hadoop 3.x?

Motivation

Major release with lot of features and improvements!

HDFS

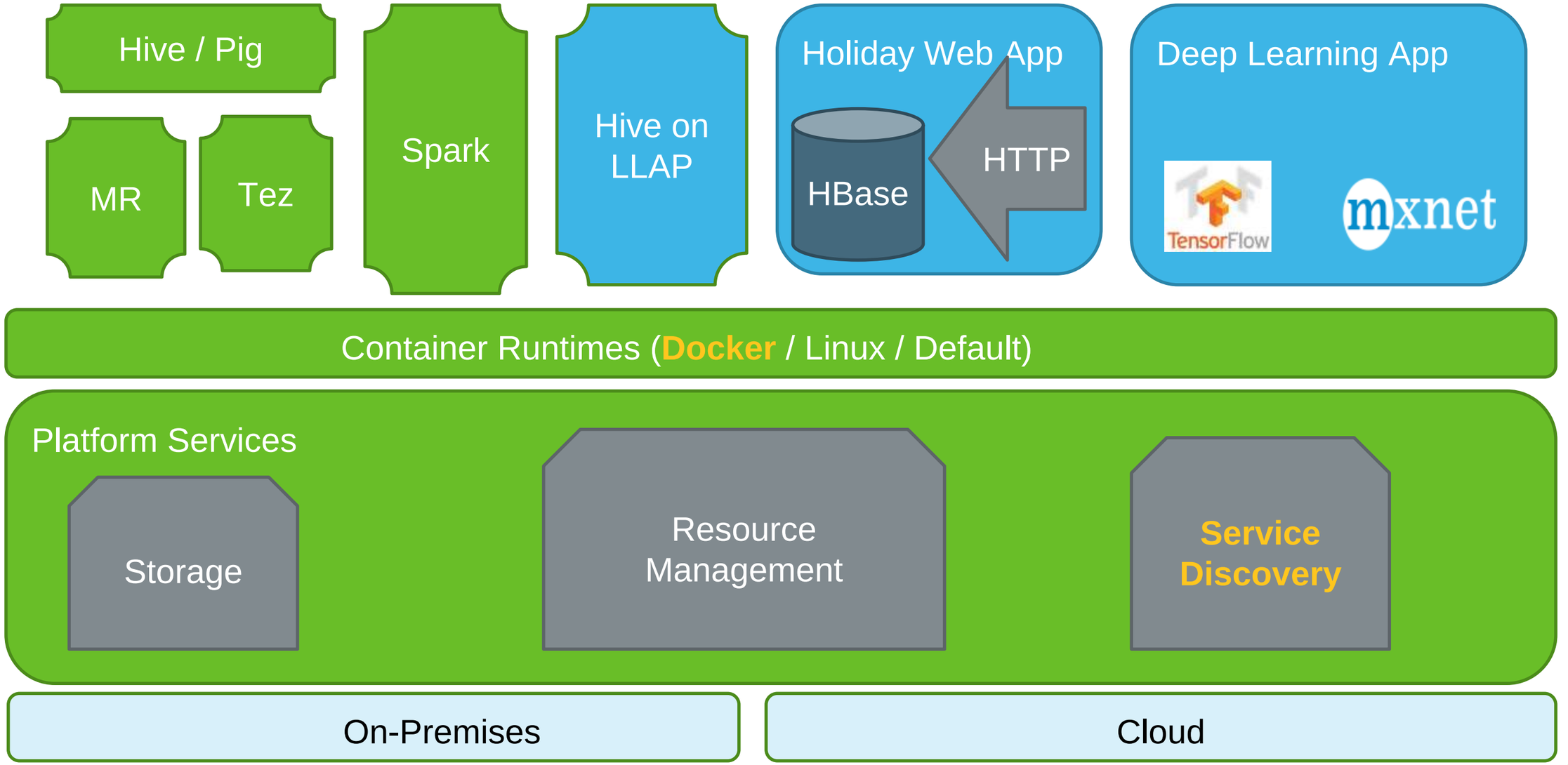
- Federation GA
- Erasure Coding
  - Significant cost savings in storage
  - Reduction of overhead from 200% to 50%
- Intra-DataNode Disk Balancer

YARN

- Scheduler Improvements
  - New Resource types - GPUs, FPGAs
  - Fast and Global scheduling
- Containerization - Docker
- Long running Services rehash
- New UI2
- Timeline Server v2

HORTONWORKS

Hadoop-3



HORTONWORKS

Things to consider before upgrade

Upgrades involve many things

- Upgrade mechanism
- Recommendation for 3.x - Express or Rolling ?
- Compatibility
- Source & Target versions
- Tooling
- Cluster Environment
- Configuration changes
- Script changes
- Classpath changes

HORTONWORKS

Upgrade mechanism: Express/Rolling Upgrades

Express Upgrades

- "Stop the world" Upgrades
- Cluster downtime
- Less stringent prerequisites
- Process
  - Upgrade masters and workers in one shot

Rolling Upgrades

- Preserve cluster operation
- Minimizes Service impact and downtime
- Can take longer to complete
- Process
  - Upgrades masters and workers in batches

HORTONWORKS

Recommendation for 3.x - Express or Rolling ?

- **Major version upgrade**
  - Challenges and issues in supporting Rolling Upgrades
- **Why rolling upgrades can't be done?**
  - HDFS-13596
    - Change in edit log format
  - HADOOP-15502
    - MetricsPlugin API In-compatibility change
  - HDFS-6440
    - Incompatible changes in image transfer protocol
- **Recommended**
  - **'Express Upgrade'** from Hadoop 2 to 3

HORTONWORKS

Compatibility

- **Wire compatibility**
  - Preserves compatibility with Hadoop 2 clients
  - Distcp/WebHDFS compatibility preserved
- **API compatibility**

  Not fully!
  - Dependency version bumps
  - Removal of deprecated APIs and tools
  - Shell script rewrite, rework of Hadoop tools scripts
  - Incompatible bug fixes!

HORTONWORKS

Source & Target versions

- Upgrades Tested with

Hadoop 2 Base version	Hadoop 3 Base version
Apache Hadoop 2.8.4	Apache Hadoop 3.1.x

- Why 2.8.4 release?
  - Most of production deployments are close to 2.8.x

- What should users of 2.6.x and 2.7.x do?
  - Recommend upgrading at least to Hadoop 2.8.4 before migrating to Hadoop 3!

HORTONWORKS

Tooling

- **Fresh Install**
  - Fully automated via **Apache Ambari**
  - Manual installation of RPMs/Tar balls
- **Upgrade**
  - Fully automated via **Apache Ambari 2.7**
  - Manual upgrade

HORTONWORKS

Cluster Environment

Java

- **>= Java 8**
- Java 7 EOL in April 2015
- Lot of libraries support only Java 8

Shell

- **>= Bash V3**
- POSIX shell NOT supported

Docker

- If you want to use containerized apps in 3.x
- **>= 1.12.5**
- Also corresponding stable OS

HORTONWORKS

Configuration changes: Hadoop Env files

hadoop-env.sh

- Common placeholder
- Precedence rule
  - *yarn/hdfs-env.sh > hadoop-env.sh > hard-coded defaults*

hdfs-env.sh

- HDFS_* replaces HADOOP_*
- Precedence rule
  - *hdfs-env.sh > hadoop-env.sh > hard-coded defaults*

yarn-env.sh

- YARN_* replaces HADOOP_*
- Precedence rule
  - *yarn-env.sh > hadoop-env.sh > hard-coded defaults*

HORTONWORKS

Configuration changes: Hadoop Env files Contd..

Daemon Heap Size HADOOP-10950

- Deprecated
  - **HADOOP_HEAPSIZE**
- Replaced with
  - **HADOOP_HEAPSIZE_MAX and HADOOP_HEAPSIZE_MIN**
- Units support in heap size
  - Default unit is MB
    - *Ex: HADOOP_HEAPSIZE_MAX=4096*
    - *Ex: HADOOP_HEAPSIZE_MAX=4g*
- Auto-tuning
  - Based on memory size of the host

HORTONWORKS

Configuration changes: YARN

Modified Defaults

- RM Max Completed Applications in State Store/Memory

Configuration	Previous	Current
yarn.resourcemanager.max-completed-applications	10000	1000
yarn.resourcemanager.state-store.max-completed-applications	10000	1000

HORTONWORKS

Configurations Changes: HDFS

Change in Default Daemon Ports (HDFS-9427)

Service	Previous	Current Port
NameNode	50470 50070	9871 9870
DataNode	50020 50010 50475 50075	9867 9866 9865 9864
Secondary NameNode	50091 50090	9869 9868
KMS	16000	9600

HORTONWORKS

Script changes: Starting/Stopping Hadoop Daemons

Daemon scripts

- *-daemon.sh deprecated
- Use bin/hdfs or bin/yarn commands with --daemon option
  - Ex: bin/*hdfs --daemon start/stop/status namenode*
  - Ex: bin/*yarn --daemon start/stop/status resourcemanager*

Debuggability

- Scripts support –debug
  - Construction of env
  - Java options and classpath

Logs/Pid

- Created as hadoop-yarn* instead of yarn-yarn*
- Log4j settings in the *-daemon.sh have been removed. Instead set via *_OPT in*-env.sh
  - Eg: *YARN_RESOURCEMANAGER_OPTS in yarn-env.sh*

HORTONWORKS

Classpath Changes

Classpath isolation now!
Users should rebuild their applications with shaded hadoop-client jars

- Hadoop Dependencies leaked to application's classpath - Guava, protobuf,jackson,jetty...
- Shaded jars available - isolates downstream clients from any third party dependencies HADOOP-11804
  - *hadoop-client-api* For compile time dependencies
  - *hadoop-client-runtime* For runtime third-party dependencies
  - *hadoop-minicluster* For test scope dependencies
- HDFS-6200 hadoop-hdfs jar contained both the hdfs server and the hdfs client.
  - Clients should instead depend on hadoop-hdfs-client instead to isolate themselves from server-side dependencies
- No YARN/MR shaded jars

HORTONWORKS

Upgrade process

Hadoop Pre-Upgrade Steps

STACK

- Backup Configuration files
- Stop users/services using YARN/HDFS
- Other metadata backup – Hive MetaStore, Oozie etc

YARN

- Stop all YARN queues
- Stop/Wait for Running applications to complete

HDFS

- Run fsck and fix any errors
  - *hdfs fsck / -files –blocks –locations > dfs-old-fsck.1.log*
- Checkpoint Metadata
  - *hdfs dfsadmin -safemode enter*
  - *hdfs dfsadmin -saveNamespace*
- Backup checkpoint files
  - *${dfs.namenode.name.dir}/current*
- Get Cluster DataNode reports
  - *hdfs dfsadmin -report > dfs-old-report-1.log*
- Capture Namespace
  - *hdfs dfs –ls –R / > dfs-old-lsr-1.log*
- Finalize previous upgrade
  - *hdfs dfsadmin –finalizeUpgrade*

HORTONWORKS

Upgrade Steps

Install new packages → Stop Services → Configuration Updates → Link to new versions → Start Services

Additional HDFS Upgrade Steps

https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.3/bk_command-line-upgrade/content/start-hadoop-core-25.html

HORTONWORKS

Upgrade Validation

HDFS

- Run HDFS Service checks
- Verify NameNode gets out of Safe Mode
  hdfs dfsadmin -safeMode wait
- FileSystem Health
- Compare with Previous State
  - Node list
  - Full NameSpace
- Let Cluster run production workloads for a while
- When ready to discard backup, finalize HDFS upgrade
  hdfs dfsadmin –upgrade finalize/query

YARN

- Run YARN Service checks
- Submit test applications – MR, TEZ, …

HORTONWORKS

Enable New features

- **Erasure Coding**
  - *https://hadoop.apache.org/docs/r3.0.0/hadoop-project-dist/hadoop-hdfs/HDFSErasureCoding.html*
- **YARN UI2**
  - *https://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-site/YarnUI2.html*
- **ATSv2**
  - New Daemon – Timeline Reader
  - https://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/TimelineServiceV2.html
- **YARN DNS**
  - Service Discovery of YARN Services
  - *http://hadoop.apache.org/docs/r3.1.0/hadoop-yarn/hadoop-yarn-site/yarn-service/RegistryDNS.html*
- ***HDFS Federation***
  - https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/Federation.html

HORTONWORKS

Migrating workloads

MapReduce (1/2)

Compatibility

- Full Binary compatibility of mapreduce APIs
- hadoop-streaming related deprecated IO Formats removed HADOOP-10485
  - *XMLRecordInput/Output*
  - *CSVRecordInput*

Configuration

- yarn.app.mapreduce.client.job.max-retries
  - Default changed from 0 to 3
  - Protects clients from failures that are transient.

HORTONWORKS

MapReduce (2/2) - Task Heap Management MAPREDUCE-5785

Heap size no longer needs to be specified in task configuration and Java options.

mapreduce.map.memory.mb	mapreduce.map.java.opts	Xmx Behaviour
Configured **2048MB**	Configured **1638 MB**	No Change **1638MB**
Configured **2048 M**B	Not Configured	Derived from mapreduce.map.memory.mb **1638MB**
Not Configured	Con**f**igure **1638 M**B	Automatically inferred from Xmx in mapreduce.map.java.opts. **1638MB**
Not Configured	Not Configured	Default : **1024 MB**

HORTONWORKS

Hive on Tez

- **Hive 3.0.0** Hive version supporting Hadoop 3 HIVE-16531
- Does NOT support rolling upgrades
  - Acid table format changes are not compatible with 2.x
- Tez version support for Hadoop 3
  - Planned for release **0.10.0**
    - TEZ-3923 Move master to Hadoop 3+ and create separate 0.9.x line
    - TEZ-3252 - [Umbrella] Enable support for Hadoop-3.x

HORTONWORKS

Spark

Ongoing efforts in community to build/validate Spark with Hadoop 3

- SPARK-23534 Umbrella jira to Build/test with Hadoop 3

HORTONWORKS

Apache HBase

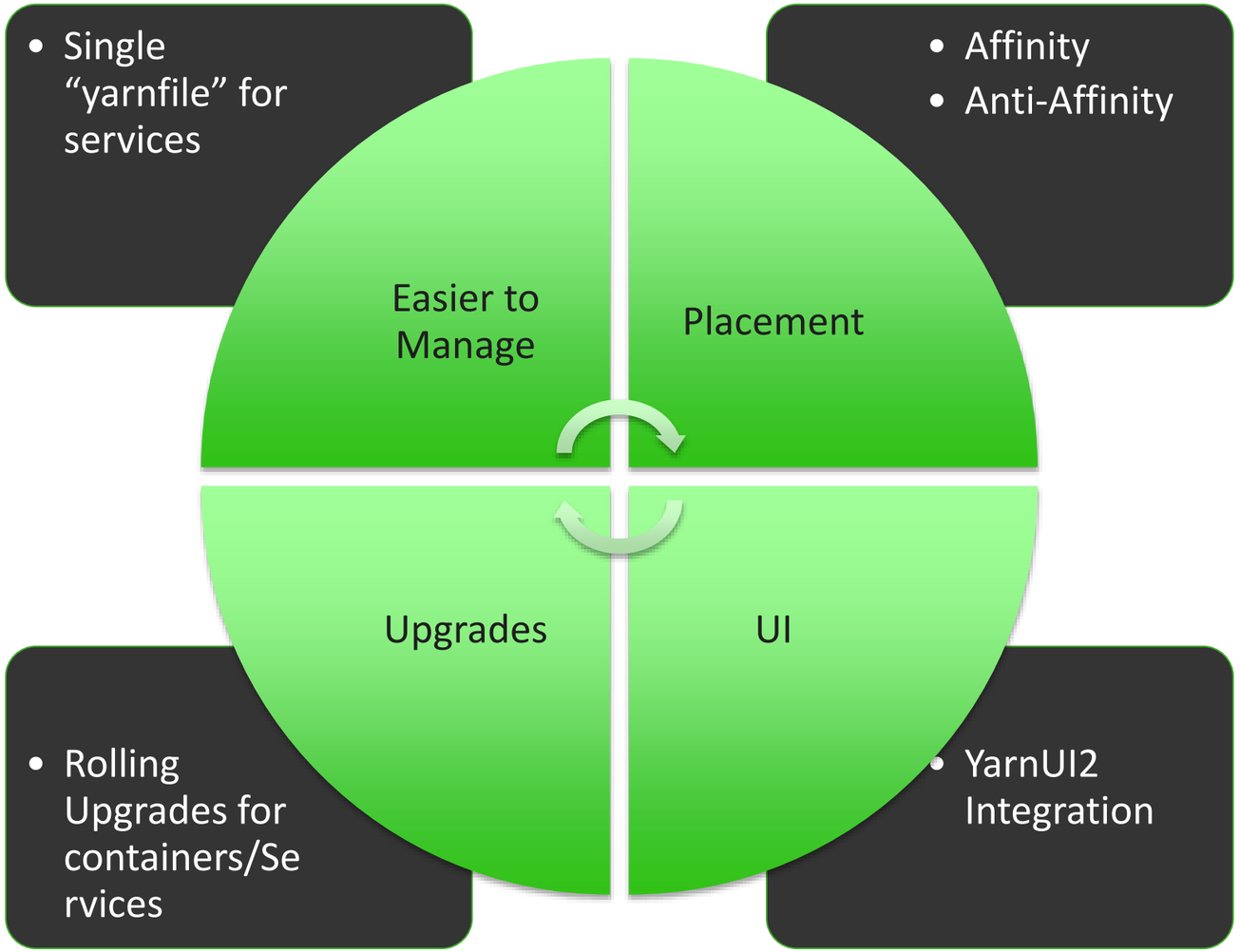
- **HBase 2.0** supports Hadoop 3
- Does NOT support Rolling Upgrades in major version upgrades (1.x to 2.x)
- Refer Upgrade documentation for further details

  https://github.com/apache/hbase/blob/master/src/main/asciidoc/_chapters/upgrading.adoc#upgrade2.0

HORTONWORKS

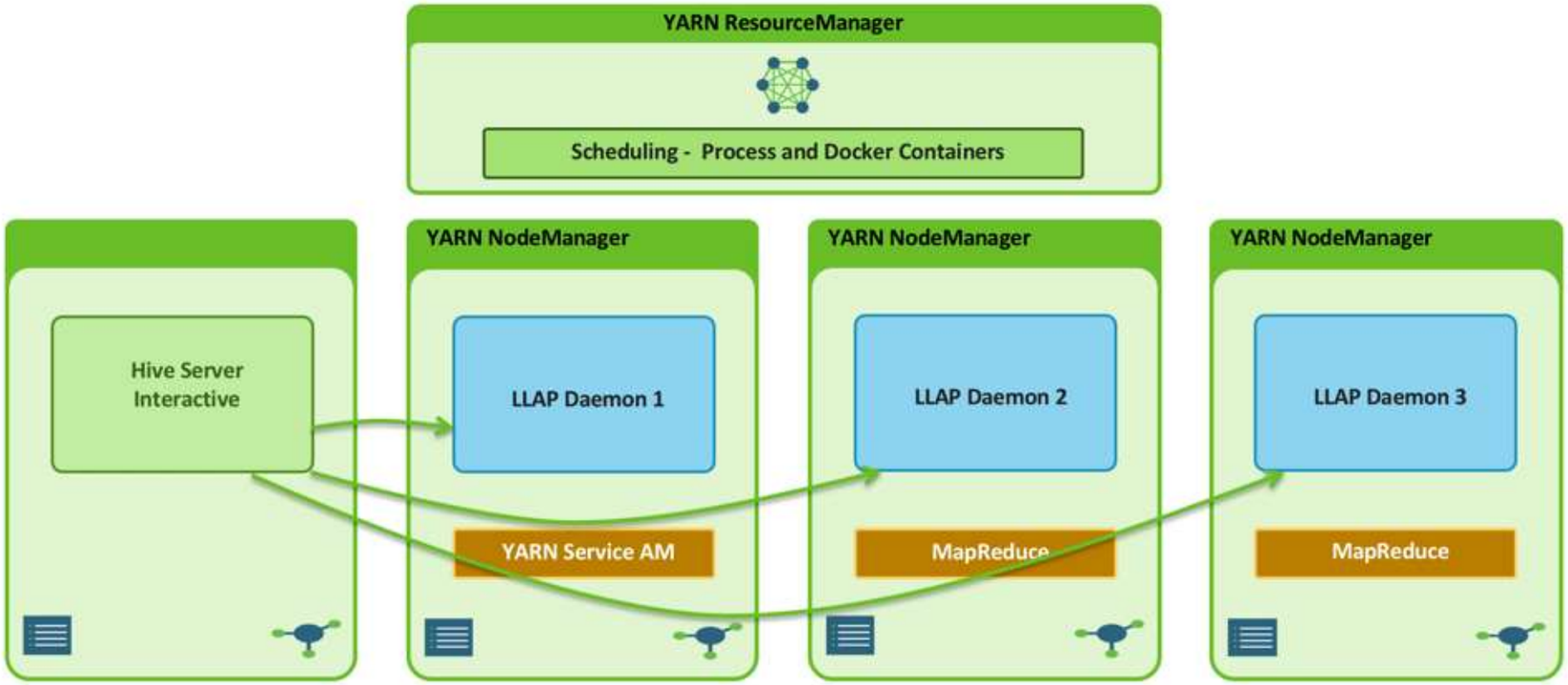
Apache Slider Applications

- Apache Slider is retiring from Apache Incubator
- Superseded by YARN Services.
- Port your Slider apps to Yarn Services
- Benefits of Yarn Services
  - Easier to manage and deploy
    - Single “yarnfile” to configure a Yarn Service
  - Supports container placement scheduling such as affinity and anti-affinity YARN-6592
  - Rolling upgrades for containers and service YARN-7512 and YARN-4726.
  - Services UI in YARN UI2 improving debuggability and log access.



Hive on LLAP

- Now runs as a Yarn Service Application instead of a Slider App
- Version that supports LLAP as a YARN service is not released yet.
  - Planned for release **Hive-4.0.0/3.1.0**
- Refer https://hortonworks.com/blog/apache-hive-llap-as-a-yarn-service



PIG/Oozie

Support for Hadoop 3 In-Progress in the community

- **PIG**
  - Planned for release – **0.18.0**
  - PIG-5253 Pig Hadoop 3 support
- **OOZIE**
  - Planned for release – **5.1.0**
  - OOZIE-2973 Make sure Oozie works with Hadoop 3

HORTONWORKS

Other Aspects

Other Aspects

Validations In-progress

- Performance testing
- Scale testing for HDFS/YARN
- OSes compatibility

HORTONWORKS

Summary

- **Hadoop 3**
  - Eagerly awaited release with lots of new features and optimizations !
  - 3.1.1 will be released soon with some bug fixes identified since 3.1.0
- **Express Upgrades** are recommended
- **Admins**
  - A bit of work
- **Users**
  - Should work mostly **as-is**
- **Community effort**
  - HADOOP-15501 Upgrade efforts to Hadoop 3.x
  - Wiki - https://cwiki.apache.org/confluence/display/HADOOP/Hadoop+2.x+to+3.x+Upgrade+Efforts
  - Volunteers needed for validating workload upgrades on Hadoop 3 !

HORTONWORKS


Questions?

Thank you