

特此声明

本视频由whoami博主（博客：itweet.cn）
Hadoop十周年庆开始公益录制，内容涉及
《Enterprise_Hadoop_Solutions》，声明此
教程使用图片大部分来自网络，版权属于广大
网友；任何个人或企业未经授权不得作商业用
途。想要持续关注更新或参与讨论请加入QQ
群：457710636！

Apache Hadoop 安装



新浪微博 @Mr-Robot1992

作者: whoami



Apache Hadoop集群

- 版本说明
 - zookeeper-3.4.6
 - Hadoop 2.4
 - Hive 0.13.1
 - Spark 1.3
 - Hbase 0.98.6.1-hadoop2
 - Phoenix 4.x - HBase 0.98.1+
 - 集群监控工具ganglia, Graphite, Nagios, Cacti
- 安装方式
 - 源码安装



Cloudera Hadoop安装

- 版本说明
 - CDH 5.2.0
 - http://www.cloudera.com/content/cloudera/en/downloads/quickstart_vms/cdh-5-3-x.html
- 安装方式
 - ClouderaManager离线安装
 - Yum 离线安装



Hortonworks Hadoop安装

- 版本说明
 - HDP 2.2
 - <http://zh.hortonworks.com/hdp/downloads/>
- 安装方式
 - Ambari 2.0方式安装

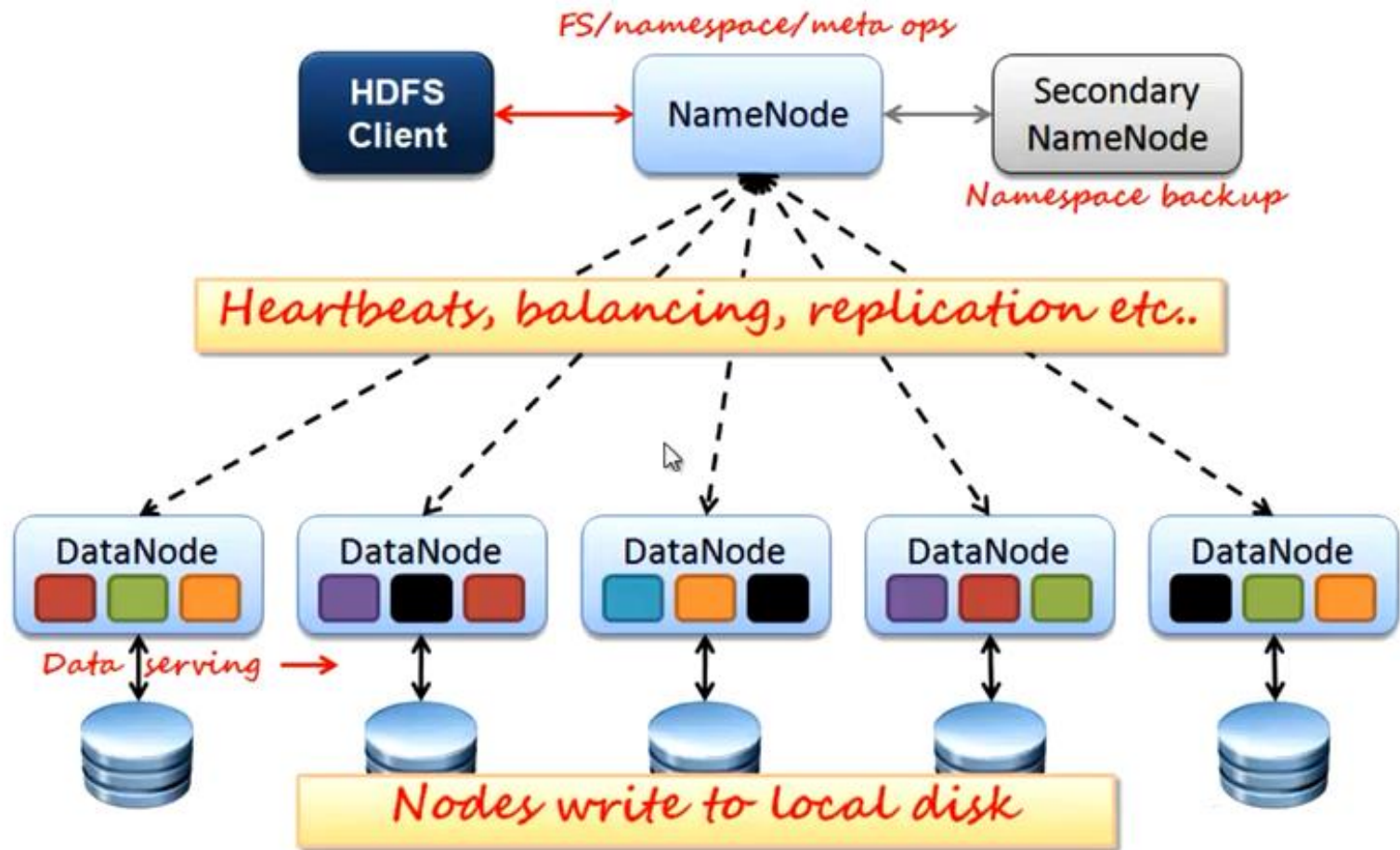


MapR Hadoop

- <https://www.mapr.com/>



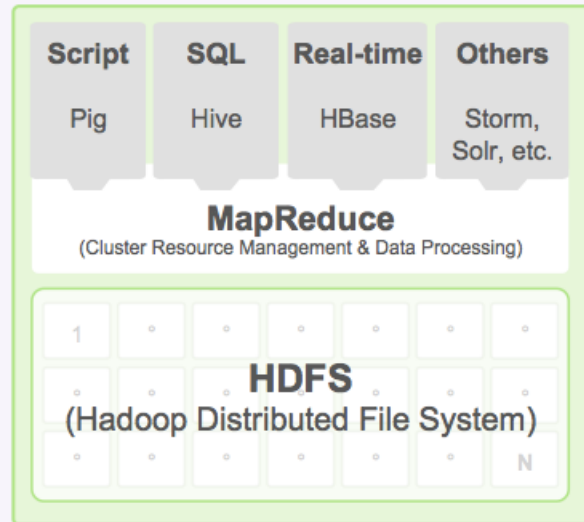
分布式存储



Why?

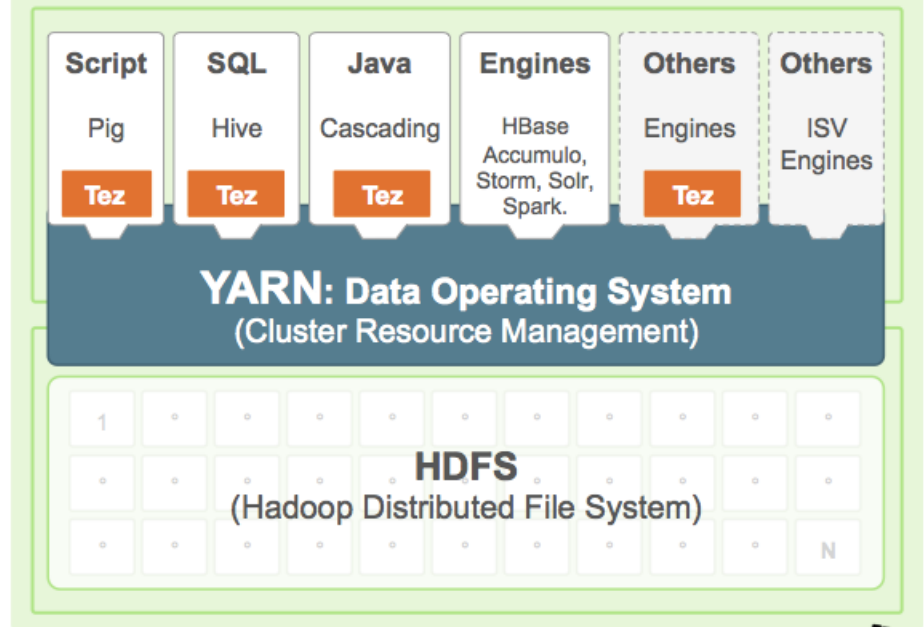
Hadoop 1

- Silos & Largely batch
- Single Processing engine

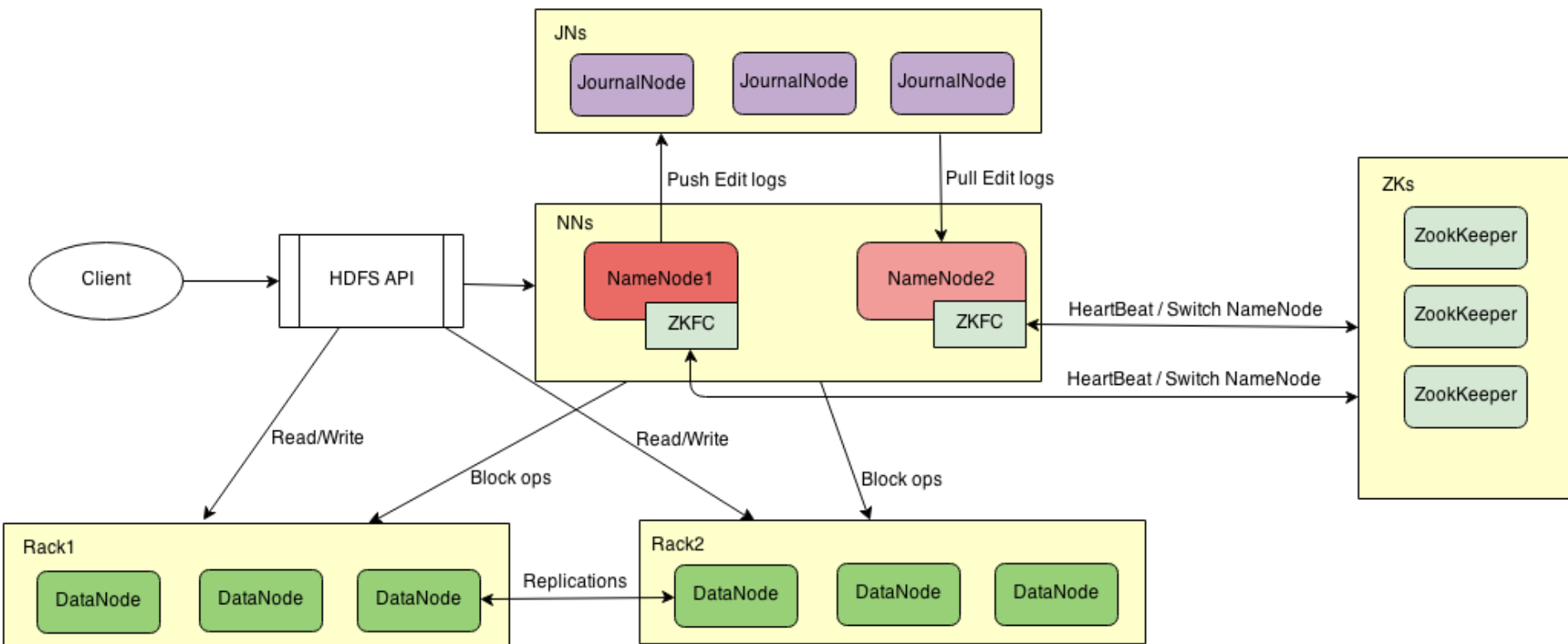


Hadoop 2 w/ Tez

- Multiple Engines, Single Data Set
- Batch, Interactive & Real-Time

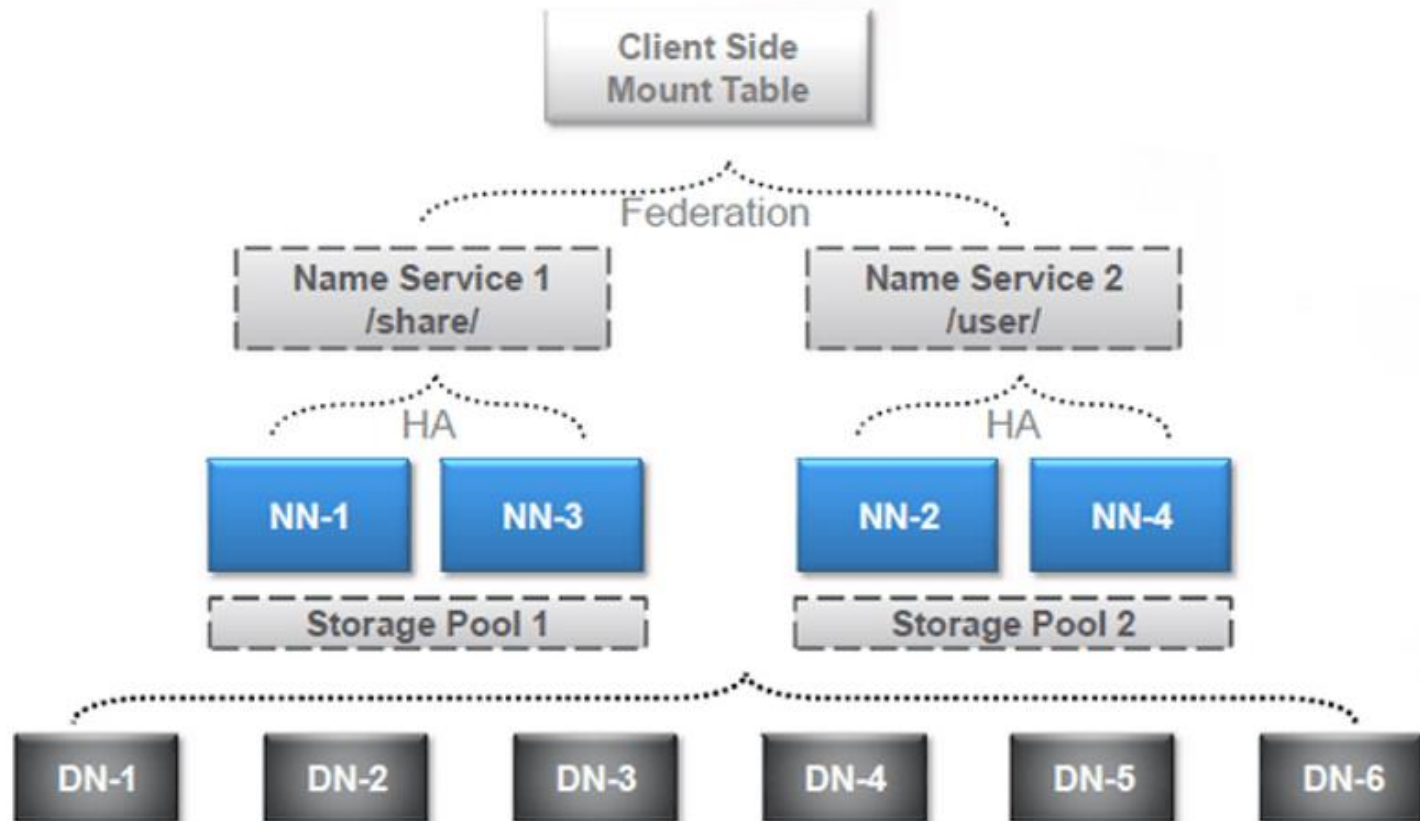


Hadoop Cluster HA



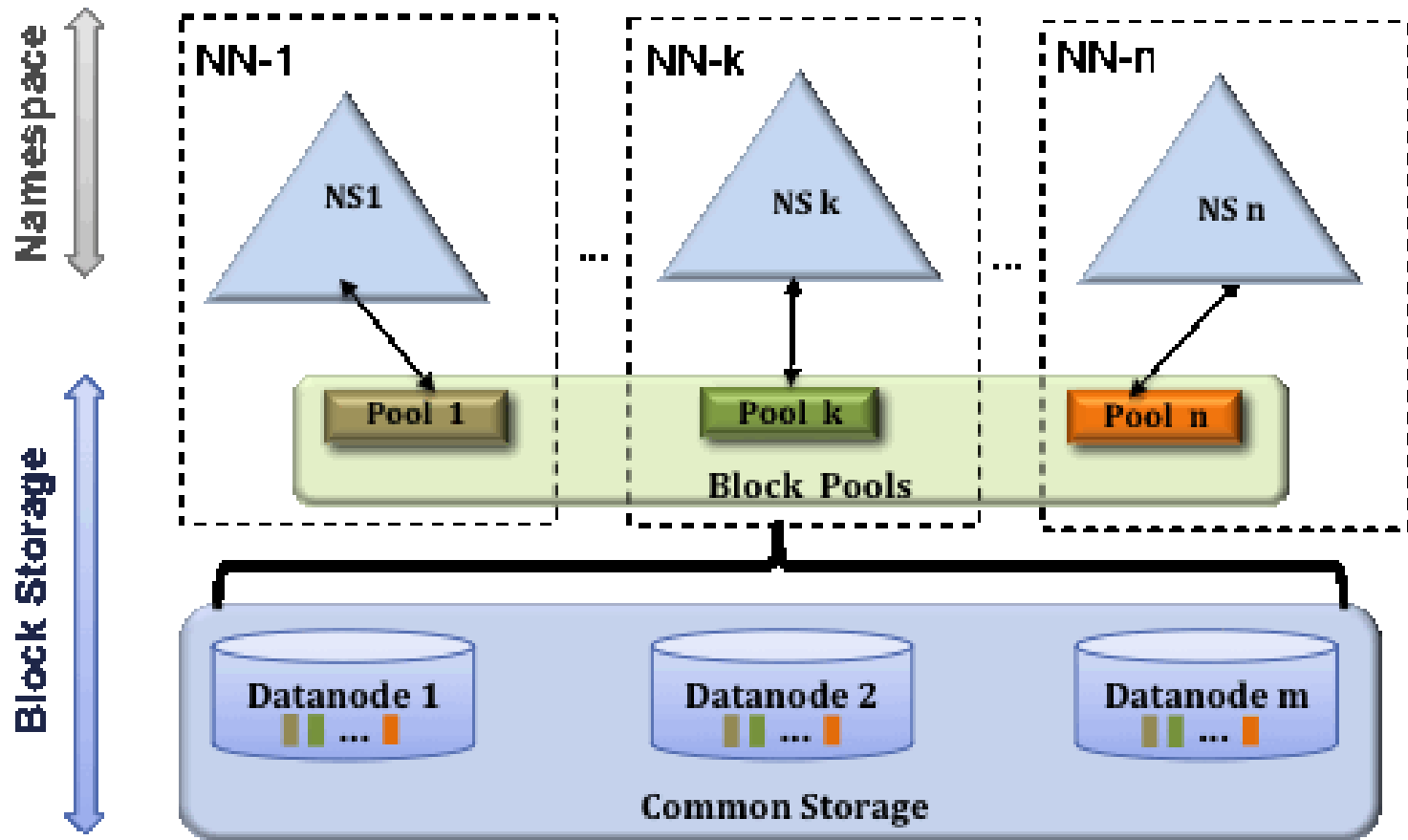
参考: <http://hadoop.apache.org/docs/r2.6.0/hadoop-project-dist/hadoop-hdfs/HDFSHighAvailabilityWithNFS.html>

Hadoop federation

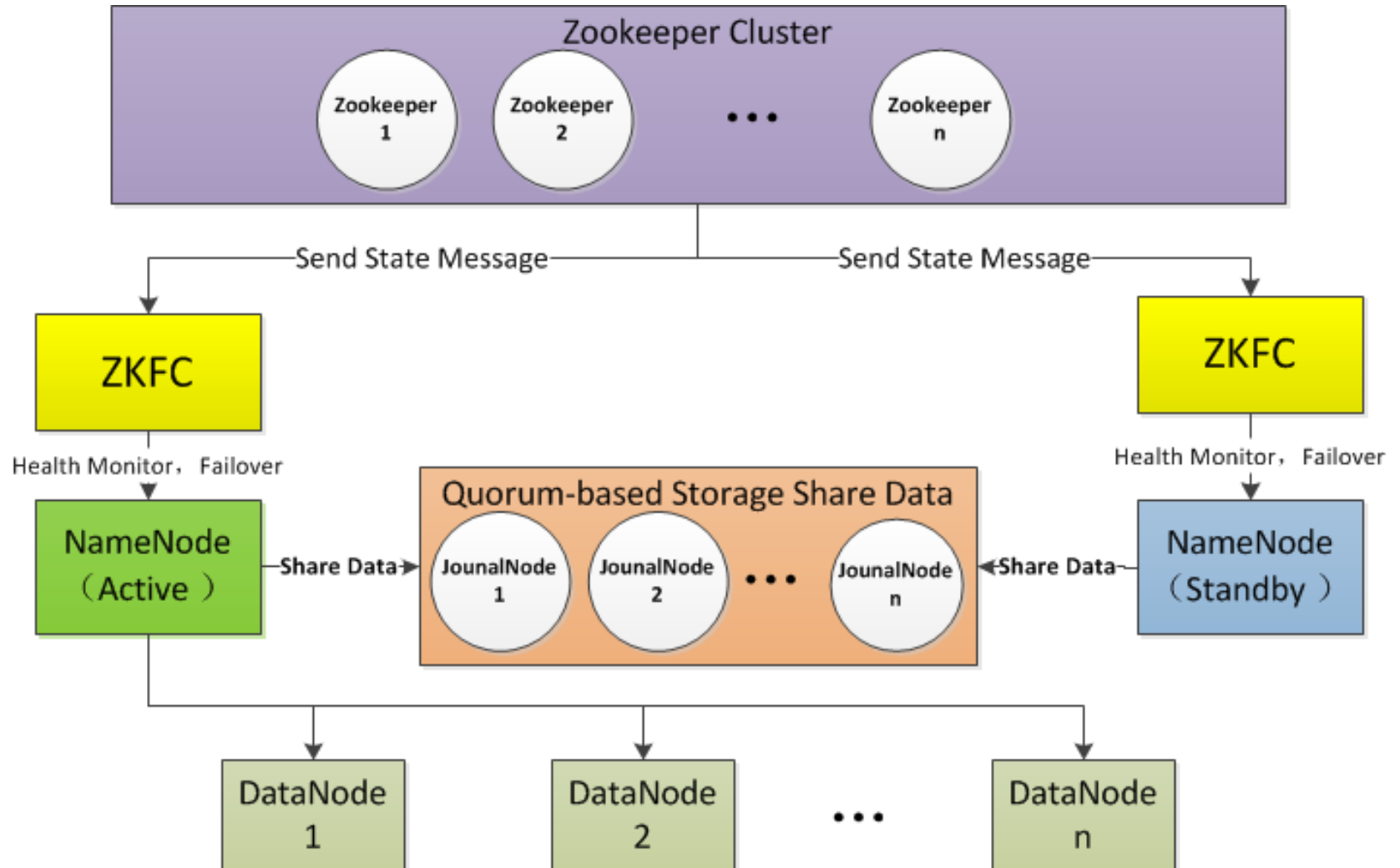


参考: <http://hadoop.apache.org/docs/r2.6.0/hadoop-project-dist/hadoop-hdfs/Federation.html>

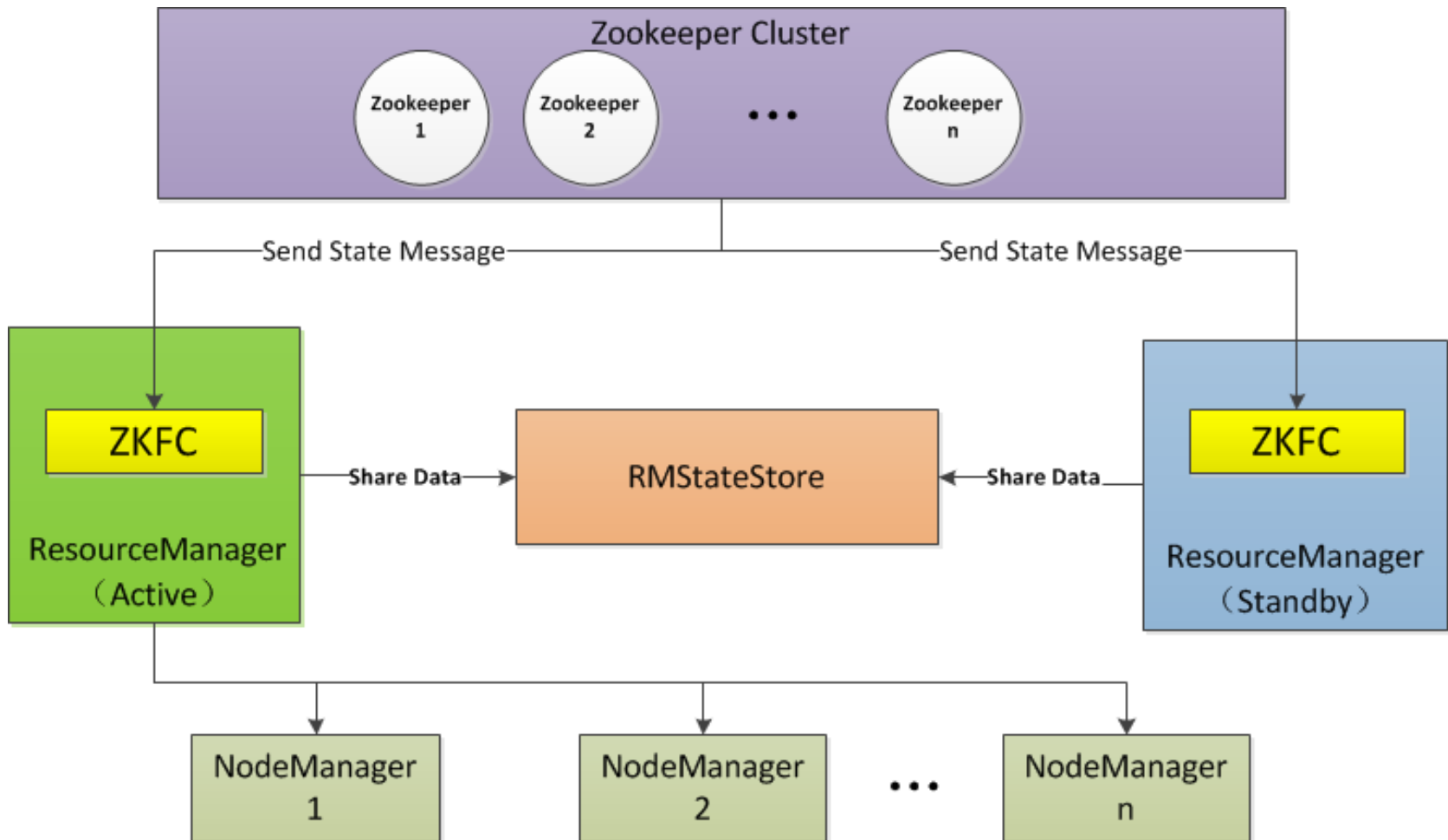
多重Namespaces/Namespaces



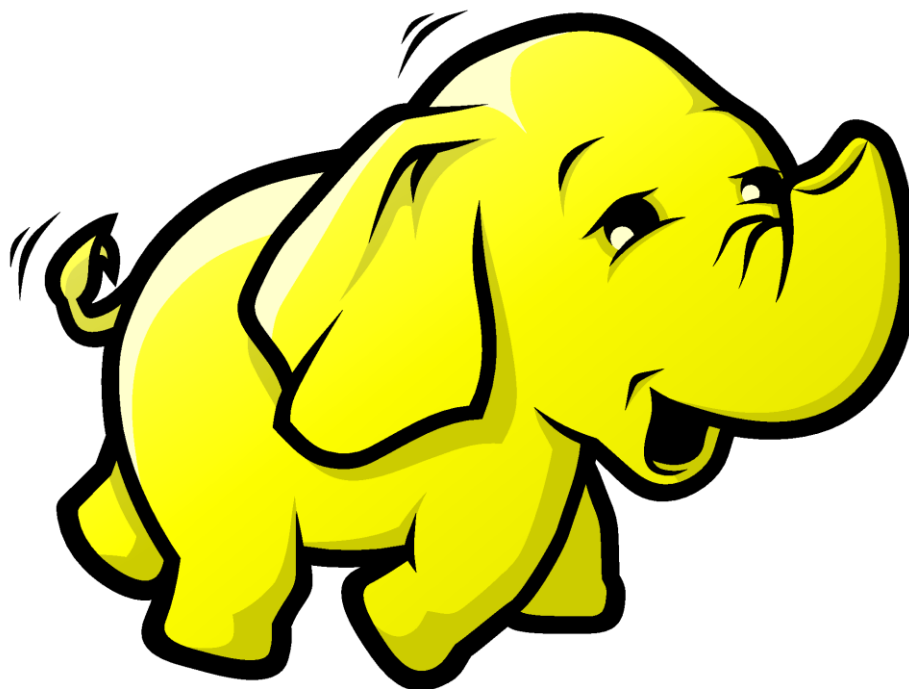
HDFS HA



YARN HA



Apache Hadoop的安装



Hadoop2.0+

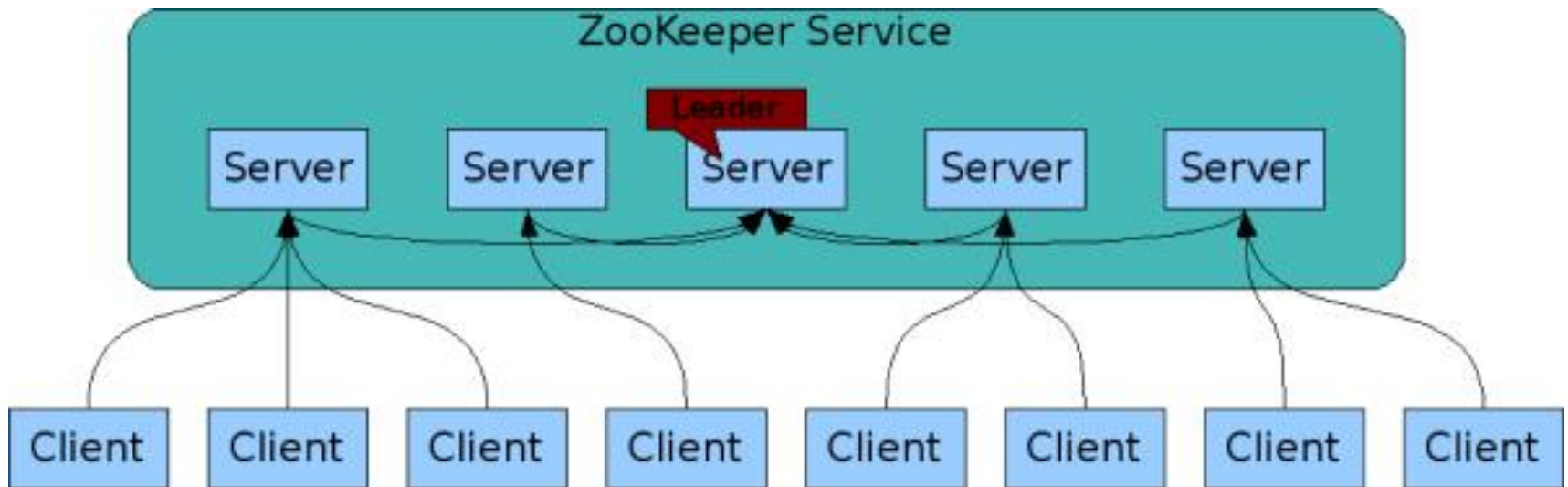
. HDFS

namenode(HA),
SecondaryNamenode,journalnode,datanode

.Yarn

ResourceManager,NodeManager

Zookeeper



<http://zookeeper.apache.org/doc/r3.5.0-alpha/zookeeperOver.html>

Hadoop HA安装

- core-site.xml
- hdfs-site.xml
- yarn-site.xml
- mapred-site.xml

<http://hadoop.apache.org/docs/r2.5.2/>

1、core-site.xml

```
<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://mycluster</value>
</property>

<property>
<name>hadoop.tmp.dir</name>
<value>/usr/local/hadoop-2.4.0/tmp</value>
</property>

<property>
<name>ha.zookeeper.quorum</name>
<value>itr-mastertest01:2181,itr-mastertest02:2181,itr-nodetest01:2181</value>
</property>

<property>
<name>fs.trash.interval</name>
<value>2000</value>
</property>
</configuration>
```

2、hdfs-site.xml

<configuration>

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

....

3、 yarn-site.xml

yarn-site.xml

```
<configuration>
```

```
<property>
```

```
  <name>yarn.resourcemanager.hostname</name>
```

```
  <value>itr-mastertest01</value>
```

```
</property>
```

```
<property>
```

```
  <name>yarn.nodemanager.aux-services</name>
```

```
  <value>mapreduce_shuffle</value>
```

```
</property>
```

```
</configuration>
```

4、mapred-site.xml

mapred-site.xml

```
<configuration>
```

```
  <property>
```

```
    <name>mapreduce.framework.name</name>
```

```
      <value>yarn</value>
```

```
    </property>
```

```
</configuration>
```

演示安装过程

HDFS | command

```
$ hadoop
Usage: hadoop [--config confdir] COMMAND
```

where COMMAND is one of:

```
fs                run a generic filesystem user client
version           print the version
jar <jar>         run a jar file
checknative [-a|-h] check native hadoop and compression libraries availability
distcp <srcurl> <desturl> copy file or directories recursively
archive -archiveName NAME -p <parent path> <src>* <dest> create a hadoop archive
classpath         prints the class path needed to get the
credential        interact with credential providers
                  Hadoop jar and the required libraries
daemonlog         get/set the log level for each daemon
```

or

```
CLASSNAME        run the class named CLASSNAME
```

```
$ hadoop version
Hadoop 2.5.0-cdh5.2.0
Subversion http://github.com/cloudera/hadoop -r e1f20a08bde76a33b79df026d00a0c91b2298387
Compiled by jenkins on 2014-10-11T21:00Z
Compiled with protoc 2.5.0
From source with checksum 309bccd135b199bdfdd6df5f3f4153d
```

```
$ hadoop fs
Usage: hadoop fs [generic options]
[-appendToFile <localsrc> ... <dst>]
[-cat [-ignoreCrc] <src> ...]
[-checksum <src> ...]
[-chgrp [-R] GROUP PATH...]
[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
[-chown [-R] [OWNER][:[GROUP]] PATH...]
[-copyFromLocal [-f] [-p] <localsrc> ... <dst>]
[-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
[-count [-q] <path> ...]
[-cp [-f] [-p | -p[topax]] <src> ... <dst>]
[-createSnapshot <snapshotDir> [<snapshotName>]]
[-deleteSnapshot <snapshotDir> <snapshotName>]
[-df [-h] [<path> ...]]
[-du [-s] [-h] <path> ...]
[-expunge]
[-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
```

HDFS | command jar

```
$ hadoop jar  
RunJar jarFile [mainClass] args...
```

提交一个jar程序到集群

```
hadoop jar hadoop-mapreduce-examples-2.4.0.jar wordcount /testdata /output
```


大数据



Thank you

提问时间?

Blog: <http://www.itweet.cn>

PPT: <https://github.com/itweet/course>

Video: <http://www.tudou.com/home/sparkjvm/>

