



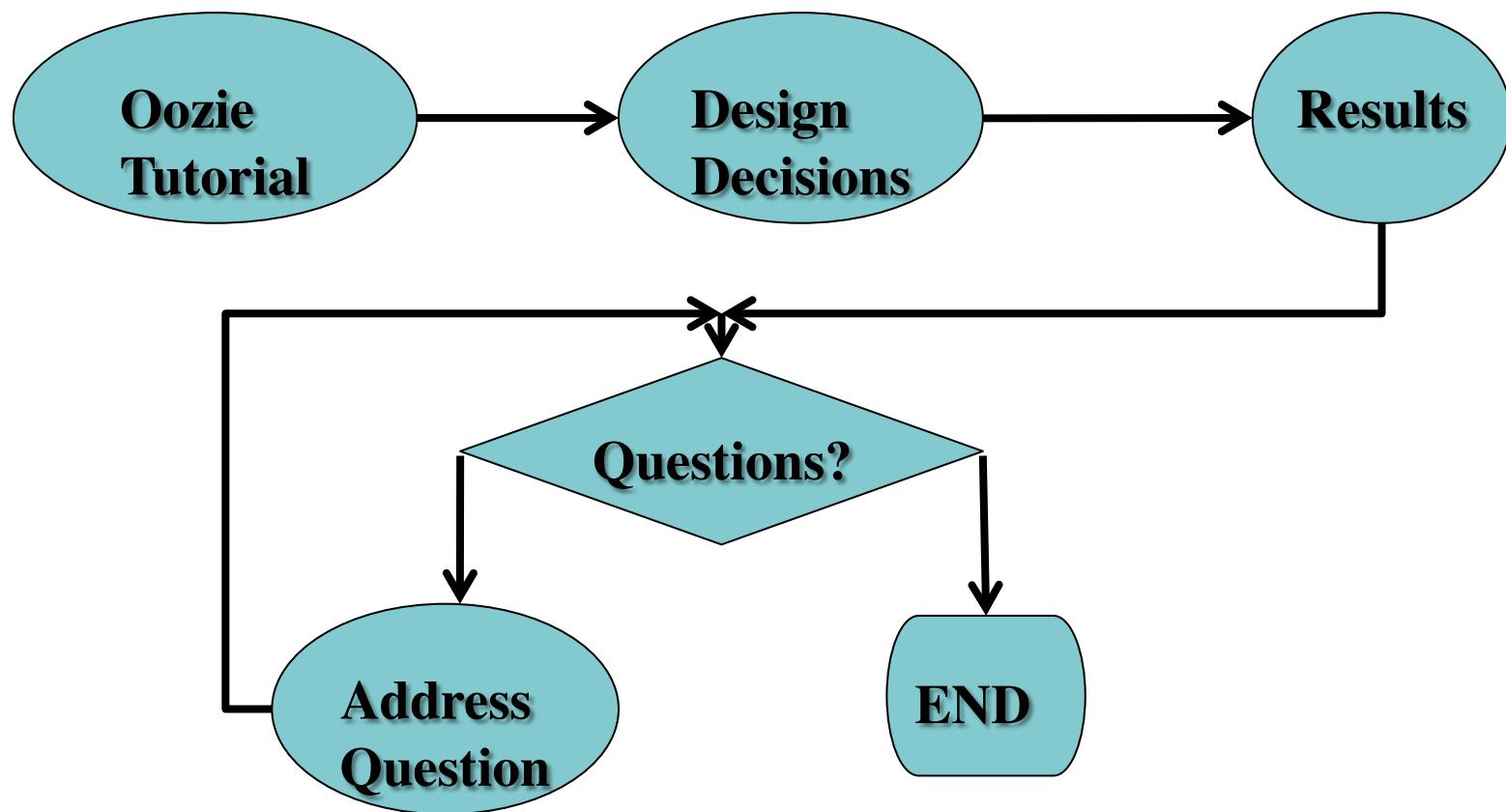
Oozie: A Workflow Scheduling For Hadoop

Mohammad Islam





Presentation Workflow



What is Hadoop?

- A framework for very large scale data processing in distributed environment
- Main Idea came from Google
- Implemented at Yahoo and open sourced to Apache.
- Hadoop is free!

YAHOO!



What is Hadoop? Contd.

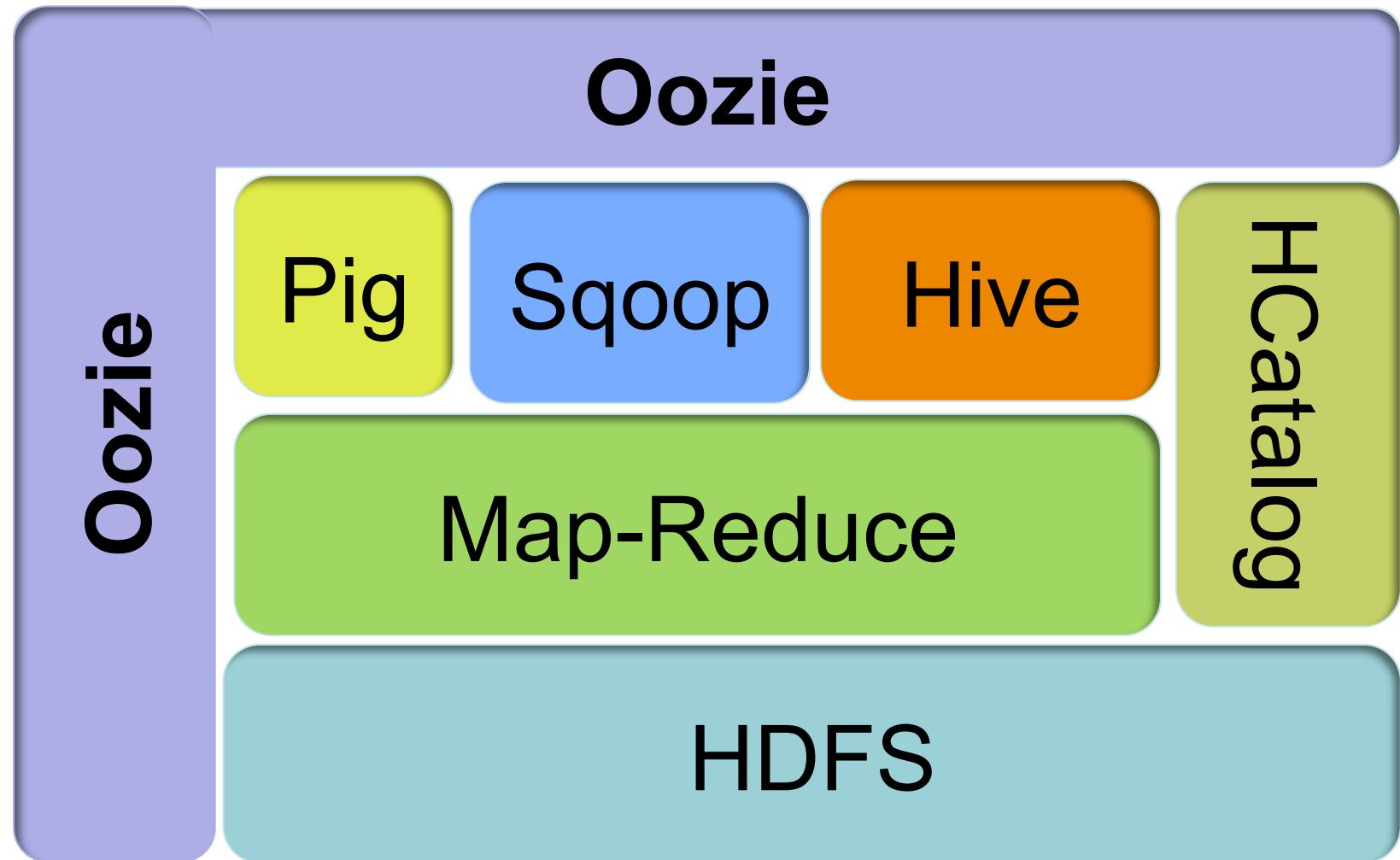
- Main components:
 - HDFS
 - Map-Reduce
- Highly scalable, fault-tolerant system
- Built on commodity hardware.

Yet Another WF System?

- Workflow management is a matured field.
- A lot of WF systems are already available
- Existing WF system could be *hacked* to use for Hadoop
- But no WF system built-for Hadoop.
- Hadoop has some benefits and shortcomings.
- Oozie was designed *only for* Hadoop considering those features



Oozie in Hadoop Eco-System



Oozie : The Conductor

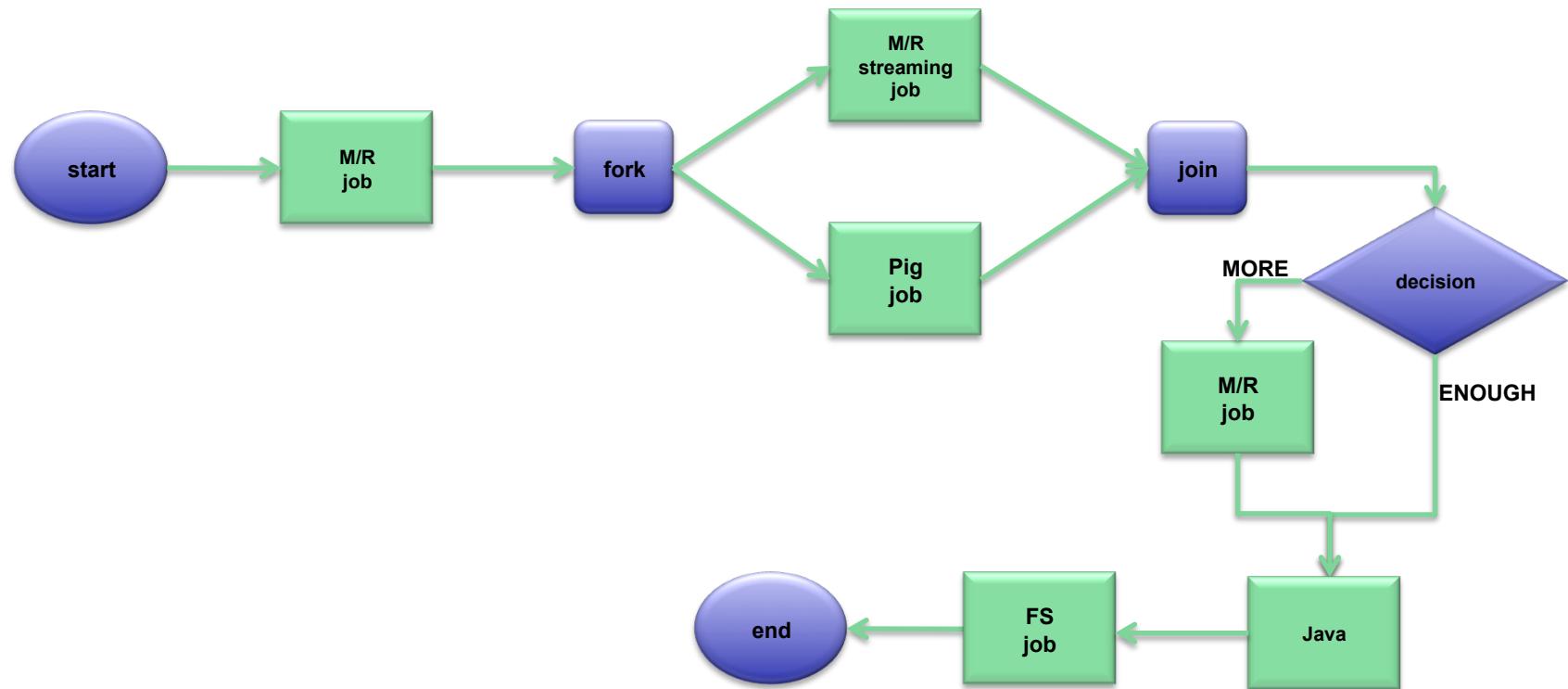


YAHOO!

Y!

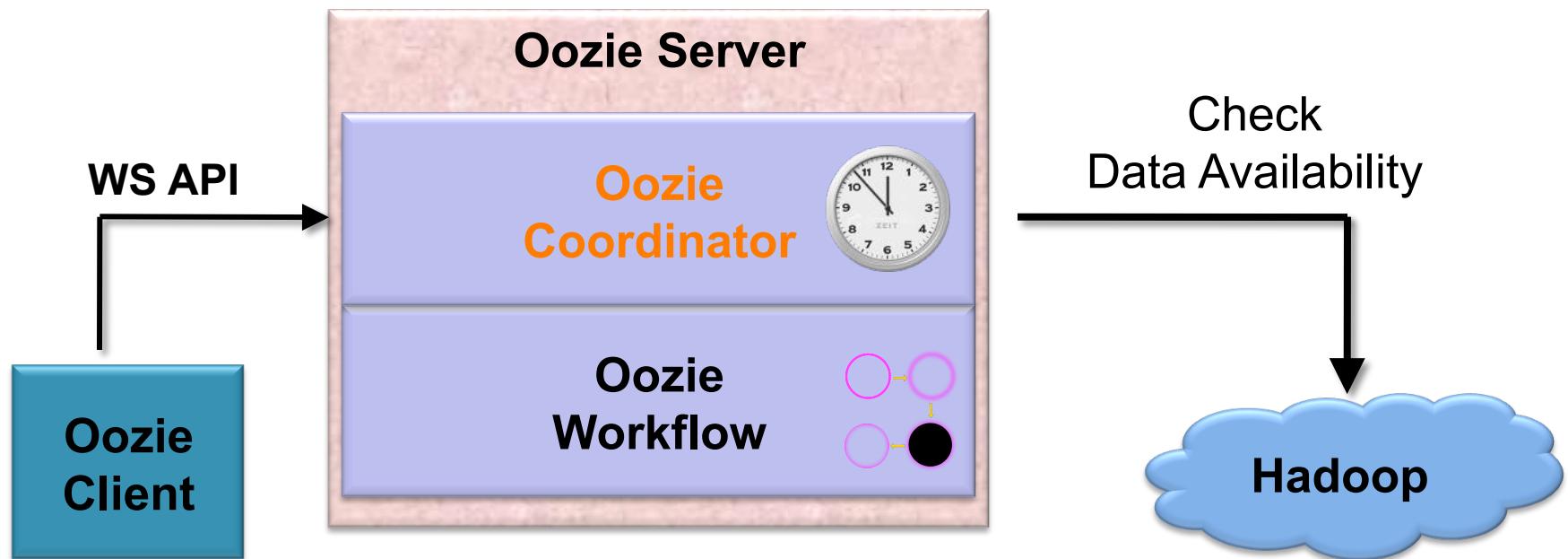
A Workflow Engine

- Oozie executes workflow defined as DAG of jobs
- The job type includes: Map-Reduce/Pig/Hive/Any script/Custom Java Code etc



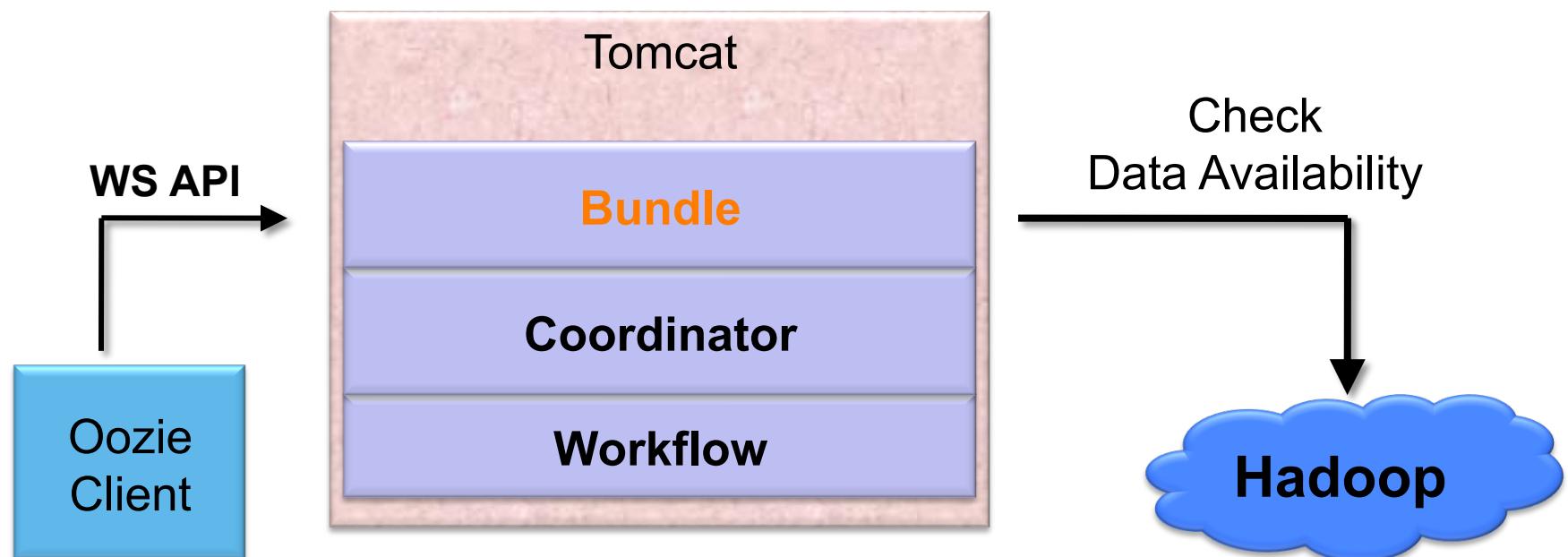
A Scheduler

- Oozie executes workflow based on:
 - Time Dependency (Frequency)
 - Data Dependency

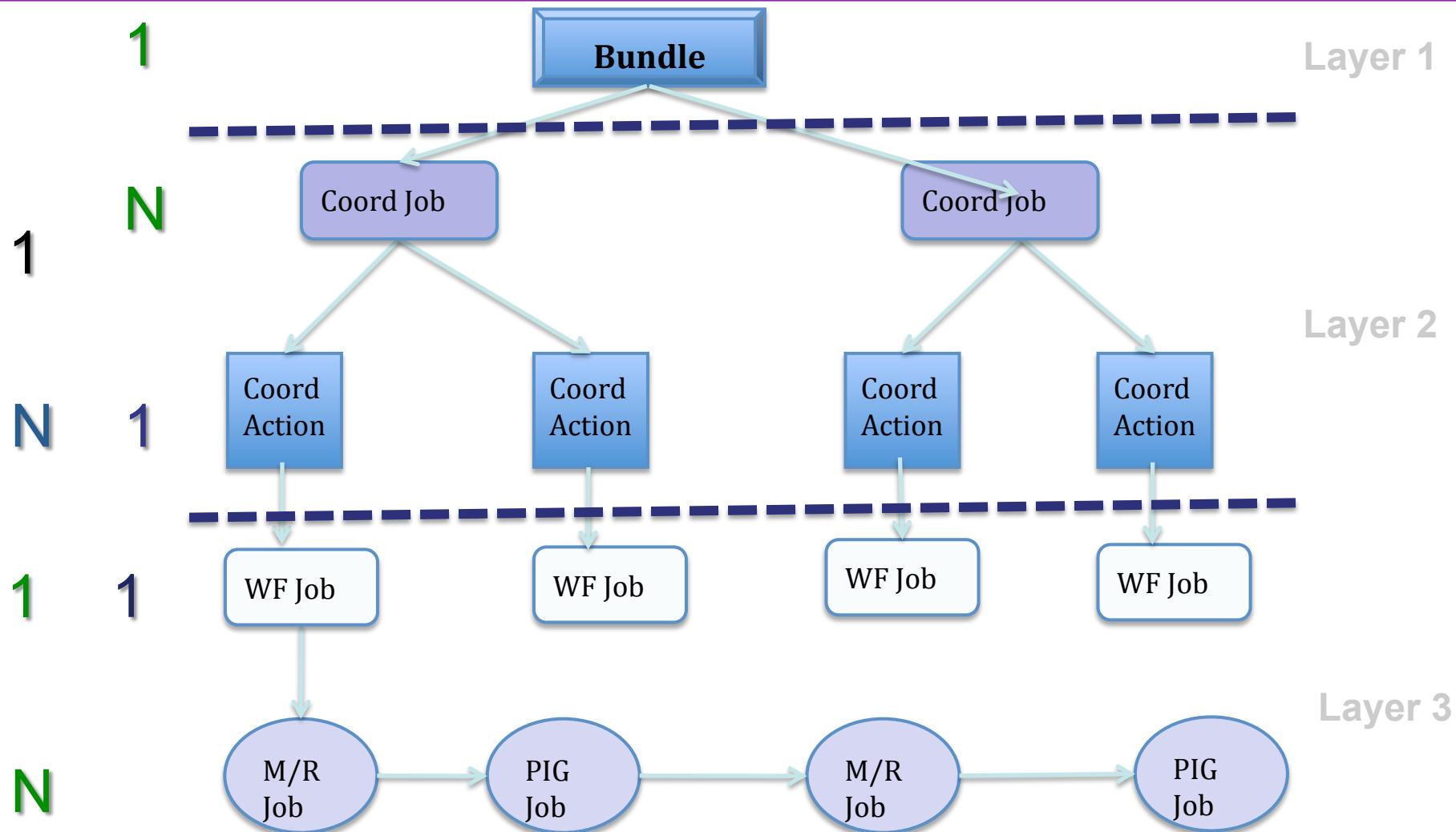


Bundle

- A new abstraction layer on top of Coordinator.
- Users can define and execute *a bunch of* coordinator applications.
- Bundle is *optional*



Oozie Abstraction Layers



YAHOO!



Access to Oozie Service

- Four different ways:
 - Using Oozie CLI client
 - Java API
 - REST API
 - Web Interface (read-only)

Installing Oozie

Step 1: Download the Oozie tarball

```
curl -O http://mirrors.sonic.net/apache/incubator/oozie/oozie-3.1.3-incubating/  
oozie-3.1.3-incubating-distro.tar.gz
```

Step 2: Unpack the tarball

```
tar -xzvf <PATH_TO_OOZIE_TAR>
```

Step 3: Run the setup script

```
bin/oozie-setup.sh -hadoop 0.20.200 ${HADOOP_HOME} -extjs /tmp/ext-2.2.zip
```

Step 4: Start oozie

```
bin/oozie-start.sh
```

Step 5: Check status of oozie

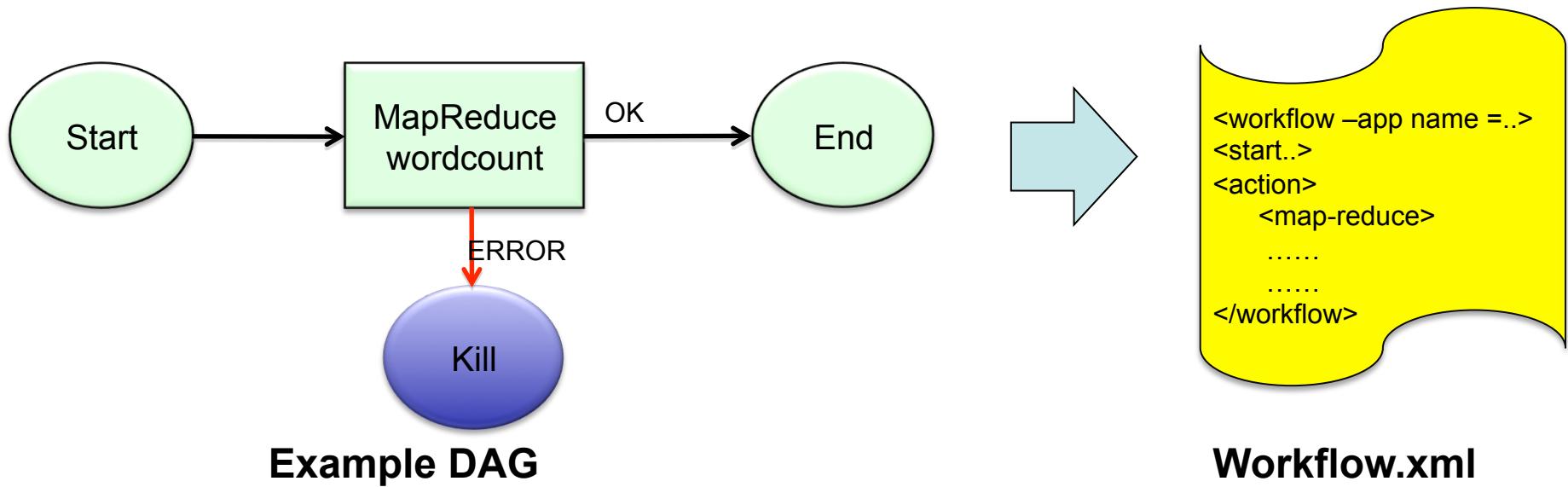
```
bin/oozie admin -oozie http://localhost:11000/oozie -status
```

Running an Example

- Standalone Map-Reduce job

```
$ hadoop jar /usr/joe/hadoop-examples.jar org.myorg.wordcount inputDir outputDir
```

- Using Oozie



Example Workflow

```
<action name='wordcount'>
  <map-reduce>
    <configuration>
      <property>
        <name>mapred.mapper.class</name>
        <value>org.myorg.WordCount.Map</value>
      </property>
      <property>
        <name>mapred.reducer.class</name>
        <value>org.myorg.WordCount.Reduce</value>
      </property>
      <property>
        <name>mapred.input.dir</name>
        <value>usr/joe/inputDir </value>
      </property>
      <property>
        <name>mapred.output.dir</name>
        <value>/usr/joe/outputDir</value>
      </property>
    </configuration>
  </map-reduce>
</action>
```

mapred.mapper.class =
org.myorg.WordCount.Map

mapred.reducer.class =
org.myorg.WordCount.Reduce

mapred.input.dir = inputDir

mapred.output.dir = outputDir

A Workflow Application

Three components required for a Workflow:

1) Workflow.xml:

Contains job definition

2) Libraries:

optional 'lib/' directory contains .jar/.so files

3) Properties file:

- Parameterization of Workflow xml
- Mandatory property is *oozie.wf.application.path*

Workflow Submission

Deploy Workflow to HDFS

```
$ hadoop fs –put wf_job hdfs://bar.com:9000/usr/abc/wf_job
```

Run Workflow Job

```
$ oozie job –run -config job.properties -oozie http://localhost:11000/oozie/  
Workflow ID: 00123-123456-oozie-wrkf-W
```

Check Workflow Job Status

```
$ oozie job –info 00123-123456-oozie-wrkf-W -oozie http://localhost:11000/  
oozie/
```

Oozie Web Console

 Apache Documentation [Yahoo Documentation](#)

| Oozie Web Console (v1) [/oozie/] | | | | | | | | | | |
|--|--------------------|------------------|-----|-------------|-------|-------------------------------|-------------------------------|------------------------------|---------------|--|
| Workflow Jobs | | Coordinator Jobs | | Bundle Jobs | | System Info | | Instrumentation | | |
| All Jobs Active Jobs Done Jobs Custom Filter ▾ | | | | | | | | | | |
| Job Id | Name | Status | ... | User | Group | Created | Started | | Last Modified | |
| 1 0000129-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 23:24:45 GMT | Thu, 17 May 2012 23:24:45 GMT | Thu, 17 May 2012 23:25:46 GM | | |
| 2 0000127-120516024608778-oozie... | StarlingProcessor | RUNNING | 0 | gmon | users | Thu, 17 May 2012 23:24:00 GMT | Thu, 17 May 2012 23:24:01 GMT | Thu, 17 May 2012 23:30:01 GM | | |
| 3 0000126-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 22:24:45 GMT | Thu, 17 May 2012 22:24:46 GMT | Thu, 17 May 2012 22:25:28 GM | | |
| 4 0000124-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 21:24:44 GMT | Thu, 17 May 2012 21:24:44 GMT | Thu, 17 May 2012 21:25:38 GM | | |
| 5 0000122-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 20:24:44 GMT | Thu, 17 May 2012 20:24:44 GMT | Thu, 17 May 2012 20:25:24 GM | | |
| 6 0000120-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 19:24:45 GMT | Thu, 17 May 2012 19:24:45 GMT | Thu, 17 May 2012 19:25:26 GM | | |
| 7 0000118-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 18:24:44 GMT | Thu, 17 May 2012 18:24:44 GMT | Thu, 17 May 2012 18:25:35 GM | | |
| 8 0000116-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 17:24:44 GMT | Thu, 17 May 2012 17:24:45 GMT | Thu, 17 May 2012 17:25:31 GM | | |
| 9 0000114-120516024608778-oozie... | StarlingProcessor | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 17:24:00 GMT | Thu, 17 May 2012 17:24:00 GMT | Thu, 17 May 2012 22:39:58 GM | | |
| 10 0000113-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 16:24:45 GMT | Thu, 17 May 2012 16:24:45 GMT | Thu, 17 May 2012 16:25:42 GM | | |
| 11 0000111-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 15:24:44 GMT | Thu, 17 May 2012 15:24:44 GMT | Thu, 17 May 2012 15:25:38 GM | | |
| 12 0000109-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 14:24:44 GMT | Thu, 17 May 2012 14:24:44 GMT | Thu, 17 May 2012 14:25:35 GM | | |
| 13 0000107-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 13:24:44 GMT | Thu, 17 May 2012 13:24:44 GMT | Thu, 17 May 2012 13:25:23 GM | | |
| 14 0000105-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 12:24:44 GMT | Thu, 17 May 2012 12:24:45 GMT | Thu, 17 May 2012 12:25:39 GM | | |
| 15 0000103-120516024608778-oozie... | hive-yflow-proc... | SUCCEE... | 0 | amdgard_1 | users | Thu, 17 May 2012 12:00:00 GMT | Thu, 17 May 2012 12:00:00 GMT | Thu, 17 May 2012 12:54:55 GM | | |
| 16 0000102-120516024608778-oozie... | rise-dni-gcc4_6... | KILLED | 0 | rise_scm | users | Thu, 17 May 2012 11:50:46 GMT | Thu, 17 May 2012 11:50:47 GMT | Thu, 17 May 2012 11:50:47 GM | | |
| 17 0000101-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 11:24:44 GMT | Thu, 17 May 2012 11:24:45 GMT | Thu, 17 May 2012 11:25:31 GM | | |
| 18 0000099-120516024608778-oozie... | StarlingProcessor | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 11:24:00 GMT | Thu, 17 May 2012 11:24:01 GMT | Thu, 17 May 2012 14:15:34 GM | | |
| 19 0000098-120516024608778-oozie... | map-reduce-wf | SUCCEE... | 0 | gmon | users | Thu, 17 May 2012 10:24:46 GMT | Thu, 17 May 2012 10:24:46 GMT | Thu, 17 May 2012 10:25:28 GM | | |

Oozie Web Console: Job Details

Job (Name: map-reduce-wf/JobId: 0000129-120516024608778-oozie-wrkf-W)

Job Info Job Definition Job Configuration Job Log

Job Id: 0000129-120516024608778-oozie
Name: map-reduce-wf
App Path: hdfs://axoniteblue-nn1.blue.ygrid.y
Run: 0
Status: SUCCEEDED
User: gmon
Group: users
Create Time: Thu, 17 May 2012 23:24:45 GMT
Nominal Time:
Start Time: Thu, 17 May 2012 23:24:45 GMT
Last Modified: Thu, 17 May 2012 23:25:46 GMT
End Time: Thu, 17 May 2012 23:25:46 GMT

Actions

| Action Id | Name | Type | Status | Transition | StartTime | EndTime |
|--|---------|------------|--------|------------|--------------------------------|--------------------------------|
| 1 0000129-120516024608778-oozie-wrkf-W@... | hadoop1 | map-reduce | OK | end | Thu, 17 May 2012 23:24:46 G... | Thu, 17 May 2012 23:25:46 G... |

Oozie Web Console: Action Details

Job (Name: map-reduce-wf/JobId: 0000129-120516024608778-oozie-wrkf-W)

Job Info Job Definition Job Configuration Job Log

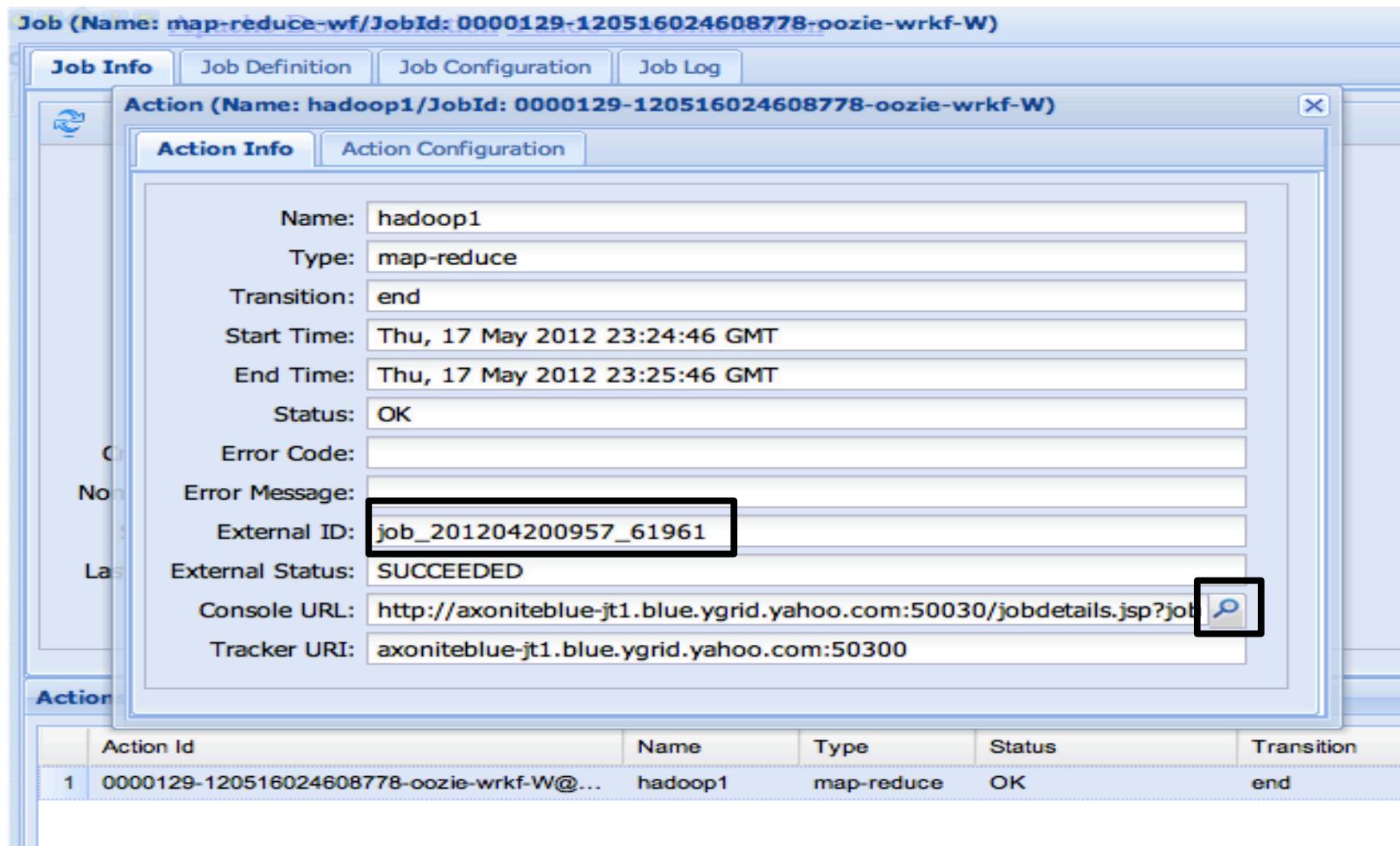
Action (Name: hadoop1/JobId: 0000129-120516024608778-oozie-wrkf-W)

Action Info Action Configuration

| | |
|------------------|---|
| Name: | hadoop1 |
| Type: | map-reduce |
| Transition: | end |
| Start Time: | Thu, 17 May 2012 23:24:46 GMT |
| End Time: | Thu, 17 May 2012 23:25:46 GMT |
| Status: | OK |
| Error Code: | |
| Error Message: | |
| External ID: | job_201204200957_61961 |
| External Status: | SUCCEEDED |
| Console URL: | http://axoniteblue-jt1.blue.ygrid.yahoo.com:50030/jobdetails.jsp?jobid=0000129-120516024608778-oozie-wrkf-W&action=hadoop1 |
| Tracker URI: | axoniteblue-jt1.blue.ygrid.yahoo.com:50300 |

Action

| Action Id | Name | Type | Status | Transition |
|--|---------|------------|--------|------------|
| 0000129-120516024608778-oozie-wrkf-W@... | hadoop1 | map-reduce | OK | end |



Hadoop Job Details

Hadoop Job job_201204200957_61961 on [History Viewer](#)

User: gmon

JobName: oozie:action:T=map-reduce:W=map-reduce-wf:A=hadoop1:ID=0000129-120516024608778-oozie-wrkf-W

JobConf: [/user/gmon/staging/job_201204200957_61961/job.xml](#)

Job-ACLs:

mapreduce.job.acl-view-job: No users are allowed

mapreduce.job.acl-modify-job: Users [users] are allowed

Submitted At: 17-May-2012 23:25:04

Launched At: 17-May-2012 23:25:11 (7sec)

Finished At: 17-May-2012 23:25:46 (34sec)

Status: SUCCESS

Failure Info:

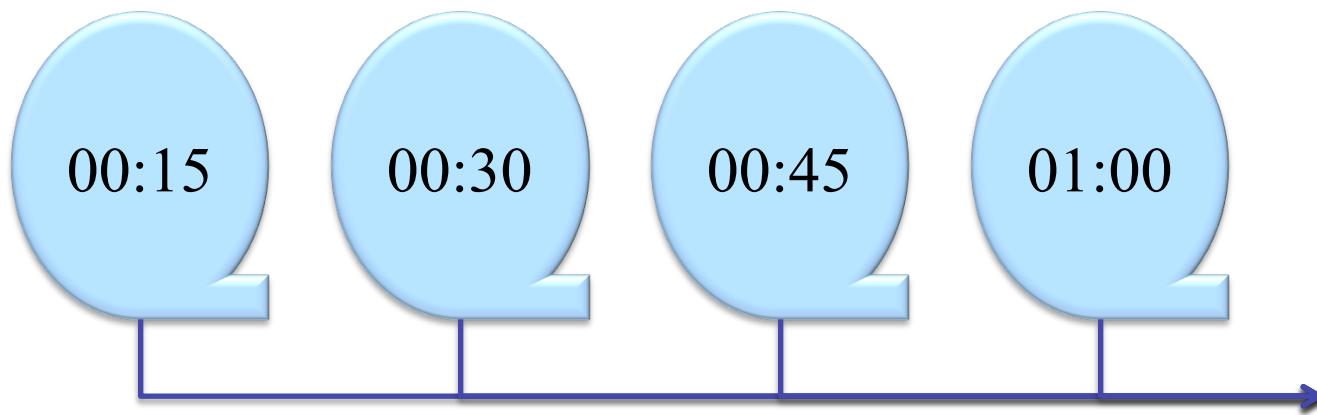
[Analyse This Job](#)

| Kind | Total Tasks(successful+failed+killed) | Successful tasks | Failed tasks | Killed tasks | Start Time | Finish Time |
|---------|---------------------------------------|-------------------|-------------------|-------------------|----------------------|------------------------------|
| Setup | 1 | 1 | 0 | 0 | 17-May-2012 23:25:16 | 17-May-2012 23:25:21 (4sec) |
| Map | 2 | 2 | 0 | 0 | 17-May-2012 23:25:21 | 17-May-2012 23:25:30 (9sec) |
| Reduce | 1 | 1 | 0 | 0 | 17-May-2012 23:25:25 | 17-May-2012 23:25:41 (16sec) |
| Cleanup | 1 | 1 | 0 | 0 | 17-May-2012 23:25:41 | 17-May-2012 23:25:46 (4sec) |

| | Counter | Map | Reduce | Total |
|----------------------------|--|-------|--------|--------|
| Job Counters | Launched reduce tasks | 0 | 0 | 1 |
| | SLOTS_MILLIS_MAPS | 0 | 0 | 17,801 |
| | Total time spent by all reduces waiting after reserving slots (ms) | 0 | 0 | 0 |
| | Total time spent by all maps waiting after reserving slots (ms) | 0 | 0 | 0 |
| | Launched map tasks | 0 | 0 | 2 |
| | SLOTS_MILLIS_REDUCES | 0 | 0 | 32,728 |
| File Input Format Counters | Bytes Read | 1,464 | 0 | 1,464 |

Use Case : Time Triggers

- Execute your workflow every 15 minutes (*CRON*)

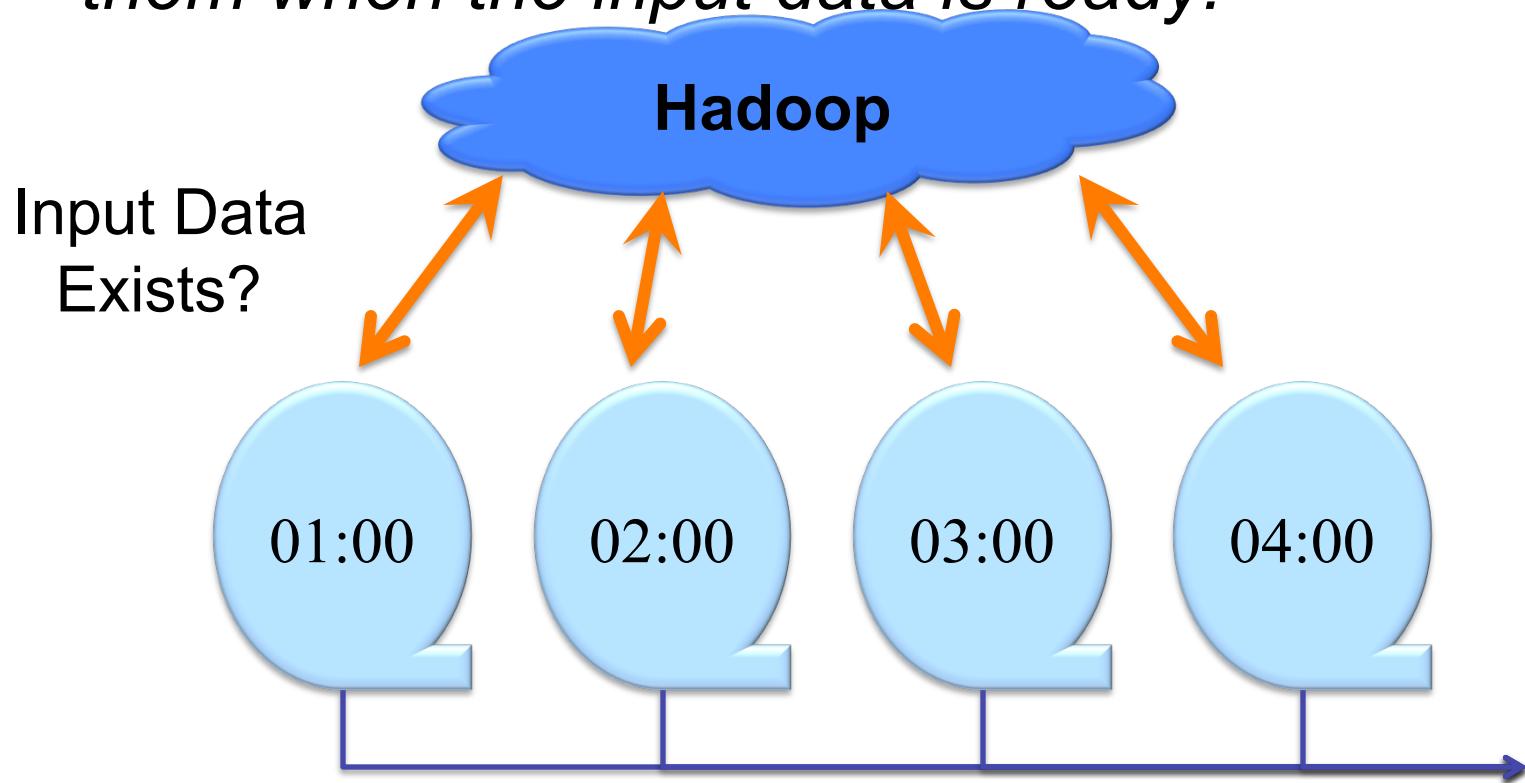


Run Workflow every 15 mins

```
<coordinator-app name="Clicks15Mins" start="2012-01-01T00:00Z"  
    end="2012-12-31T00:00Z" frequency="15">   
    <action>  
        <workflow>  
            <app-path>hdfs://bar:9000/usr/abc/logsprocessor-wf</app-path>  
            <configuration>  
                <property> <name>key1</name><value>value1</value> </property>  
            </configuration>  
        </workflow>  
    </action>  
</coordinator-app>
```

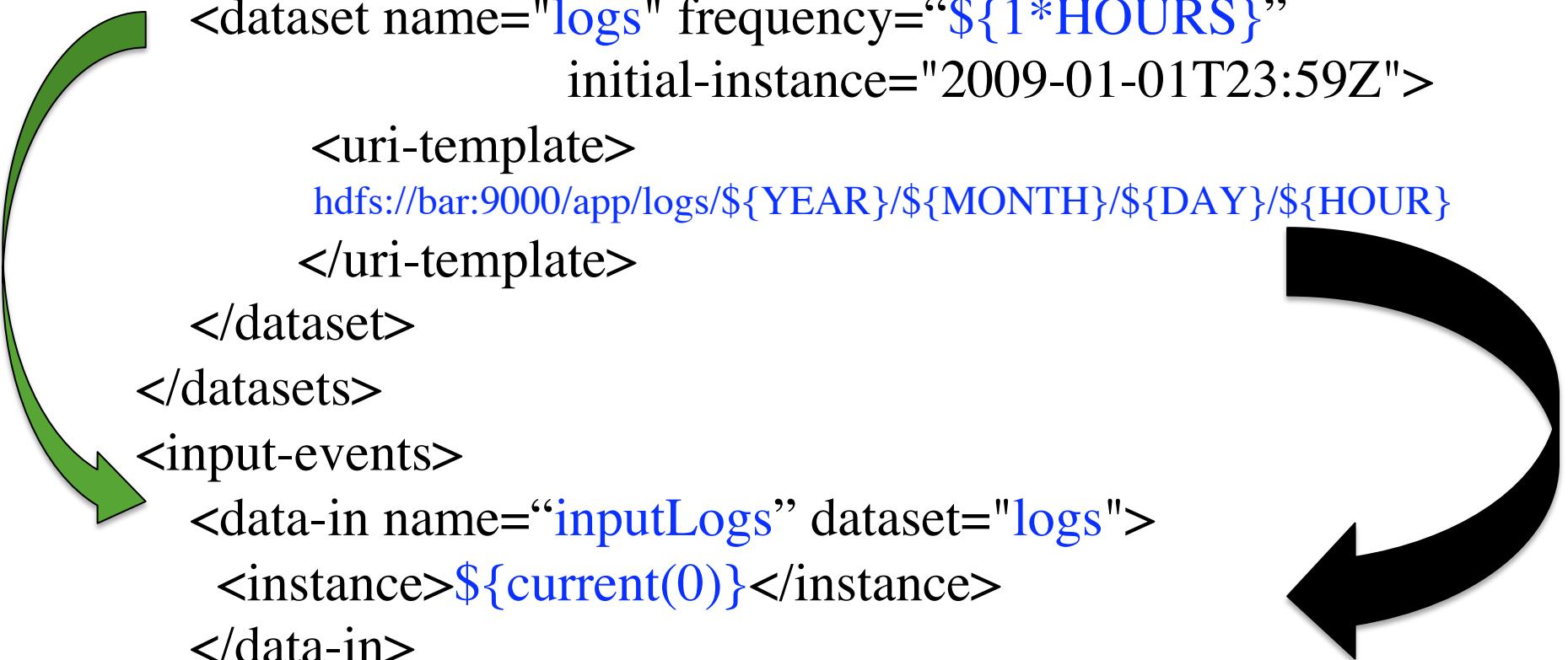
Use Case: Time and Data Triggers

- Execute your workflow every hour, *but only run them when the input data is ready.*



Data Triggers

```
<coordinator-app name="coord1" frequency="${1*HOURS}"...>
  <datasets>
    <dataset name="logs" frequency="${1*HOURS}"
            initial-instance="2009-01-01T23:59Z">
      <uri-template>
        hdfs://bar:9000/app/logs/${YEAR}/${MONTH}/${DAY}/${HOUR}
      </uri-template>
    </dataset>
  </datasets>
  <input-events>
    <data-in name="inputLogs" dataset="logs">
      <instance>${current(0)}</instance>
    </data-in>
  </input-events>
</coordinator-app>
```



Running an Coordinator Job

Application Deployment

```
$ hadoop fs –put coord_job hdfs://bar.com:9000/usr/abc/coord_job
```

Coordinator Job Parameters:

```
$ cat job.properties
```

```
oozie.coord.application.path = hdfs://bar.com:9000/usr/abc/coord_job
```

Job Submission

```
$ oozie job –run -config job.properties
```

```
job: 000001-20090525161321-oozie-xyz-C
```

Debugging an Coordinator Job

Coordinator Job Information

```
$ oozie job -info 000001-20090525161321-oozie-xyz-C
```

Job Name : wordcount-coord

App Path : hdfs://bar.com:9000/usr/abc/coord_job

Status : RUNNING

Coordinator Job Log

```
$ oozie job -log 000001-20090525161321-oozie-xyz-C
```

Coordinator Job Definition

```
$ oozie job -definition 000001-20090525161321-oozie-xyz-C
```

Three Questions ...

Do you need Oozie?

Q1 : Do you have *multiple* jobs with dependency?

Q2 : Does your job start based on *time* or *data* availability?

Q3 : Do you need monitoring and operational support for your jobs?

If *any one* of your answers is YES,
then you should consider Oozie!



What Oozie is NOT

- Oozie is not a resource scheduler
- Oozie is not for off-grid scheduling
 - Note: Off-grid *execution* is possible through SSH action.
- If you want to submit your job occasionally, Oozie is NOT a must.
 - Oozie provides REST API based submission.

Oozie in Apache

Main Contributors

YAHOO!®

eBay®

cloudera

IBM

YAHOO!

Y!



Oozie in Apache

- Y! internal usages:
 - Total number of user : **375**
 - Total number of processed jobs **≈ 750K/month**
- External downloads:
 - **2500+** in last year from GitHub
 - A large number of downloads maintained by 3rd party packaging.

Oozie Usages Contd.

- User Community:
 - Membership
 - Y! internal - **286**
 - External – **163**
 - No of Messages (approximate)
 - Y! internal – **7/day**
 - External – **10+/day**



Oozie: Towards a Scalable Workflow Management System for Hadoop

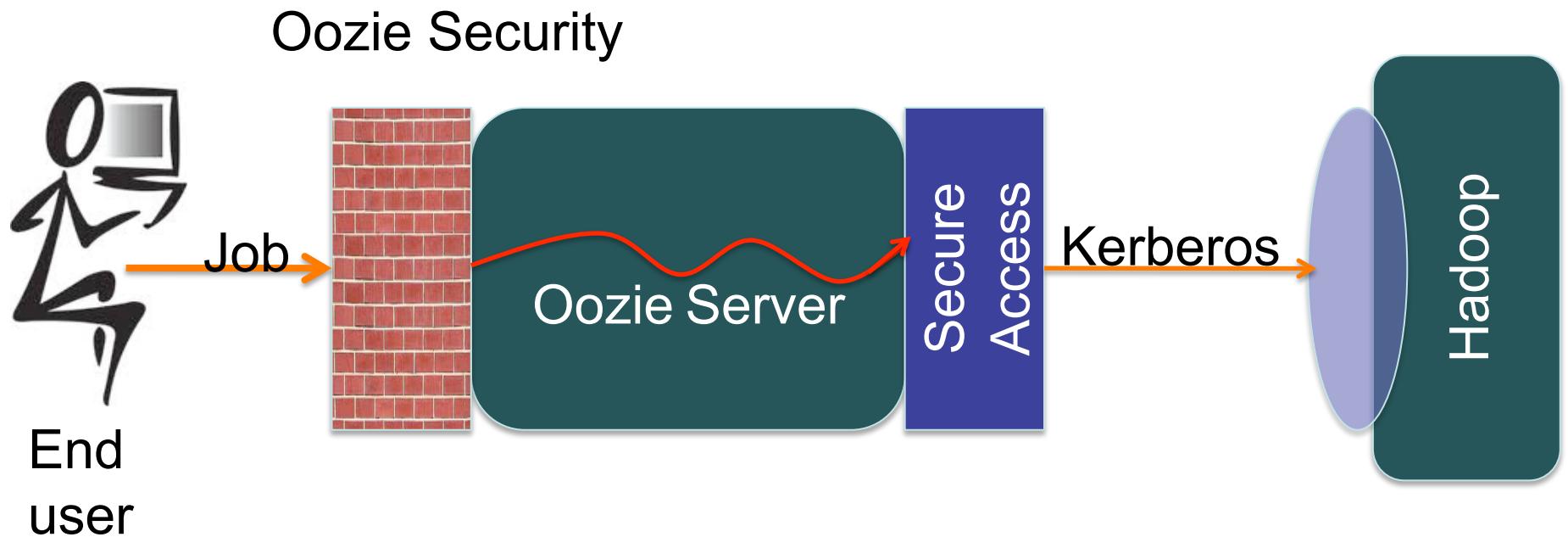
Mohammad Islam



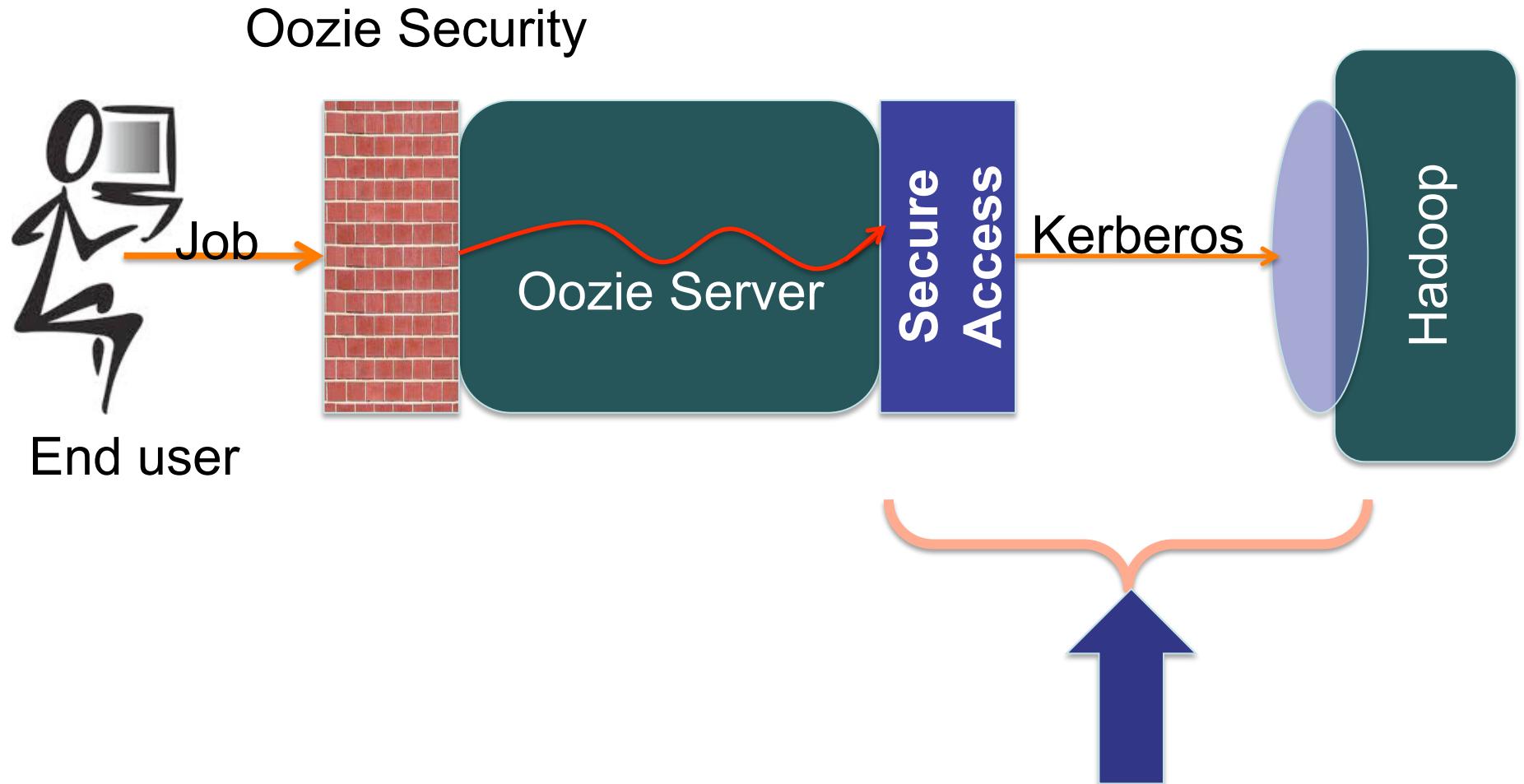
Key Features and Design Decisions

- Multi-tenant
- Security
 - Authenticate every request
 - Pass appropriate token to Hadoop job
- Scalability
 - Vertical: Add extra memory/disk
 - Horizontal: Add machines

Oozie Job Processing



Oozie-Hadoop Security



YAHOO!



Oozie-Hadoop Security

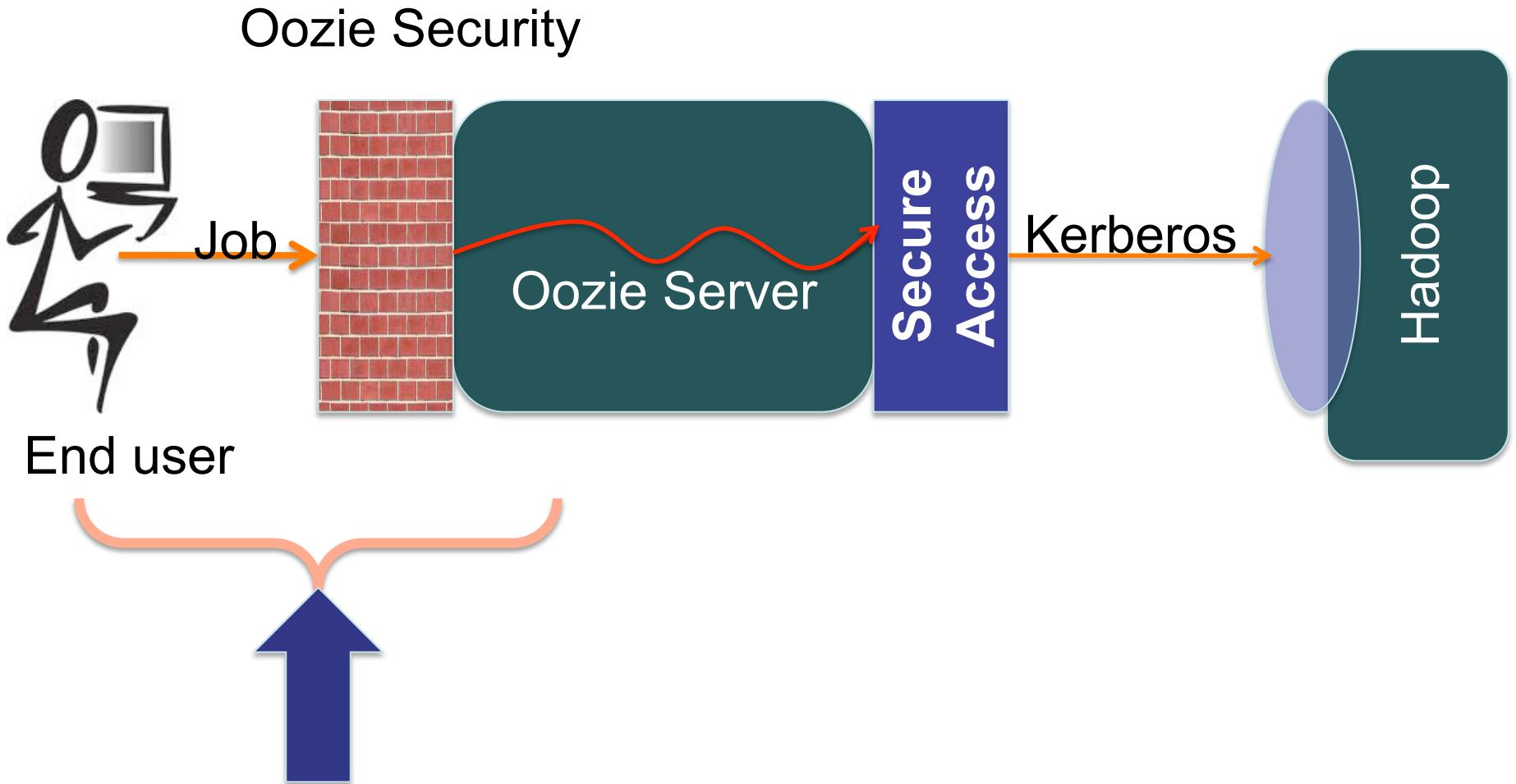
- Oozie is a multi-tenant system
- Job can be scheduled to run *later*
- Oozie submits/maintains the hadoop jobs
- Hadoop needs *security token* for each request

Question: Who should provide the security token to hadoop and how?

Oozie-Hadoop Security Contd.

- **Answer:** Oozie
- How?
 - Hadoop considers *Oozie* as a **super-user**
 - Hadoop does not check *end-user* credential
 - Hadoop only checks the credential of *Oozie* process
- BUT hadoop job is executed as *end-user*.
- Oozie utilizes doAs() functionality of Hadoop.

User-Oozie Security



Why Oozie Security?

- One user should *not modify* another user's job
- Hadoop doesn't authenticate end-user
- Oozie *has to verify* its user before passing the job to Hadoop

How does Oozie Support Security?

- Built-in authentication
 - Kerberos
 - Non-secured (default)
- Design Decision
 - Pluggable authentication
 - Easy to include new type of authentication
 - Yahoo supports 3 types of authentication.

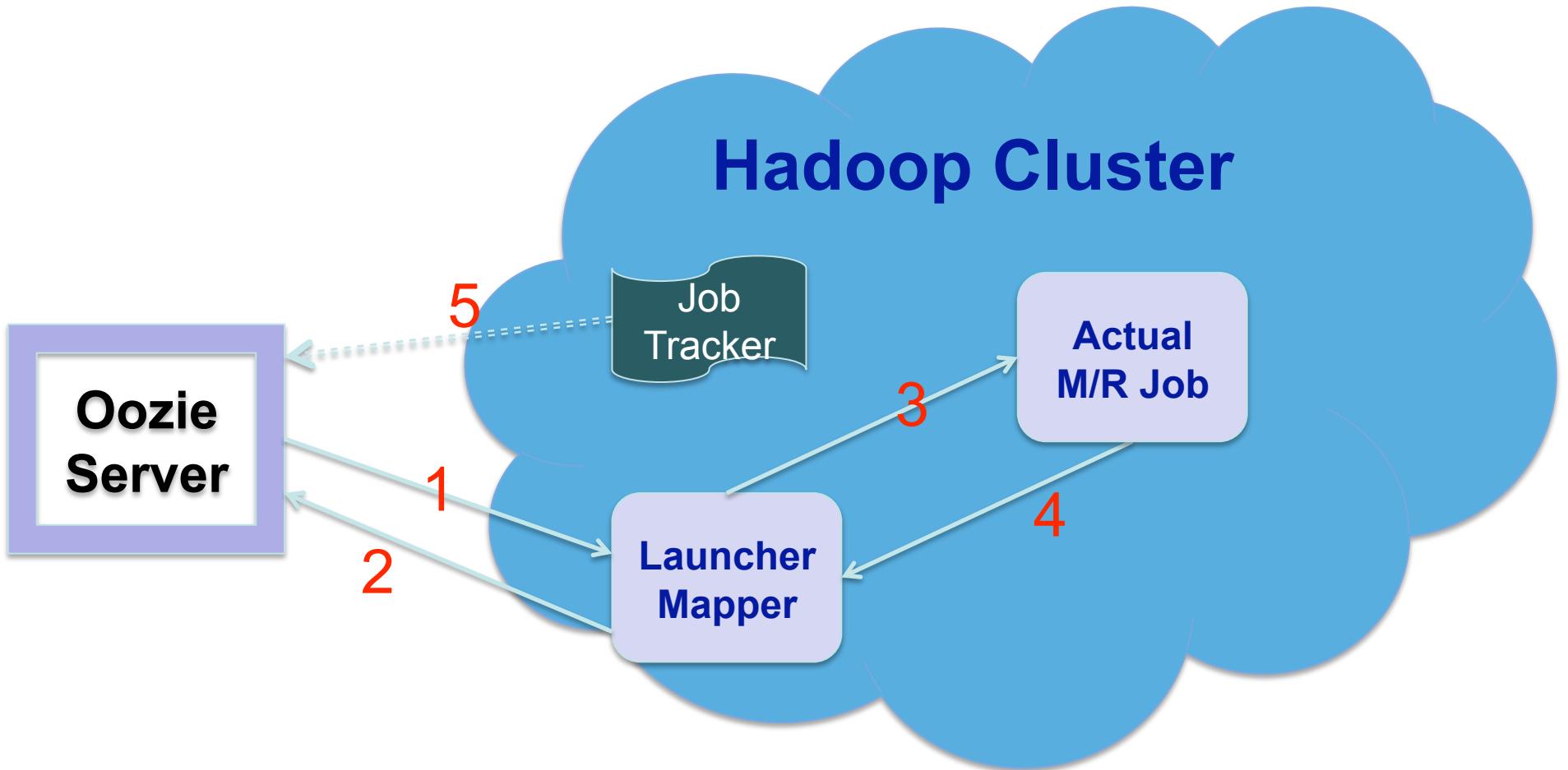
Job Submission to Hadoop

- Oozie is designed to handle thousands of jobs at the same time
- **Question :** Should Oozie server
 - Submit the hadoop job *directly*?
 - Wait for it to finish?
- **Answer:** No

Job Submission Contd.

- **Reason**
 - **Resource constraints:** A single Oozie process can't simultaneously create thousands of threads for each hadoop job. (Scaling limitation)
 - **Isolation:** Running user code on Oozie server might de-stabilize Oozie
- **Design Decision**
 - Create a launcher hadoop job
 - Execute the actual user job from the launcher.
 - Wait *asynchronously* for the job to finish.

Job Submission to Hadoop



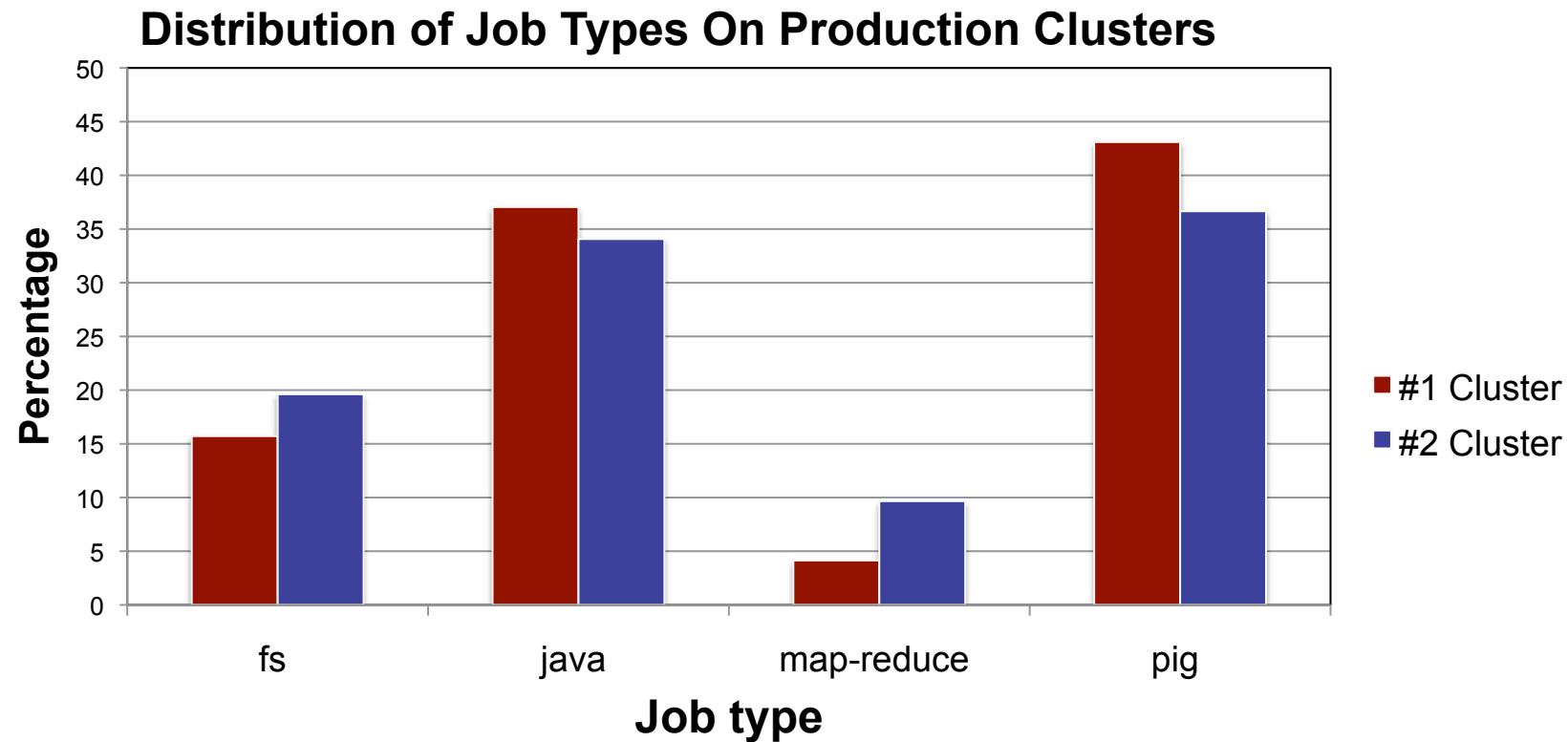
Job Submission Contd.

- **Advantages**
 - Horizontal scalability: If load increases, add machines into Hadoop cluster
 - Stability: Isolation of user code and system process
- **Disadvantages**
 - Extra map-slot is occupied by each job.

Production Setup

- Total number of nodes: 42K+
- Total number of Clusters: 25+
- Data presented from *two* clusters
- Each of them have nearly 4K nodes
- Total number of users /cluster = 50

Oozie Usage Pattern @ Y!

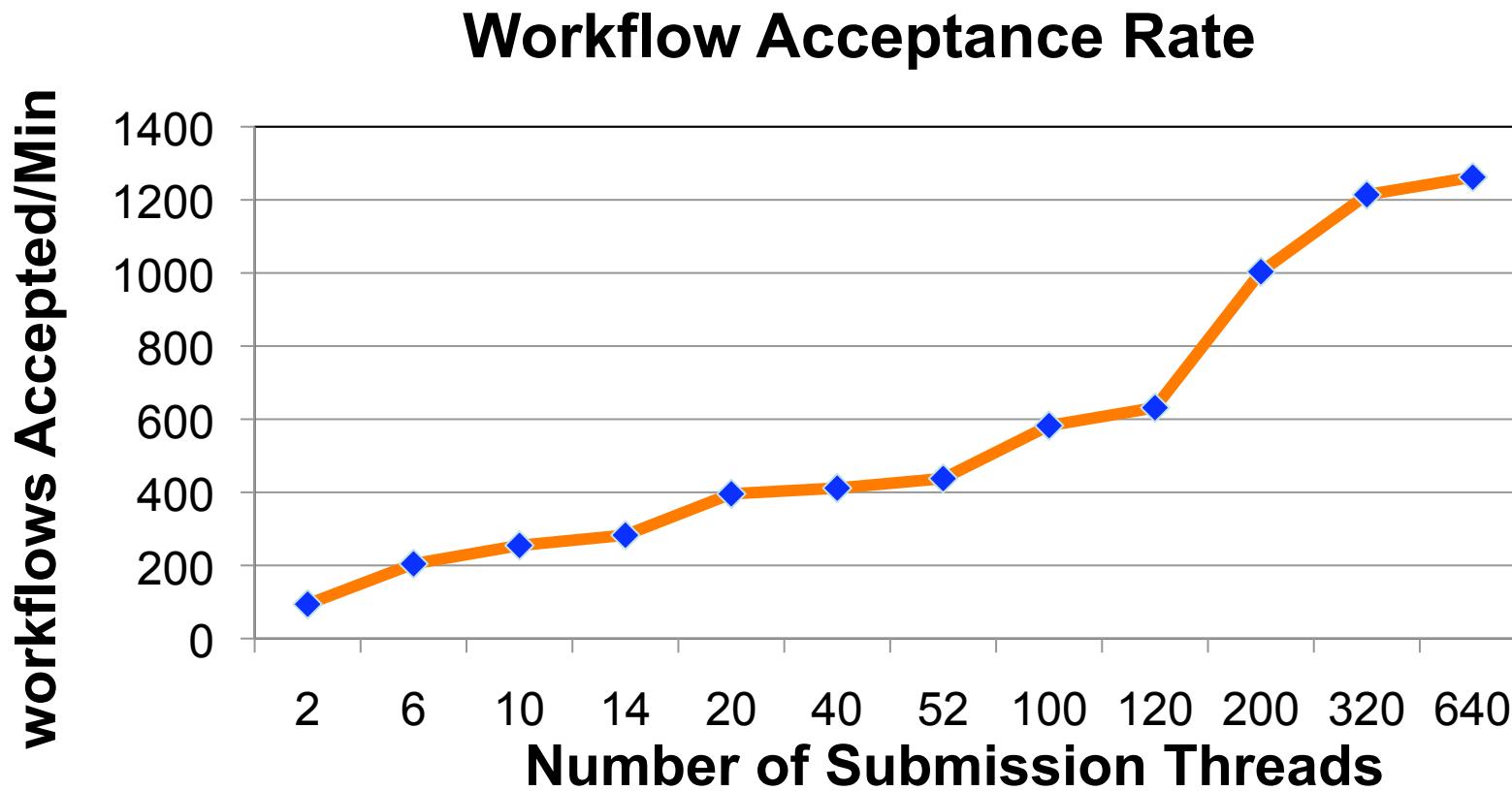


- Pig and Java are the most popular
- Number of *pure* Map-Reduce jobs are fewer

Experimental Setup

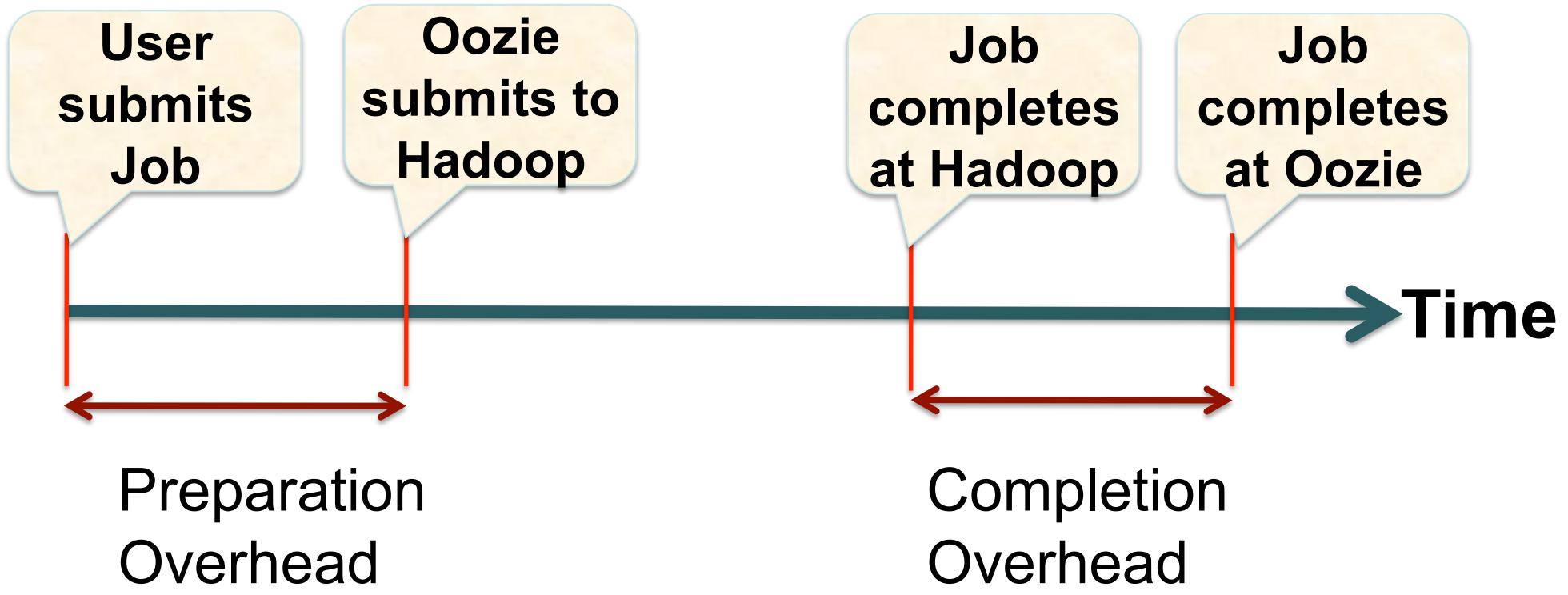
- Number of nodes: 7
- Number of map-slots: 28
- 4 Core, RAM: 16 GB
- 64 bit RHEL
- Oozie Server
 - 3 GB RAM
 - Internal Queue size = 10 K
 - # Worker Thread = 300

Job Acceptance



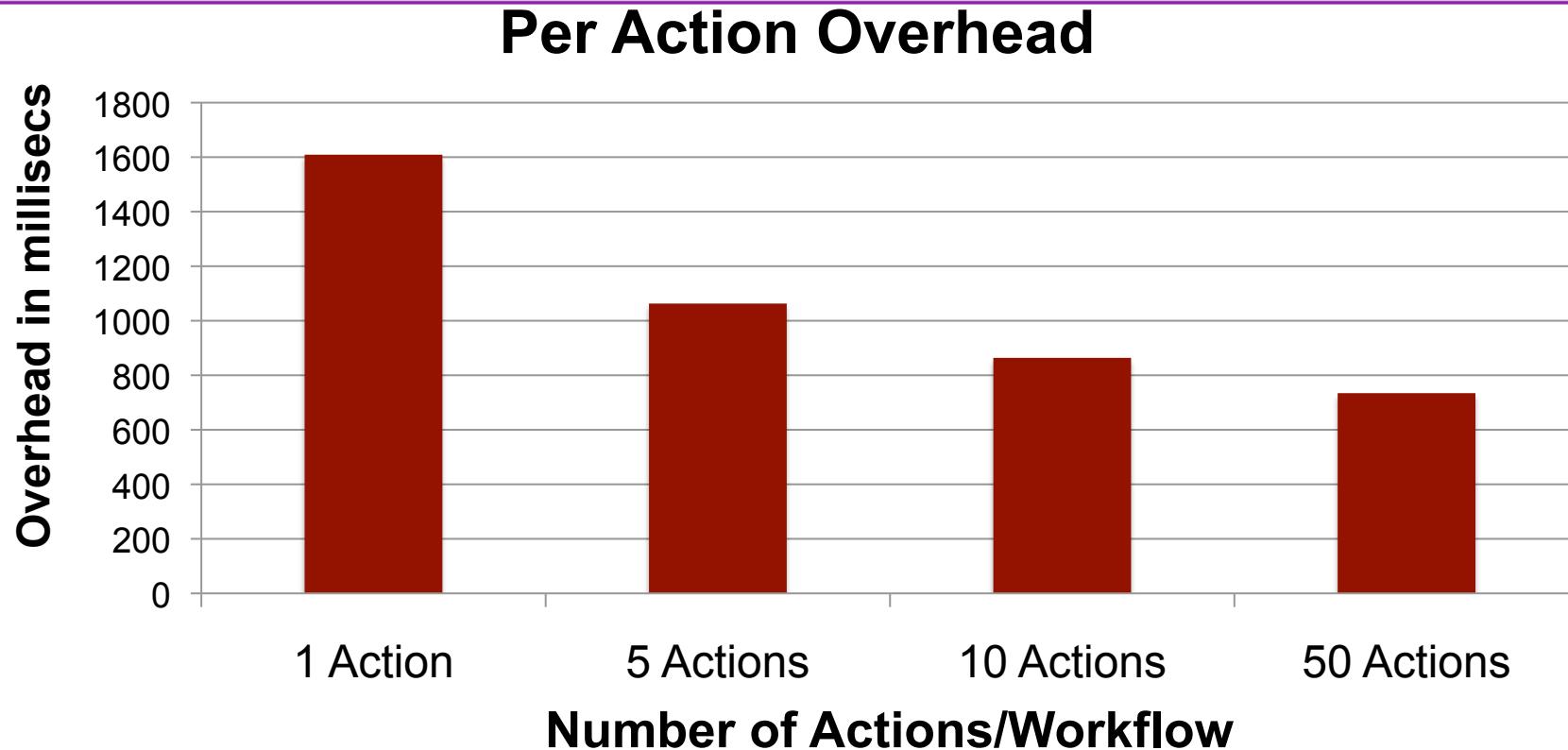
Observation: Oozie can accept a large number of jobs

Time Line of a Oozie Job



Total Oozie Overhead = Preparation + Completion

Oozie Overhead



Observation: Oozie overhead is less when multiple actions are in the same workflow.



Future Work

- Scalability
 - Hot-Hot/Load balancing service
 - Replace SQL DB with Zookeeper
- Event-based data processing
 - Asynchronous Data processing
 - User-defined event
- Extend the benchmarking scope
- Monitoring WS API



Take Away ..

- Oozie is
 - Apache project
 - Scalable, Reliable and multi-tenant
 - Deployed in production and growing fast.

Q & A



Mohammad K Islam
kamrul@yahoo-inc.com

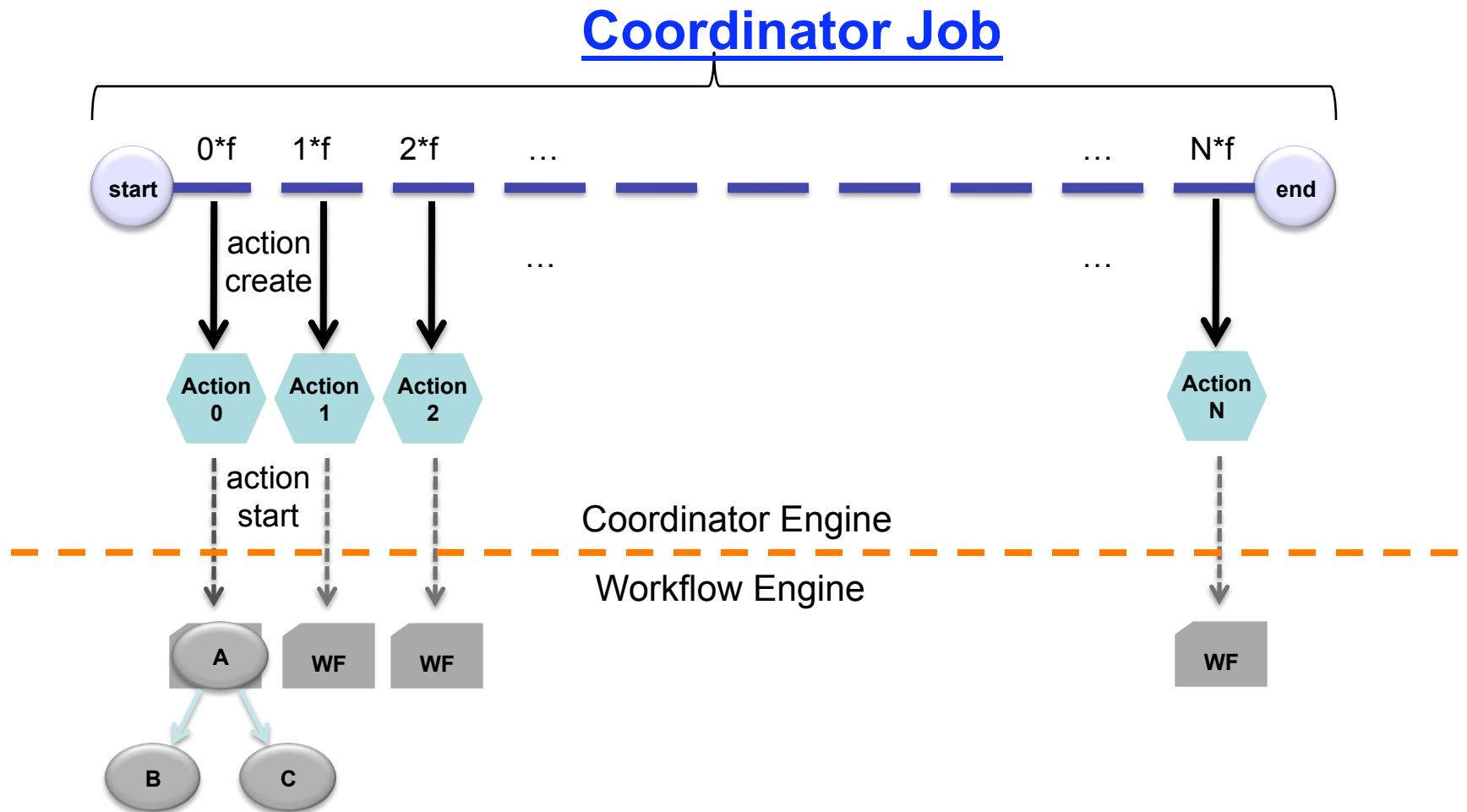
<http://incubator.apache.org/oozie/>

Backup Slides

YAHOO!



Coordinator Application Lifecycle



Vertical Scalability

- Oozie *asynchronously* processes the user request.
- **Memory resident**
 - An internal queue to store any sub-task.
 - Thread pool to process the sub-tasks.
- Both items
 - Fixed in size
 - Don't change due to load variations
 - Size can be configured if needed. Might need extra memory

Challenges in Scalability

- Centralized persistence storage
 - Currently supports any SQL DB (such as Oracle, MySQL, derby etc)
 - Each DBMS has respective limitations
 - Oozie scalability is limited by the underlying DBMS
 - Using of Zookeeper might be an option

Oozie Workflow Application

- Contents
 - A **workflow.xml** file
 - Resource files, config files and Pig scripts
 - All necessary JAR and native library files
- Parameters
 - The **workflow.xml**, is parameterized, parameters can be propagated to **map-reduce**, **pig** & **ssh** jobs
- Deployment
 - In a directory in the HDFS of the Hadoop cluster where the Hadoop & Pig jobs will run

Running a Workflow Job

Oozie
cmd

Workflow Application Deployment

```
$ hadoop fs -mkdir hdfs://usr/tucu/wordcount-wf
$ hadoop fs -mkdir hdfs://usr/tucu/wordcount-wf/lib
$ hadoop fs -copyFromLocal workflow.xml wordcount.xml hdfs://usr/tucu/wordcount-wf
$ hadoop fs -copyFromLocal hadoop-examples.jar hdfs://usr/tucu/wordcount-wf/lib
$
```

Workflow Job Execution

```
$ oozie run -o http://foo.corp:8080/oozie \
    -a hdfs://bar.corp:9000/usr/tucu/wordcount-wf \
    input=/data/2008/input output=/data/2008/output
Workflow job id [1234567890-wordcount-wf]
$
```

Workflow Job Status

```
$ oozie status -o http://foo.corp:8080/oozie -j 1234567890-wordcount-wf
Workflow job status [RUNNING]
...
$
```



Big Features (1/ 2)

- Integration with Hadoop 0.23
- Event-based data processing (3.3)
 - Asynchronous Data processing (3.3)
 - User-defined event (Next)
- HCatalog integration (3.3)
 - Non-polling approach

Big Features (2/ 2)

- DAG AM (Next)
 - WF will run as AM on hadoop cluster
 - Higher scalability
- Extend coordinator's scope (3.3)
 - Currently supports only WF
 - Any job type such as pig, hive, MR, java, distcp can be scheduled *without* WF.
- Dataflow-based processing in Hadoop(Next)
 - Create an abstraction to make things easier.

Usability (1/2)

- Easy adoption
 - Modeling tool (3.3)
 - IDE integration (3.3)
 - Modular Configurations (3.3)
- Monitoring API (3.3)
- Improved Oozie application management (next+)
 - Automate upload the application to hdfs.
 - Application versioning.
 - Seamless upgrade



Usability (2/2)

- Shell and Distcp Action
- Improved UI for coordinator
- Mini-Oozie for CI
 - Like Mini-cluster
- Support multiple versions (3.3)
 - Pig, Distcp, Hive etc.
- Allow job notification through JMS (Next++)
- Prioritization (Next++)
 - By user, system level.



Reliability

- Auto-Retry in WF Action level
- High-Availability
 - Hot-Warm through ZooKeeper (3.3)
 - Hot-Hot with Load Balancing (Next)



Manageability

- Email action
- Query Pig Stats/Hadoop Counters
 - Runtime control of Workflow based on stats
 - Application-level control using the stats

REST-API for Hadoop Components

- Direct access to Hadoop components
 - Emulates the command line through REST API.
- Supported Products:
 - Pig
 - Map Reduce