



APACHE OOZIE

DATA FLOW SCHEDULING LABS

Bin Jiang
04/08/2017

Oozie End2End Solution Lab

- **Apache Oozie Lab Instructions**

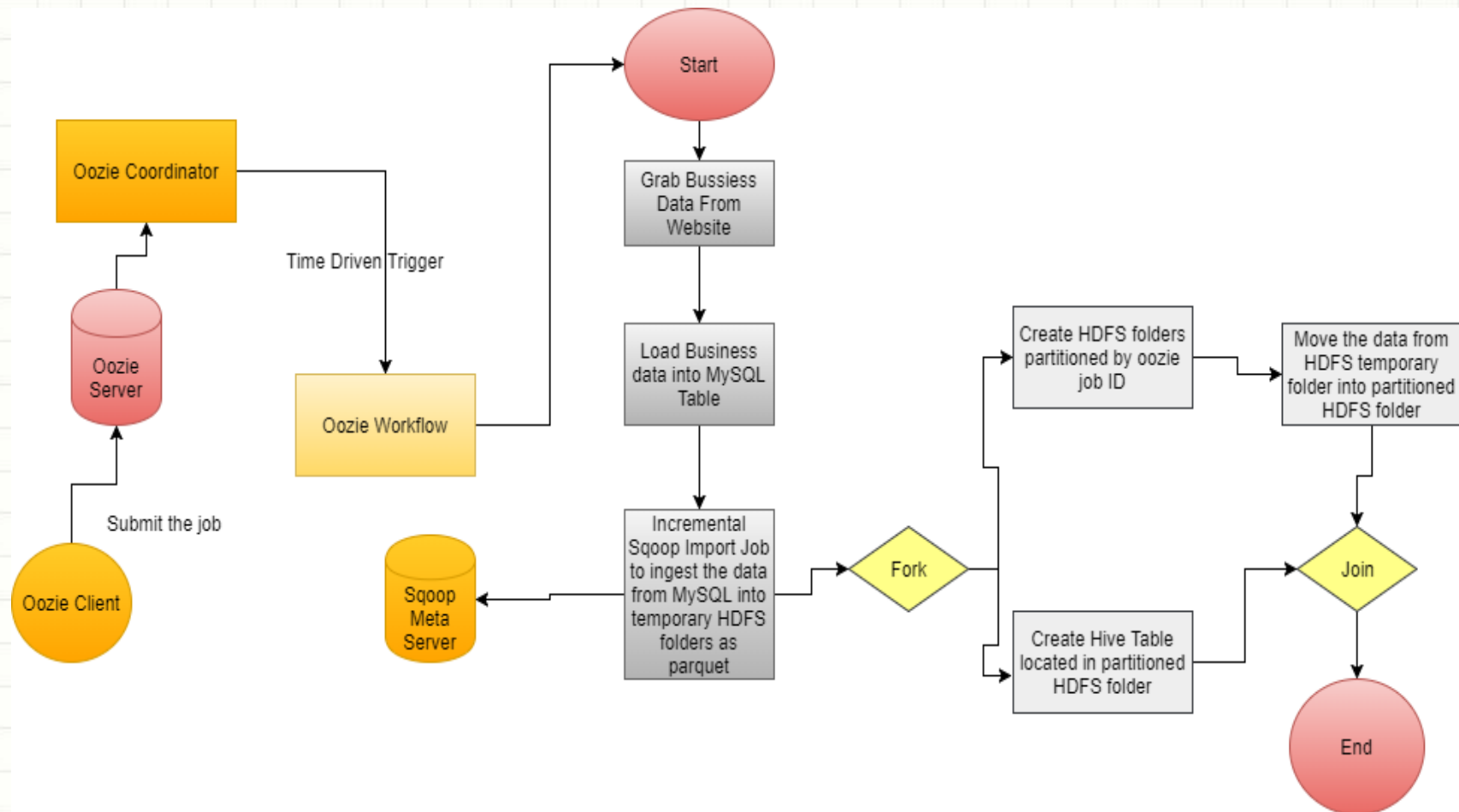
- TrainingOnHDP\docs\Class 5 - Data Workflow Scheduling and Management\lab\Apache Oozie Labs.txt

- **Apache Oozie Lab Code**

- TrainingOnHDP\DataFlowSchedulingOnOozie\end2end

Oozie End2End Solution Lab

- **Apache Oozie Lab Diagram**



Oozie End2End Solution Lab

- Hive Configuration Change via Ambari – Run Out Of Memory Issue

hive.tez.java.opts=-Xmx1024m

hive.server2.thrift.http.path	<input type="text" value="cliservice"/>			
hive.server2.thrift.http.port	<input type="text" value="10001"/>			
hive.server2.thrift.max.worker.threads	<input type="text" value="500"/>			
HiveServer2 Port	<input type="text" value="10000"/>			
hive.server2.thrift.sasl.qop	<input type="text" value="auth"/>			
hive.server2.transport.mode	<input type="text" value="binary"/>			
hive.tez.java.opts	<input type="text" value="-server -Xmx1024m -Djava.net.preferIPv4Stack=true"/>			
hive.tez.log.level	<input type="text" value="INFO"/>			
Transaction Manager	<input type="text" value="org.apache.hadoop.hive.ql.lockmgr.DbTxnManager"/>			
hive.txn.max.open.batch	<input type="text" value="1000"/>			
hive.txn.timeout	<input type="text" value="300"/>			

Oozie End2End Solution Lab

- Change the startTime and endTime in coordinator.properties if needed

```
nameNode=hdfs://sandbox-hdp.hortonworks.com:8020
jobTracker=sandbox-hdp.hortonworks.com:8050
queueName=default

oozie.libpath=${nameNode}/user/oozie/share/lib
oozie.use.system.libpath=true
oozie.wf.rerun.failnodes=true

oozieProjectRoot=${nameNode}/user/${user.name}/oozie
appRoot=${oozieProjectRoot}/coordinator
oozie.coord.application.path=${appRoot}/oozie_in_action_coord.xml

workflowAppPath=${oozieProjectRoot}/workflow/oozie_in_action.xml

inputDir=${oozieProjectRoot}/data/**/*/*/*/*/*
outputDirJavaMain=${appRoot}/output-JavaMain

startTime=2017-12-28T05:30Z
endTime=2017-12-29T20:57Z
timeZoneDef=UTC

inputDirRecordCount=`cat ${inputDir} | wc -l`
minRequiredRecordCount=1
```

Oozie End2End Solution Lab


- **Launch Oozie Job**

```
CoordAction ID: 0000002-171228045254275-oozie-oozi-C@1
[root@sandbox-hdp TrainingOnHDP]# oozie job -oozie http://127.0.0.1:11000/oozie -config /root/TrainingOnHDP/DataFlowSchedulingOnOozie/end2end/coordin
ator.properties -run
job: 0000010-171228060637208-oozie-oozi-C
[root@sandbox-hdp TrainingOnHDP]#
```

Oozie End2End Solution Lab


- **Oozie Coordinator Job is running**

← → ↻ ⓘ 127.0.0.1:11000/oozie/

 [Documentation](#)

Oozie Web Console

[Workflow Jobs](#) **[Coordinator Jobs](#)** [Bundle Jobs](#) [System Info](#) [Instrumentation](#) [Settings](#)

 [All Jobs](#) **[Active Jobs](#)** [Done Jobs](#) [Custom Filter](#) ▼ S

	Job Id	Name	Status	User	Group	Frequency	Unit	Started	Next Materialization
1	0000010-171228060637208-oozie-oozi-C	Oozie In Action	RUNNING	root		30	MINUTE	Thu, 28 Dec 2017 05:30:00 G...	Thu, 28 Dec 2017 06:30:00 G...
2	0000007-171228060637208-oozie-oozi-C	Oozie In Action	KILLED	root		30	MINUTE	Thu, 28 Dec 2017 05:30:00 G...	Thu, 28 Dec 2017 07:30:00 G...
3	0000003-171228060637208-oozie-oozi-C	Oozie In Action	KILLED	root		30	MINUTE	Thu, 28 Dec 2017 05:30:00 G...	Thu, 28 Dec 2017 07:30:00 G...
4	0000000-171228060637208-oozie-oozi-C	Oozie In Action	KILLED	root		30	MINUTE	Thu, 28 Dec 2017 05:30:00 G...	Thu, 28 Dec 2017 06:30:00 G...
5	0000002-171228045254275-oozie-oozi-C	Oozie In Action	KILLED	root		30	MINUTE	Thu, 28 Dec 2017 05:30:00 G...	Thu, 28 Dec 2017 06:30:00 G...
6	0000001-171228045254275-oozie-oozi-C	Oozie In Action	KILLED	root		30	MINUTE	Thu, 28 Dec 2017 05:30:00 G...	

Oozie End2End Solution Lab

- **Oozie Workflow Job is running**

 [Documentation](#)

Oozie Web Console

Workflow Jobs

Coordinator Jobs

Bundle Jobs

System Info

Instrumentation

Settings



All Jobs

Active Jobs

Done Jobs

Custom Filter ▼

Server version [4.2.0.2]

	Job Id	Name	Status	R...	User	Group	Created	Started	Last Modified	Ended
1	0000011-171228060637208-oozie-oozi-W	Oozie In Action	RUNNING	0	root		Thu, 28 Dec 2017 06:57:48 G...	Thu, 28 Dec 2017 06:57:48 G...	Thu, 28 Dec 2017 06:57:51 G...	
2	0000009-171228060637208-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 06:52:15 G...	Thu, 28 Dec 2017 06:52:16 G...	Thu, 28 Dec 2017 06:54:00 G...	Thu, 28 Dec 2017 06:53:54 G...
3	0000008-171228060637208-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 06:49:16 G...	Thu, 28 Dec 2017 06:49:16 G...	Thu, 28 Dec 2017 06:52:15 G...	Thu, 28 Dec 2017 06:52:15 G...
4	0000006-171228060637208-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 06:37:08 G...	Thu, 28 Dec 2017 06:37:08 G...	Thu, 28 Dec 2017 06:37:30 G...	Thu, 28 Dec 2017 06:37:27 G...
5	0000005-171228060637208-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 06:34:01 G...	Thu, 28 Dec 2017 06:34:01 G...	Thu, 28 Dec 2017 06:37:07 G...	Thu, 28 Dec 2017 06:37:07 G...
6	0000004-171228060637208-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 06:31:01 G...	Thu, 28 Dec 2017 06:31:01 G...	Thu, 28 Dec 2017 06:34:00 G...	Thu, 28 Dec 2017 06:34:00 G...
7	0000002-171228060637208-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 06:15:18 G...	Thu, 28 Dec 2017 06:15:18 G...	Thu, 28 Dec 2017 06:18:39 G...	Thu, 28 Dec 2017 06:18:39 G...
8	0000001-171228060637208-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 06:09:58 G...	Thu, 28 Dec 2017 06:09:58 G...	Thu, 28 Dec 2017 06:15:17 G...	Thu, 28 Dec 2017 06:15:16 G...
9	0000003-171228045254275-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 05:41:19 G...	Thu, 28 Dec 2017 05:41:19 G...	Thu, 28 Dec 2017 06:07:37 G...	Thu, 28 Dec 2017 06:07:36 G...
10	0000000-171228045254275-oozie-oozi-W	Oozie In Action	KILLED	0	root		Thu, 28 Dec 2017 04:56:03 G...	Thu, 28 Dec 2017 04:56:03 G...	Thu, 28 Dec 2017 05:37:08 G...	Thu, 28 Dec 2017 05:37:07 G...

Oozie End2End Solution Lab

- **Oozie Workflow Job is completed**

Job Id: 0000011-171228060637208-oozie-oozi-W
Name: Oozie In Action
App Path: hdfs://sandbox-hdp.hortonworks.com:8020/user/root/oozie/workflow/oozi
Run: 0
Status: SUCCEEDED
User: root
Group:
Parent Coord: 0000010-171228060637208-oozie-oozi-C@1
Create Time: Thu, 28 Dec 2017 06:57:48 GMT
Start Time: Thu, 28 Dec 2017 06:57:48 GMT
Last Modified: Thu, 28 Dec 2017 07:10:29 GMT
End Time: Thu, 28 Dec 2017 07:10:29 GMT

Actions

	Action Id	Name	Type	Status	Transition	StartTime	EndTime
1	0000011-171228060637208-oozie-oozi-W@:start:	:start:	:START:	OK	WGetFiles	Thu, 28 Dec 2017 06:57:48 G...	Thu, 28 Dec 2017 06:57:48 G...
2	0000011-171228060637208-oozie-oozi-W@WGetFiles	WGetFiles	shell	OK	PrepareMy...	Thu, 28 Dec 2017 06:57:49 G...	Thu, 28 Dec 2017 06:59:20 G...
3	0000011-171228060637208-oozie-oozi-W@PrepareMyS...	PrepareMy...	shell	OK	SqoopImpo...	Thu, 28 Dec 2017 06:59:20 G...	Thu, 28 Dec 2017 07:00:53 G...
4	0000011-171228060637208-oozie-oozi-W@SqoopImpor...	SqoopImpo...	sqoop	OK	Fork	Thu, 28 Dec 2017 07:00:54 G...	Thu, 28 Dec 2017 07:06:00 G...
5	0000011-171228060637208-oozie-oozi-W@Fork	Fork	:FORK:	OK	*	Thu, 28 Dec 2017 07:06:01 G...	Thu, 28 Dec 2017 07:06:01 G...
6	0000011-171228060637208-oozie-oozi-W@LoadIntoHive	LoadIntoHive	hive	OK	Join	Thu, 28 Dec 2017 07:06:02 G...	Thu, 28 Dec 2017 07:10:28 G...
7	0000011-171228060637208-oozie-oozi-W@CreatePartiti...	CreatePartit...	fs	OK	MoveFiles	Thu, 28 Dec 2017 07:06:02 G...	Thu, 28 Dec 2017 07:06:17 G...
8	0000011-171228060637208-oozie-oozi-W@MoveFiles	MoveFiles	fs	OK	Join	Thu, 28 Dec 2017 07:06:18 G...	Thu, 28 Dec 2017 07:06:18 G...
9	0000011-171228060637208-oozie-oozi-W@Join	Join	:JOIN:	OK	end	Thu, 28 Dec 2017 07:10:28 G...	Thu, 28 Dec 2017 07:10:29 G...
10	0000011-171228060637208-oozie-oozi-W@end	end	:END:	OK		Thu, 28 Dec 2017 07:10:29 G...	Thu, 28 Dec 2017 07:10:29 G...

Oozie End2End Solution Lab

- Sqoop job has imported the data from mysql into partitioned HDFS folders

```
[root@sandbox-hdp end2end]# hadoop fs -ls /user/yarn/business
Found 2 items
drwxr-xr-x  - root hdfs      0 2017-12-28 07:06 /user/yarn/business/jobid=0000011-171228060637208-oozie-oozi-W
drwxr-xr-x  - root hdfs      0 2017-12-28 07:19 /user/yarn/business/jobid=0000012-171228060637208-oozie-oozi-W
[root@sandbox-hdp end2end]#
```

Oozie End2End Solution Lab

- **MySQL Business Table which is imported into Hive**

```
mysql> show tables;
+-----+
| Tables_in_oozie |
+-----+
| business        |
+-----+
1 row in set (0.00 sec)

mysql> select count(*) from business;
+-----+
| count(*) |
+-----+
|    230471 |
+-----+
1 row in set (0.90 sec)

mysql> select count(*) from business;
+-----+
| count(*) |
+-----+
|    460942 |
+-----+
1 row in set (4.15 sec)
```

Oozie End2End Solution Lab

- The data is imported into Hive Table

```
>
> select count(*) from business;
Query ID = root_20171228074144_9598ffcc-c122-4326-bd29-0256bd360236
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1514224417003_0026)

-----
      VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 .....  SUCCEEDED      4         4         0         0         0         0
Reducer 2 .....  SUCCEEDED      1         1         0         0         0         0
-----
VERTICES: 02/02  [=====>>] 100%  ELAPSED TIME: 113.90 s
-----
OK
460942
Time taken: 118.094 seconds, Fetched: 1 row(s)
hive>
```


Oozie End2End Solution Lab

- The data is imported into Hive Table

```
Time taken: 118.094 seconds, Fetched: 1 row(s)
hive> select * from business limit 5;
OK
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
0007594-01-001 7594 Devincenzi Trucking Inc Devincenzi Trucking Inc 1598 Carroll Av San Francisco CA 94124 07/01/1987 07/01
1987 1598 Carroll Av San Francisco 94124 CA 4800-4999 Transportation and Warehousing false false 10 B
Oceanview Hunters Point 1598 CARROLL AV
San Francisco, CA 94124
(37.725018, -122.393442) 1 0000011-171228060637208-oozie-oozi-W
0313066-03-999 313066 Johnson Sarah H Sarah Johnson 27 Meadow Wy Fairfax CA 94930 01/18/1998 02/01/2009 06/01/2015 P
Box 61 Fairfax 94978 CA false false 27 MEADOW WY
FAIRFAX, CA 94930
(37.975269, -122.60045) 2 0000011-171228060637208-oozie-oozi-W
0360413-02-001 360413 Chen Jian Q J & I Cycle Trading Co 455 Faxon Ave San Francisco CA 94112 10/21/2002 08/07/2007 2
2 Howth St San Francisco 94112 CA false false 7 Oceanview/Merced/Ingleside 455 F
FAXON AVE
San Francisco, CA 94112
(37.722844, -122.460239) 3 0000011-171228060637208-oozie-oozi-W
0360414-01-001 360414 Aew Engineering Inc Aew Engineering Inc 55 New Montgomery St 722 San Francisco CA 94105 08/19/2002 1
0/01/2008 55 New Montgomery St #722 San Francisco 94105 CA false false 6 Finan
cial District/South Beach 55 NEW MONTGOMERY ST
San Francisco, CA 94105
(37.787833, -122.400946) 4 0000011-171228060637208-oozie-oozi-W
0360417-03-001 360417 Trager Richard S Trager Photography 68 Mccoppin St San Francisco CA 94103 07/25/2013 07/25
2013 68 Mccoppin St San Francisco 94103 CA false false 6 Mission 68 MCCOPPIN S
San Francisco, CA 94103
(37.771634, -122.421839) 5 0000011-171228060637208-oozie-oozi-W
Time taken: 1.106 seconds, Fetched: 5 row(s)
```

Oozie End2End Solution Lab


- **Import Oozie in Action into Zeppelin**


 **Zeppelin** Notebook ▾ Job

Welcome to Zeppelin!

Zeppelin is web-based notebook that enables interactive data analytics.
You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!

Notebook ↻

 Import note

 Create new note

☐ Getting Started

☐ Labs

☒ Mobile Phone Vendors Twitter Dashboard

☒ **Oozie In Action**

☐ R (SparkR)

☐ Real time Data Visualization via NiFi and Web Socket


☐ Zeppelin Tutorial (Basic Features)


Help


Get started with [Zeppelin documentation](#)

Community

Please feel free to help us to improve Zeppelin,
Any contribution are welcome!


 Mailing list


 Issues tracking

 Github

Oozie End2End Solution Lab

- **Oozie in Action Note in Zeppelin**

 **Zeppelin** Notebook ▾ Job Search your No

Oozie In Action 

Oozie In Action

Goal

Demonstrate how Enterprise Oozie applications works (sample data, scripts, configurations, libraries, HDFS and Oozie commands) covered in the training course. In the class, We will create one miniproject to help the students to understand:

- ETL data from RDMS - ingested using [Apache Sqoop](#)
- Schedul the jobs to ingest the DATA - Workflow Scheduler is [Apache Oozie](#)
- Create External Hive Table and Dynamical Partitioned Data - [Apache Hive](#)

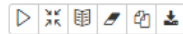
Actions/Functionalities covered:

1. Oozie workflow - hdfs and email actions
2. Oozie workflow - hdfs, email and hive actions
3. Oozie workflow - sqoop action (Hive-mysql)
4. Oozie workflow - Pig action
5. Oozie sub-workflow
6. Oozie coordinator job - time-triggered sub-workflow, fork-join control and decision control
7. Oozie coordinator jobs - file triggered
8. Oozie coordinator jobs - dataset availability triggered
9. Oozie bundle jobs
10. Oozie workflow - shell action +passing output from one action to another
11. Oozie workflow - SSH action

Oozie End2End Solution Lab

- Oozie in Action Note in Zeppelin

Oozie In Action



Head



default

City Geography Statistics

FINISHED

```
%jdbc(hive)
select city, count(*) num_per_city from oozie.business group by city
```



settings



Output is truncated to 1000 rows. Learn more about [common.max_count](#)



Oozie End2End Solution Lab

- **Oozie in Action Note in Zeppelin**

Oozie In Action



Head



default

State Geography Statistics

FINISHED

```
%jdbc(hive)
select state, count(*) num_per_tate from oozie.business group by state
```



settings

