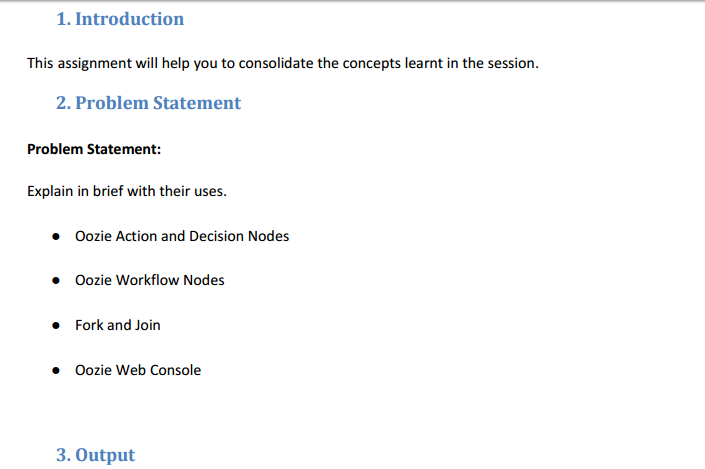
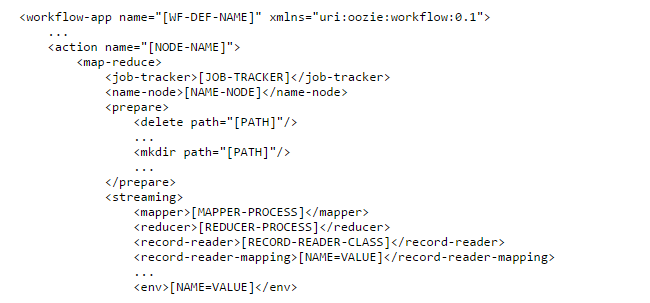
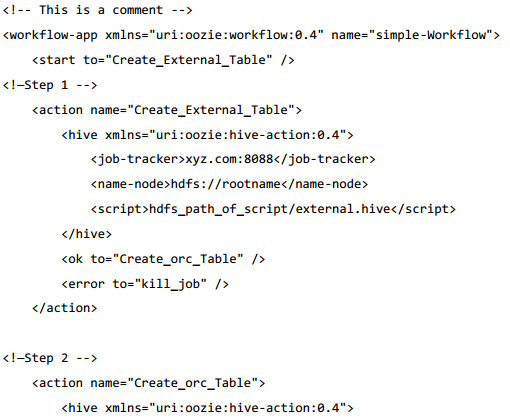
**Assignment 34.3**

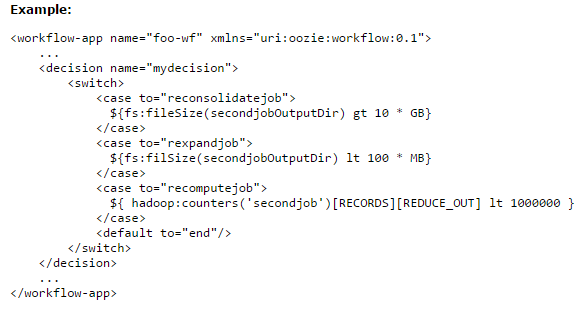


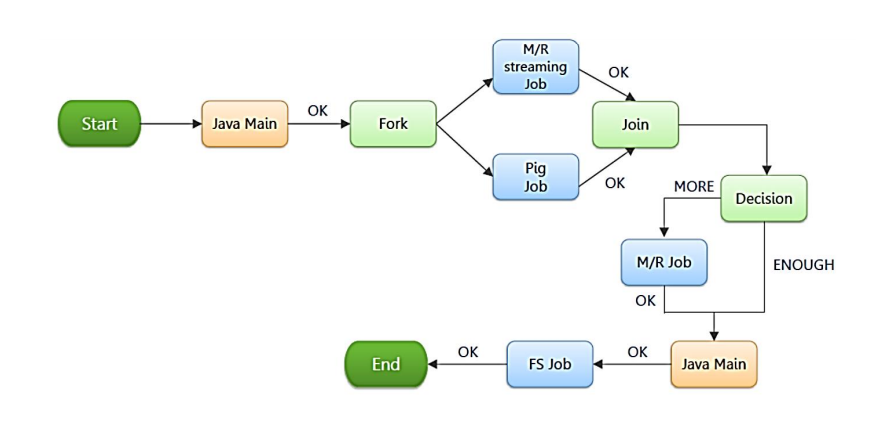
* **Oozie action and decision nodes**
* **Action Node**
* An action node represents a workflow task, e.g., moving files into HDFS, running a MapReduce, Pig or[Hive](http://www.guru99.com/hive-tutorials.html) jobs, importing data using Sqoop or running a shell script of a program written in Java.
* **Example of map reduce action**



* **Example of Hive Action node**



* **Decision Node**
* We can add decision tags to check if we want to run an action based on the output of decision. In the above example, if we already have the hive table we won’t need to create it again. In such a scenario, we can add a decision tag to not run the Create Table steps if the table already exists.
* The behavior of a decision node can be seen as a switch-case statement.
* A decision node consists of a list of predicates-transition pairs plus a default transition. Predicates are evaluated in order or appearance until one of them evaluates to true and the corresponding transition is taken. If none of the predicates evaluates to true the default transition is taken.
* **=**
* **Oozie workflow nodes**



Oozie workflow nodes contains two nodes –

* **Control Flow Nodes**

Control nodes, which are used to define job chronology, provide the rules for beginning and ending a workflow and control the workflow execution path with possible decision points known as fork and join nodes.

**Control node achieves this by using 3 nodes**

**• Start/end/kill**

**• Decision**

**• Fork/join**

* **Start Control Node**

The start node is the entry point for a workflow job, it indicates the first workflow node the workflow job must transition to.When a workflow is started, it automatically transitions to the node specified in the start .

A workflow definition must have one start node.

Syntax:

<workflow-app name="[WF-DEF-NAME]" xmlns="uri:oozie:workflow:0.1">

...

<start to="[NODE-NAME]"/>

...

</workflow-app>

The to attribute is the name of first workflow node to execute.

Example:

<workflow-app name="foo-wf" xmlns="uri:oozie:workflow:0.1">

...

<start to="firstHadoopJob"/>

...

</workflow-app>

* **Kill control Node**

The kill node allows a workflow job to kill itself.

When a workflow job reaches the kill it finishes in error (KILLED).

If one or more actions started by the workflow job are executing when the kill node is reached, the actions will be killed.

A workflow definition may have zero or more kill nodes.

Syntax:

<workflow-app name="[WF-DEF-NAME]" xmlns="uri:oozie:workflow:0.1">

...

<kill name="[NODE-NAME]">

<message>[MESSAGE-TO-LOG]</message>

</kill>

...

</workflow-app>

The name attribute in the kill node is the name of the Kill action node.

The content of the message element will be logged as the kill reason for the workflow job.

A kill node does not have transition elements because it ends the workflow job, as KILLED .

**Example:**

<workflow-app name="foo-wf" xmlns="uri:oozie:workflow:0.1">

...

<kill name="killBecauseNoInput">

<message>Input unavailable</message>

</kill>

...

</workflow-app>

* **End Control Node**

The end node is the end for a workflow job, it indicates that the workflow job has completed successfully.

When a workflow job reaches the end it finishes successfully (SUCCEEDED).

If one or more actions started by the workflow job are executing when the end node is reached, the actions will be killed. In this scenario the workflow job is still considered as successfully run.

A workflow definition must have one end node.

Syntax:

<workflow-app name="[WF-DEF-NAME]" xmlns="uri:oozie:workflow:0.1">

...

<end name="[NODE-NAME]"/>

...

</workflow-app>

The name attribute is the name of the transition to do to end the workflow job.

Example:

<workflow-app name="foo-wf" xmlns="uri:oozie:workflow:0.1">

...

<end name="end"/>

</workflow-app>

* **Decision Control Node**

A decision node enables a workflow to make a selection on the execution path to follow.

The behavior of a decision node can be seen as a switch-case statement.

A decision node consists of a list of predicates-transition pairs plus a default transition. Predicates are evaluated in order or appearance until one of them evaluates to true and the corresponding transition is taken. If none of the predicates evaluates to true the default transition is taken.

Predicates are JSP Expression Language (EL) expressions (refer to section 4.2 of this document) that resolve into a boolean value, true or false . For example:

${fs:fileSize('/usr/foo/myinputdir') gt 10 \* GB}

Syntax:

<workflow-app name="[WF-DEF-NAME]" xmlns="uri:oozie:workflow:0.1">

...

<decision name="[NODE-NAME]">

<switch>

<case to="[NODE\_NAME]">[PREDICATE]</case>

...

<case to="[NODE\_NAME]">[PREDICATE]</case>

<default to="[NODE\_NAME]"/>

</switch>

</decision>

...

</workflow-app>

The name attribute in the decision node is the name of the decision node.

Each case elements contains a predicate an a transition name. The predicate ELs are evaluated in order until one returns true and the corresponding transition is taken.

The default element indicates the transition to take if none of the predicates evaluates to true .

All decision nodes must have a default element to avoid bringing the workflow into an error state if none of the predicates evaluates to true.

**Example:**

<workflow-app name="foo-wf" xmlns="uri:oozie:workflow:0.1">

...

<decision name="mydecision">

<switch>

<case to="reconsolidatejob">

${fs:fileSize(secondjobOutputDir) gt 10 \* GB}

</case>

<case to="rexpandjob">

${fs:filSize(secondjobOutputDir) lt 100 \* MB}

</case>

<case to="recomputejob">

${ hadoop:counters('secondjob')[RECORDS][REDUCE\_OUT] lt 1000000 }

</case>

<default to="end"/>

</switch>

</decision>

...

</workflow-app>

* **Fork and Join Control Nodes**

A fork node splits one path of execution into multiple concurrent paths of execution.

A join node waits until every concurrent execution path of a previous fork node arrives to it.

The fork and join nodes must be used in pairs. The join node assumes concurrent execution paths are children of the same fork node.

Syntax:

<workflow-app name="[WF-DEF-NAME]" xmlns="uri:oozie:workflow:0.1">

...

<fork name="[FORK-NODE-NAME]">

<path start="[NODE-NAME]" />

...

<path start="[NODE-NAME]" />

</fork>

...

<join name="[JOIN-NODE-NAME]" to="[NODE-NAME]" />

...

</workflow-app>

The name attribute in the fork node is the name of the workflow fork node. The start attribute in the path elements in the fork node indicate the name of the workflow node that will be part of the concurrent execution paths.

The name attribute in the join node is the name of the workflow join node. The to attribute in the join node indicates the name of the workflow node that will executed after all concurrent execution paths of the corresponding fork arrive to the join node.

Example:

<workflow-app name="sample-wf" xmlns="uri:oozie:workflow:0.1">

...

<fork name="forking">

<path start="firstparalleljob"/>

<path start="secondparalleljob"/>

</fork>

<action name="firstparallejob">

<map-reduce>

<job-tracker>foo:9001</job-tracker>

<name-node>bar:9000</name-node>

<job-xml>job1.xml</job-xml>

</map-reduce>

<ok to="joining"/>

<error to="kill"/>

</action>

<action name="secondparalleljob">

<map-reduce>

<job-tracker>foo:9001</job-tracker>

<name-node>bar:9000</name-node>

<job-xml>job2.xml</job-xml>

</map-reduce>

<ok to="joining"/>

<error to="kill"/>

</action>

<join name="joining" to="nextaction"/>

...

</workflow-app>

* **Action Nodes**
* An action node represents a workflow task, e.g., moving files into HDFS, running a MapReduce, Pig or[Hive](http://www.guru99.com/hive-tutorials.html) jobs, importing data using Sqoop or running a shell script of a program written in Java.

• Map-reduce

• Pig

• Hdfs

• Sub-workflow

• Java-run custom java code

* To run oozie workflows, two files are needed.

**1. workflow.xml (stored in HDFS**)

• It contains the structure of workflow.

**2. job.properties (stored in local)**

• It contains the configuration properties.

* **Fork and join Control Nodes**

A fork node splits one path of execution into multiple concurrent paths of execution.

A join node waits until every concurrent execution path of a previous fork node arrives to it.

The fork and join nodes must be used in pairs. The join node assumes concurrent execution paths are children of the same fork node.

Syntax:

<workflow-app name="[WF-DEF-NAME]" xmlns="uri:oozie:workflow:0.1">

...

<fork name="[FORK-NODE-NAME]">

<path start="[NODE-NAME]" />

...

<path start="[NODE-NAME]" />

</fork>

...

<join name="[JOIN-NODE-NAME]" to="[NODE-NAME]" />

...

</workflow-app>

The name attribute in the fork node is the name of the workflow fork node. The start attribute in the path elements in the fork node indicate the name of the workflow node that will be part of the concurrent execution paths.

The name attribute in the join node is the name of the workflow join node. The to attribute in the join node indicates the name of the workflow node that will executed after all concurrent execution paths of the corresponding fork arrive to the join node.

**Example:**

<workflow-app name="sample-wf" xmlns="uri:oozie:workflow:0.1">

...

<fork name="forking">

<path start="firstparalleljob"/>

<path start="secondparalleljob"/>

</fork>

<action name="firstparallejob">

<map-reduce>

<job-tracker>foo:9001</job-tracker>

<name-node>bar:9000</name-node>

<job-xml>job1.xml</job-xml>

</map-reduce>

<ok to="joining"/>

<error to="kill"/>

</action>

<action name="secondparalleljob">

<map-reduce>

<job-tracker>foo:9001</job-tracker>

<name-node>bar:9000</name-node>

<job-xml>job2.xml</job-xml>

</map-reduce>

<ok to="joining"/>

<error to="kill"/>

</action>

<join name="joining" to="nextaction"/>

...

</workflow-app>

* **Oozie web console**

Oozie provides a read-only Web based console that allows to allow to monitor Oozie system status, workflow applications status and workflow jobs status.

The Web base console is implemented as a client of the Web Services API.

We can enable Oozie's web console as a means of working with Cloudera Manager. **To enable the Oozie Web Console**

1. Download [**ext-2.2**](http://extjs.com/deploy/ext-2.2.zip). Extract the contents of the file to /var/lib/oozie/libext.
2. Connect to the Cloudera Manager Admin Console. Click the **Configuration** tab for the Oozie service you are configuring and enable **Enable Oozie server web console**.
3. Ensure sure the Oozie user has read/write permissions on the directory /var/lib/oozie/oozie-server/webapps. For example, you might review the permissions as follows:
4. $ ls -l /var/lib/oozie/oozie-server/
5. total 12
6. drwxr-xr-x 3 oozie oozie 4096 Jul 13 05:42 conf
7. dr--r--r-- 4 oozie oozie 4096 Jul 13 13:15 webapps

drwxr-xr-x 3 oozie oozie 4096 Jul 13 05:42 work

1. If the permissions are set appropriately, continue to the next step. If permissions must be modified, you could do so as follows:
2. $ chmod 755 /var/lib/oozie/oozie-server/webapps
3. $ ls -l /var/lib/oozie/oozie-server/
4. drwxr-xr-x 3 oozie oozie 4096 Jul 13 05:42 conf
5. drwxr-xr-x 4 oozie oozie 4096 Jul 13 13:15 webapps

drwxr-xr-x 3 oozie oozie 4096 Jul 13 05:42 work

The Oozie web console is enabled