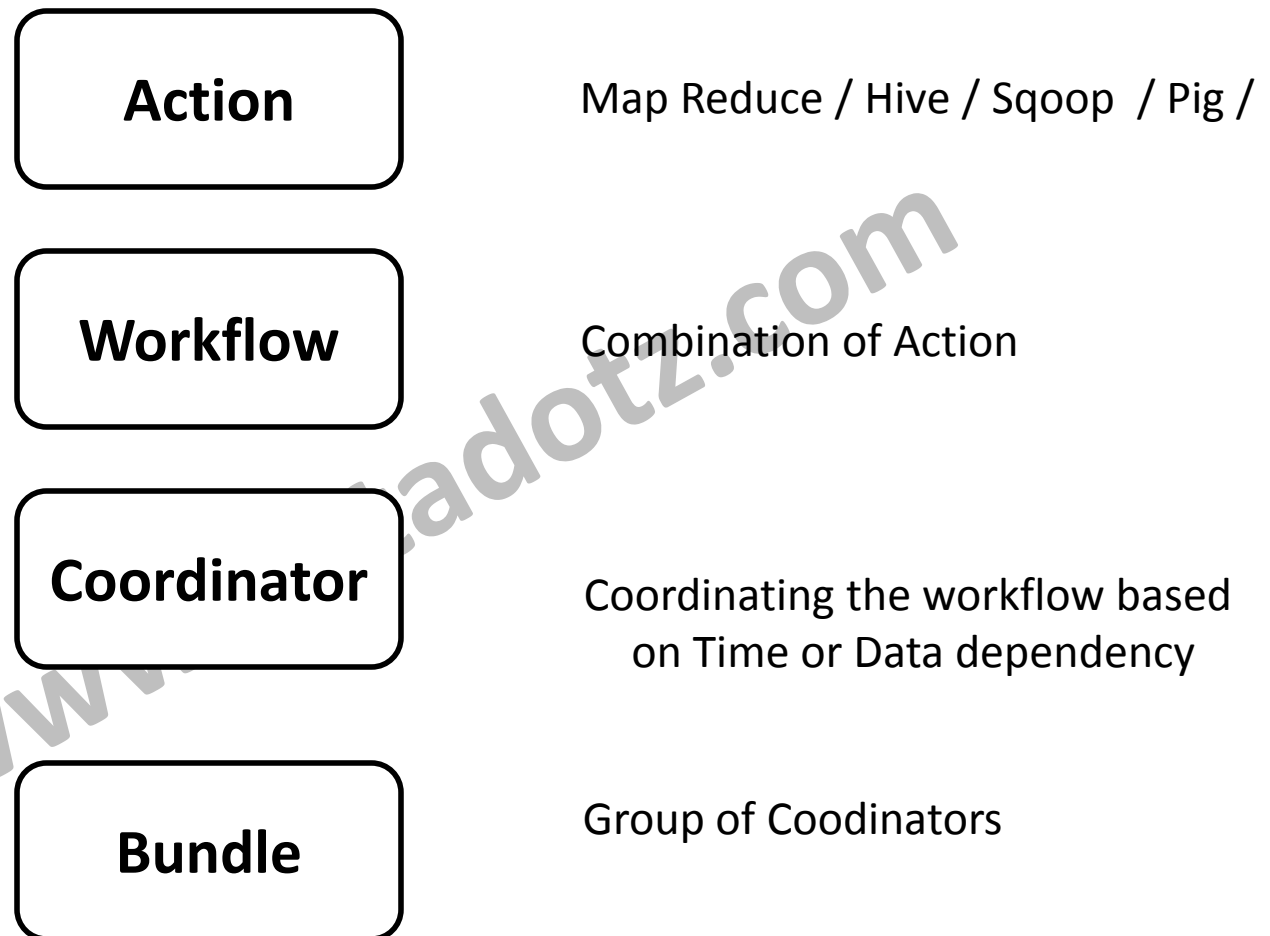# OOZIE

Senthil Kumar A

What is Oozie?

- Workflow scheduler for Hadoop jobs

- Scalable, Reliable and Extensible system designed for complex search engine workflows

- Directed Acyclic Graphs (DAG) of actions
  - A "control dependency" from one action to another
  - The second action can't run until the first action has completed.

- Developed by Yahoo and Licensed by Apache
- Latest Release
  - Oozie 4.0.1

Cont.

- Integrated with the rest of the Hadoop jobs
    - Map Reduce, Pig, Hive, Sqoop and more
    - System specific jobs (such as Java programs and shell scripts)

- Oozie Coordinator
    - Specialized in running workflows triggered by time (frequency) and data availability.

- Oozie Bundle
    - Provides higher level oozie abstraction.
    - The user will be able to start/stop/suspend/resume/rerun a set coordinator jobs in the bundle level

Layers of Oozie/ Architecture

Action — Map Reduce / Hive / Sqoop / Pig /

Workflow — Combination of Action

Coordinator — Coordinating the workflow based on Time or Data dependency

Bundle — Group of Coodinators

Installation

- tar -zxvf oozie-3.3.0.tar.gz

- chmod -R 777 oozie-3.3.0

- Add the properties to hadoop-1.2.1/conf/core-site.xml

```
<property>
        <name>hadoop.proxyuser.username.hosts</name>
        <value>*</value>
</property>
<property>
        <name>hadoop.proxyuser.username.groups</name>
         <value>*</value>
</property>
```

Cont.

- Start your hadoop cluster

- Start oozie
  - $cd oozie-3.3.0
  - $ bin/oozie-start.sh

- Web url
  - localhost:11000/oozie

How do you make a Oozie workflow?

- Workflow.xml file with all job configuration properties
  - Input, Output files and Script Files
  - Actions
  - Hadoop configuration for the action
  - Namenode, Job Tracker

- Job.properties file.
  - This file defines the values for variables assigned in workflow.xml
  - Location of workflow.xml
  - It also gives the location of the shared jars and other dependencies

DATA DOTZ

Map-Reduce WorkFlow

- Map/Reduce tags (HDFS)
  - To run a map/reduce process.

```
<start to="[mr-node]"/>
<action name="mr-node">
   <map-reduce>
   <job-tracker>${jobTracker}</job-tracker>
   <name-node>${nameNode}</name-node>
          <configuration>
   [YOUR HADOOP CONFIGURATION]
          </configuration>
   </map-reduce>
   <ok to="end"/>
   <error to="fail"/>
</action>
    <error name="[fail]"/>
    <message>"[A error message from last error node]"</message> </error>
<end name="[end]"/>
```

DATA DOTZ

Job properties

- Job.properties (Local)

nameNode=hdfs://localhost:50000

jobTracker=localhost:50001

queueName=default

examplesRoot=examples

oozie.wf.application.path=${nameNode}/[HDFS Path]

Hive Action

- Hive tags
  - Runs a Hive job.
  - Supports Hive scripts with parameter variables, their syntax is ${VARIABLES}.

- Job.properties

nameNode=hdfs://localhost:50000

jobTracker=localhost:50001

queueName=default

oozie.use.system.libpath=true

oozie.libpath=${nameNode}/oozie/sharelib/hive

oozie.wf.application.path=${nameNode}/HIVEoozie