

# Ecosystem Tools

CISC-525

Phil Grim

# Overview

- ▶ Sqoop
- ▶ Oozie
- ▶ Hue



# Sqoop

- ▶ Tool for efficiently transferring data between Hadoop and traditional data stores such as RDBMSs.
- ▶ Generates MapReduce job to accomplish transfers
- ▶ Can both import and export data
  - ▶ Sequence Files
  - ▶ Hive
  - ▶ HBase
  - ▶ Accumulo
  - ▶ Avro
- ▶ Includes tools to inspect data stores



# Sqoop

- ▶ Natively supports many database systems with JDBC drivers
  - ▶ Oracle
  - ▶ MySQL
  - ▶ PostgreSQL
  - ▶ Microsoft SQL Server
- ▶ Provides API for supporting other data sources and file types
- ▶ Supports full table import/export, incremental updates
- ▶ Generates Java code that can be re-used in MapReduce jobs.

# Sqoop

```
$ sqoop help
```

```
Running Sqoop version: 1.4.5-mapr-1410
```

```
usage: sqoop COMMAND [ARGS]
```

```
Available commands:
```

codegen	Generate code to interact with database records
create-hive-table	Import a table definition into Hive
eval	Evaluate a SQL statement and display the
results	
export	Export an HDFS directory to a database table
help	List available commands
import	Import a table from a database to HDFS
import-all-tables	Import tables from a database to HDFS
job	Work with saved jobs
list-databases	List available databases on a server
list-tables	List available tables in a database
merge	Merge results of incremental imports
metastore	Run a standalone Sqoop metastore
version	Display version information

```
See 'sqoop help COMMAND' for information on a specific command.
```

# Oozie

- ▶ Workflow scheduler system to control Hadoop jobs
- ▶ Workflows implemented as Directed Acyclical Graphs of actions
- ▶ Oozie Coordinator jobs used to schedule recurring jobs triggered by time
- ▶ Supports many ecosystem components out of the box
  - ▶ MapReduce
  - ▶ Pig
  - ▶ Hive
  - ▶ Sqoop
- ▶ Command line and Web interface, Hue integration

# Hue

- ▶ Hadoop User Experience
- ▶ Web portal to many Ecosystem components
  - ▶ Hive
  - ▶ Pig
  - ▶ HBase
  - ▶ Impala
  - ▶ Oozie
  - ▶ Many more

# Continued Reading

Sqoop

<http://sqoop.apache.org>

Oozie

<http://oozie.apache.org>

Hue Website

<http://gethue.com>

