**Overview**

* Simplified way of writing mappers and reducers
* Sits on top of MapReduce/TEZ
* Pig Latin
  + Scripting language that has a SQL-like syntax that lets you develop Map and Reduce steps
* Performance isn’t so much of an issue anymore
  + PIG can run on top of TEZ
    - TEZ is a much more efficient way of organizing MapReduce tasks
      * Uses a DAG for task management
    - TEZ runs x10 faster than MapReduce (? As claimed by Frank)
* Highly extensible UDF’s (user-defined-functions)

**Running Pig**

* Grunt
  + For experimentation
  + Run line by line
* Script
  + Script file
* Ambari/Hue (Gui)
  + GUI view

**Example**

* Relation
  + Dataset
* PigStorage
  + Use when you have a different delimiter in the input file
* Commands
  + LOAD -> Load data from disk
  + STORE -> Store datadump into a file
  + DUMP -> Dump (print) out contents of a relation (dataset)
  + FOREACH -> iterate over each record in a relation (dataset)
  + MAPREDUCE -> Lets you call specific mappers and reducers
  + STREAM -> can stream out result
  + SAMPLE -> random samples
  + AS -> Speicifies schema
  + USING
    - … Using PigSotrage(‘|’) AS …
      * Specifying delimiter
  + GENERATE
    - Create new rows
  + DESCRIBE -> Shows the schema
    - This seems similar to MySQL
  + FILTER -> Similar to a WHERE clause
  + JOIN -> Similar to JOIN clause
    - BY -> Join on
    - JOIN appends a prefix to each column in the result set
  + CROSS -> Cartesian product. Cross join
  + CUBE -> More than 2 columns?
  + RANK -> Rank column.
    - Similar to SQL
  + LIMIT -> Top n results
    - Similar to SQL
  + UNION
  + SPLIT
  + EXPLAIN -> Query planner?
  + ILLUSTRATE -> Illustrate what’s going on under the hood
* Bag
  + Tuples related
* Functions
  + Aggregations
    - AVG
    - CONCAT
    - COUNT
    - MAX
    - MIN
    - SIZE
    - SUM
  + PigStorage -> can define delimiter
  + TextLoader ->
  + JsonLoader ->
  + AvroStorage -> Serialization & Deserialization format of data
  + ParquetLoader -> Column oriented data format
  + OrcStorage ->
  + HBaseStorage -> Interage with NoSQL DB