# Hive & Pig

Two ways of doing one thing
Or
One way of doing two things

Ashutosh Chauhan

# Who am I?

- Pig Committer & PMC Member
- Hive Committer & PMC Member
- Hcatalog Committer & PPMC Member
- ASF Member
- Software Engineer at HortonWorks

# Two ways of doing same thing

- Both generate map-reduce jobs from a query written in higher level language.
- Both frees users from knowing all the little secrets of Map-Reduce & HDFS.

# Language

- PigLatin : Procedural data-flow language
  - A = load ‘mydata’;
  - Dump A;
- HiveQL : Declarative SQLish language
  - Select * from ‘mytable’;

# Different languages = Different users

- Pig : More popular among
  - Programmers
  - Researchers
- Hive : More popular among
  - Analysts

# Different users = Different usage pattern

- Pig :
  - Programmers : Writing complex data pipelines
  - Researchers : Doing ad-hoc analysis typically employing Machine Learning
- Hive :
  - Analysts Generating daily reports

# Different Usage Pattern

**Data Collection**

**Data Factory**
**Pig**

Pipelines
Iterative Processing
Research

**Data Warehouse**
**Hive**

BI Tools
Analysis

# Different usage pattern = Different future directions

- Hive is evolving towards Data-warehousing solution. Users are asking for better integration with other systems (O/JDBC)

- Pig is evolving towards a language of its own. Users are asking for better dev environment : debugger, linter, editor etc.