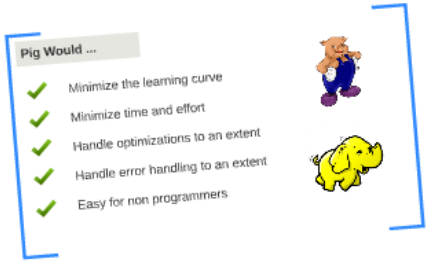
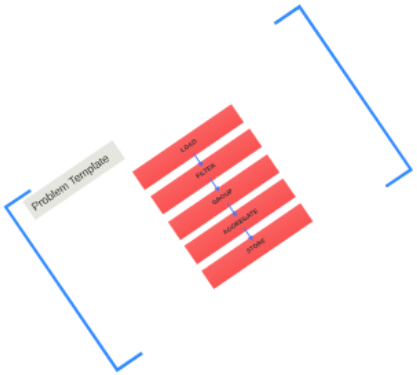
APACHE PIG





www.hadoopinrealworld.com

APACHE PIG

www.hadoopinrealworld.com

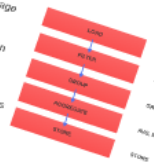
APACHE PIG

Pig Latin

Simple to use data flow language

Instructions to process data given by the user

Pig analyze, optimize the instructions and construct and submit MR jobs



Pig Would ...

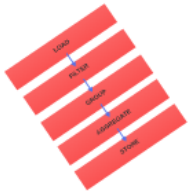
- Minimize the learning curve
- Minimize time and effort
- Handle optimizations to an extent
- Handle error handling to an extent
- Easy for non programmers

Philosophy 1 - Pigs Eat Anything

Metadata or not

Structured or not

Problem Template



Philosophy 2 - Pigs Fly

Designed with Big Data performance requirements in mind

Philosophy 3 - Pigs Are Domestic Animals

Designed to be easily controlled and modified by its users

Philosophy 4 - Pigs Live Anywhere

Pig authors envision Pig to work on any distributed processing frameworks similar to Hadoop in future.

Writing MapReduce Program - (Few) Challenges

Conceptually Visualize Problem in MapReduce

Requires lot of coding to do simple stuff for eg: Joins

Programming Language

Time & Effort

Apache Pig

- Yahoo!
- First release - Sep 2008
- Client tool
- HDFS & MapReduce
- Same implementation/phases as a MR program

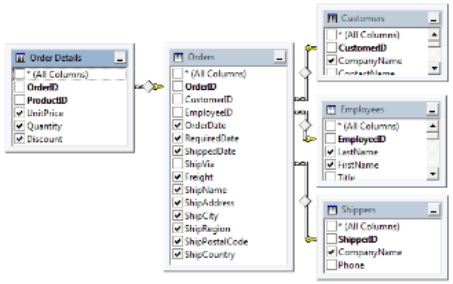
www.hadoopinrealworld.com

Writing MapReduce Program - (Few) Challenges

Conceptually Visualize Problem
in MapReduce

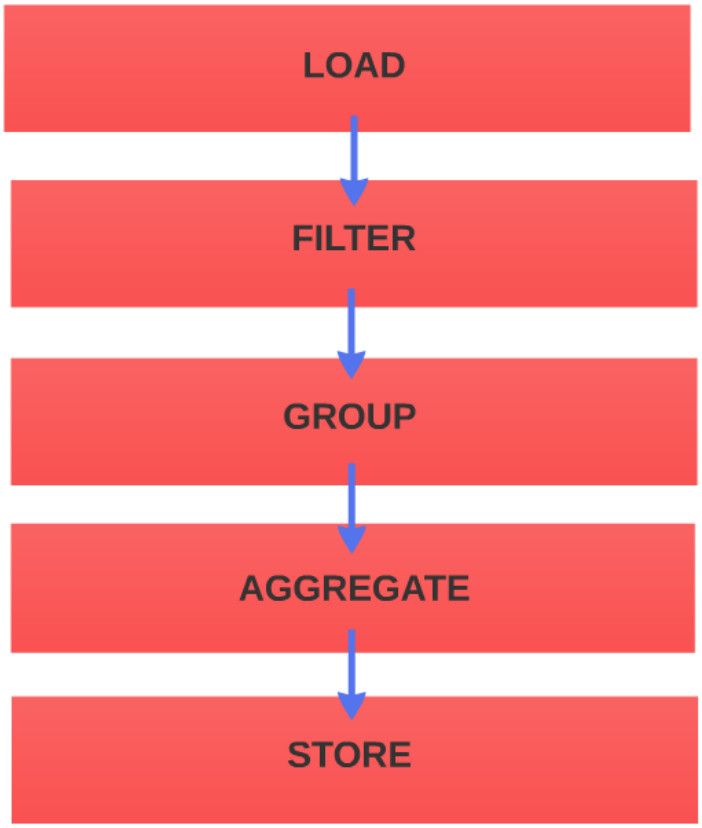
Programming Language

Requires lot of coding
to do simple stuff
for eg. Joins



Time & Effort

Problem Template



Apache Pig

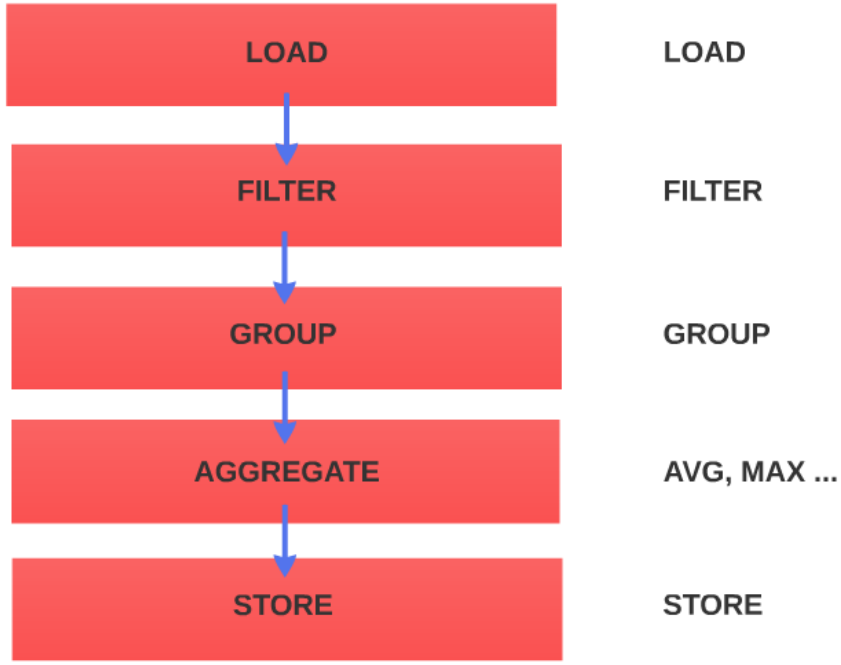
- Yahoo!
- First release - Sep 2008
- Client tool
- HDFS & MapReduce
- Same implementation/phases as a MR program

Pig Latin

Simple to use data flow language

Instructions to process data given by the user

Pig analyze, optimize the instructions and construct and submit MR jobs



Philosophy 1 - Pigs Eat Anything

Metadata or not

Structured or not

Philosophy 2 - Pigs Fly

Designed with Big Data
performance requirements in mind

Philosophy 3 -Pigs Are Domestic Animals

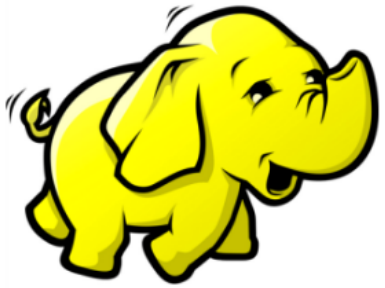
Designed to be easily controlled and
modified by its users

Philosophy 4 - Pigs Live Anywhere

Pig authors envision Pig to work on
any distributed processing frameworks
similar to Hadoop in future.

Pig Would ...

- Minimize the learning curve
- Minimize time and effort
- Handle optimizations to an extent
- Handle error handling to an extent
- Easy for non programmers



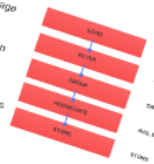
APACHE PIG

Pig Latin

Simple to use data flow language

Instructions to process data given by the user

Pig analyze, optimize the instructions and construct and submit MR jobs



Pig Would ...

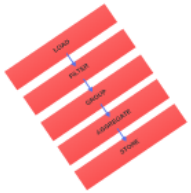
- Minimize the learning curve
- Minimize time and effort
- Handle optimizations to an extent
- Handle error handling to an extent
- Easy for non programmers

Philosophy 1 - Pigs Eat Anything

Metadata or not

Structured or not

Problem Template



Philosophy 2 - Pigs Fly

Designed with Big Data performance requirements in mind

Philosophy 3 - Pigs Are Domestic Animals

Designed to be easily controlled and modified by its users

Philosophy 4 - Pigs Live Anywhere

Pig authors envision Pig to work on any distributed processing frameworks similar to Hadoop in future.

Writing MapReduce Program - (Few) Challenges

Conceptually Visualize Problem in MapReduce

Requires lot of coding to do simple stuff for eg: Joins

Programming Language

Time & Effort

Apache Pig

- Yahoo!
- First release - Sep 2008
- Client tool
- HDFS & MapReduce
- Same implementation/phases as a MR program

www.hadoopinrealworld.com