

Apache Pig Philosophy

Table of contents

1 What does it mean to be a pig?.....	2
2 Pigs Eat Anything.....	2
3 Pigs Live Anywhere.....	2
4 Pigs Are Domestic Animals.....	2
5 Pigs Fly.....	2

1 What does it mean to be a pig?

The Apache Pig Project has some founding principles that help pig developers decide how the system should grow over time. This page presents those principles.

2 Pigs Eat Anything

Pig can operate on data whether it has metadata or not. It can operate on data that is relational, nested, or unstructured. And it can easily be extended to operate on data beyond files, including key/value stores, databases, etc.

3 Pigs Live Anywhere

Pig is intended to be a language for parallel data processing. It is not tied to one particular parallel framework. It has been implemented first on Hadoop, but we do not intend that to be only on Hadoop.

4 Pigs Are Domestic Animals

Pig is designed to be easily controlled and modified by its users.

Pig allows integration of user code where ever possible, so it currently supports user defined field transformation functions, user defined aggregates, and user defined conditionals. These functions can be written in Java or scripting languages that can compile down to Java (e.g. Jython). Pig supports user provided load and store functions. It supports external executables via its stream command and Map Reduce jars via its mapreduce command. It allows users to provide a custom partitioner for their jobs in some circumstances and to set the level of reduce parallelism for their jobs. command. It allows users to set the level of reduce parallelism for their jobs and in some circumstances to provide a custom partitioner.

Pig has an optimizer that rearranges some operations in Pig Latin scripts to give better performance, combines Map Reduce jobs together, etc. However, users can easily turn this optimizer off to prevent it from making changes that do not make sense in their situation.

5 Pigs Fly

Pig processes data quickly. We want to consistently improve performance, and not implement features in ways that weigh pig down so it can't fly.