Pig Philosophy

Table of contents

1. What does it mean to be a pig?

The 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.

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, user defined grouping functions, and user defined conditionals. In the future we want to support all the above in non-java languages, as well as streaming, user defined types, and user defined splits.

Currently pig has no optimizer, so it does not do any operation rearranging. When we add that in the future, it will always be possible for users to turn code rearranging off, so that pig does exactly what they say in the order they say it.

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.