

Welcome to Apache Pig!

Table of contents

1 News.....	2
1.1 Apache Pig 0.17.0 is released!.....	2
2 Getting Started	2
3 Getting Involved	2

Apache Pig is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets.

At the present time, Pig's infrastructure layer consists of a compiler that produces sequences of Map-Reduce programs, for which large-scale parallel implementations already exist (e.g., the Hadoop subproject). Pig's language layer currently consists of a textual language called Pig Latin, which has the following key properties:

- **Ease of programming.** It is trivial to achieve parallel execution of simple, "embarrassingly parallel" data analysis tasks. Complex tasks comprised of multiple interrelated data transformations are explicitly encoded as data flow sequences, making them easy to write, understand, and maintain.
- **Optimization opportunities.** The way in which tasks are encoded permits the system to optimize their execution automatically, allowing the user to focus on semantics rather than efficiency.
- **Extensibility.** Users can create their own functions to do special-purpose processing.

1 News

1.1 Apache Pig 0.17.0 is released!

The highlights of this release is the introduction of Pig on Spark. See details on the [release page](#).

2 Getting Started

1. [Read the documentation](#).
2. [Find training resources](#).
3. [Discuss it](#) on the mailing list.

3 Getting Involved

Pig is an open source volunteer project under the Apache Software Foundation. We encourage you to learn about the project and contribute your expertise. Here are some starter links:

1. [How to contribute](#)
2. Give us [feedback](#): What can we do better?
3. Join the [mailing list](#): Meet the community.