Apache Pig is a platform for analyzing large data sets. Pig’s language, Pig Latin, lets you specify a sequence of data transformations such as merging data sets, filtering them, and applying functions to records or groups of records. Pig comes with many built-in functions but you can also create your own user-defined functions to do special-purpose processing.

Pig Latin programs run in a distributed fashion on a cluster (programs are compiled into Map/Reduce jobs and executed using Hadoop). For quick prototyping, Pig Latin programs can also run in "local mode" without a cluster (all processing takes place in a single local JVM).

Do you Pig? At Yahoo! 40% of all Hadoop jobs are run with Pig. Come join us!

General Information

- Why Pig Latin instead of SQL? Pig Latin: A Not-So-Foreign Language ...
- Official Apache Pig Website
- Pig Talks Papers - Pig talks, papers, interviews
- PoweredBy - a (partial) list of companies using Pig
- Pig book: Programming Pig

User Documentation

- User Documentation
- PiggyBank - User-defined functions (UDFs) contributed by Pig users!
- PigTools - Tools Pig users have built around and on top of Pig.
- PigInteroperability - How to make Pig work with other platforms you may be using, such as HBase and Cassandra.
- Penny - A distributed debugging framework for Pig.
- Pig Tutorial
- FAQ

Developer Documentation

- How to
  - HowToDocument
  - HowToContribute
  - HowToCommit
  - HowToRelease
  - HowToTest
  - How to set up Eclipse environment
- Road map
  - Pig Journal
- Specification Proposals
  - PigErrorHandlingFunctionalSpecification
  - PigTestProposal
- Design proposals
  - PigInMapCombinerProposal
- Guide for new contributors

Related Resources

- Hadoop Core
- Hadoop Wiki
- Link to old Pig wiki

Thanks

YourKit is kindly supporting open source projects with its full-featured Java Profiler. YourKit, LLC is the creator of innovative and intelligent tools for profiling Java and .NET applications. Take a look at YourKit’s leading software products: YourKit Java Profiler and YourKit .NET Profiler.