



# Apache Pig

Big Data

Von:

Mario Groneick

Marius Meisenzahl

Janis Mohr

# Übersicht

- Was ist Pig
- Vorteile von Pig
- Nachteile von Pig
- Datenfluss
- Demo



# Was ist Pig

- Analyse von Datensätzen
  - Big Data
- Von Yahoo entwickelt
  - 2008
- Basiert auf MapReduce
- Besitzt eigene Skriptsprache
  - Pig Latin

# Vorteile von Pig

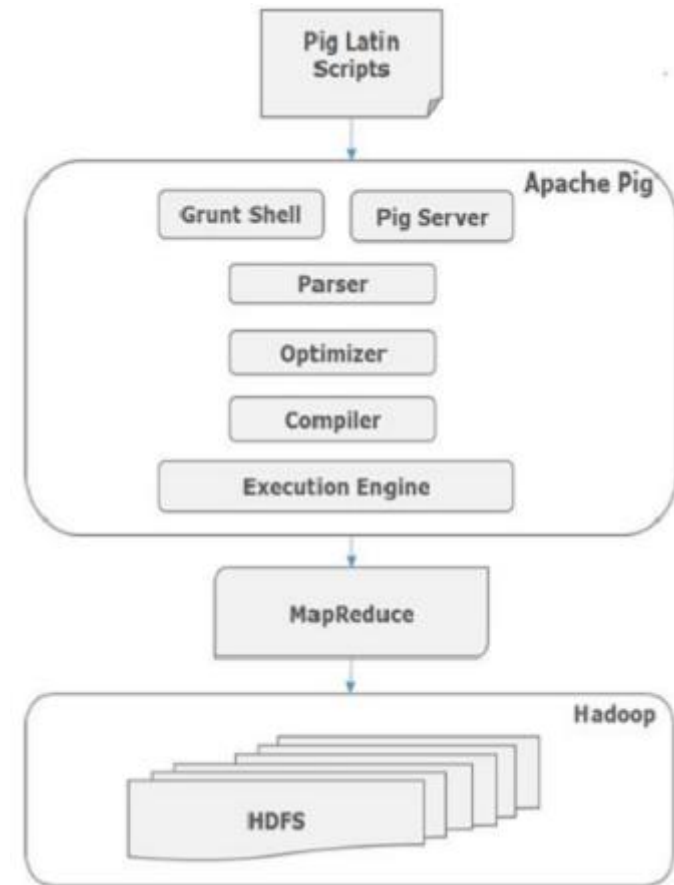
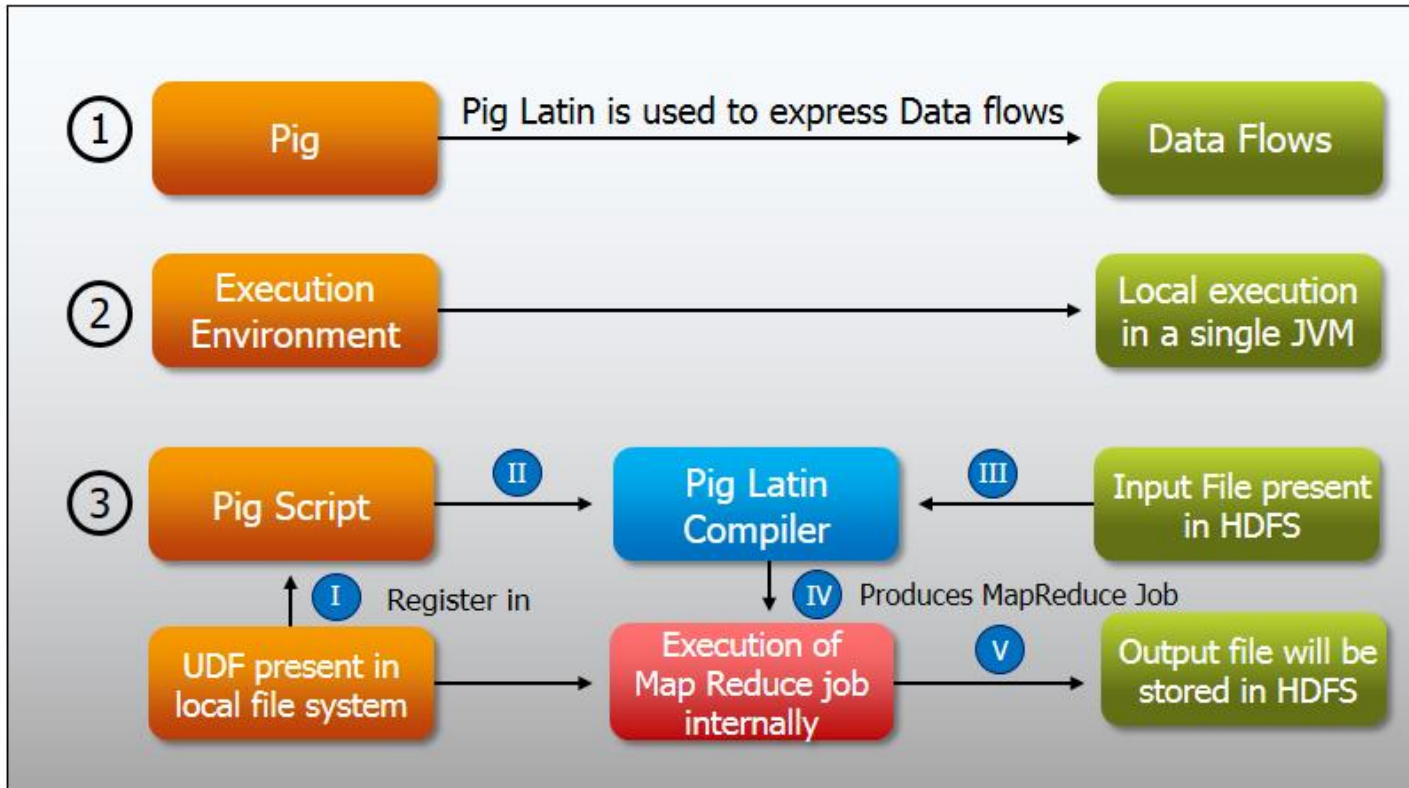
- Einfache Implementierung
  - Keine Java Kenntnisse notwendig
- Jobs lassen sich trotzdem erweitern
  - Spezielle Anforderungen lassen sich einbauen
- Line-of-code Reduction
  - Ca. 200 Zeilen MapReduce entsprechen 10 Zeilen Pig Latin

# Nachteile

- Spezielle Aufgaben müssen extern gehandelt werden
  - MapReduce ist nicht komplett Überflüssig
  - Java Kenntnisse dann notwendig



# Datenfluss





Demo





Vielen Dank für Ihre Aufmerksamkeit



# Quellen

- <https://www.youtube.com/watch?v=GG-VRm6XnNk>
- <http://www.searchenterprisesoftware.de/definition/Apache-Pig>
- <https://pig.apache.org/>
- <https://www.bigdata-insider.de/apache-pig-datenanalyse-fuer-hadoop-a-476349/>
- [https://www.tutorialspoint.com/apache\\_pig/images/apache\\_pig\\_architecture.jpg](https://www.tutorialspoint.com/apache_pig/images/apache_pig_architecture.jpg)
- <https://d1jnx9ba8s6j9r.cloudfront.net/blog/wp-content/uploads/2013/11/Presentation111.png>