



OINK: Make PIG Self-serve

Problem Statement

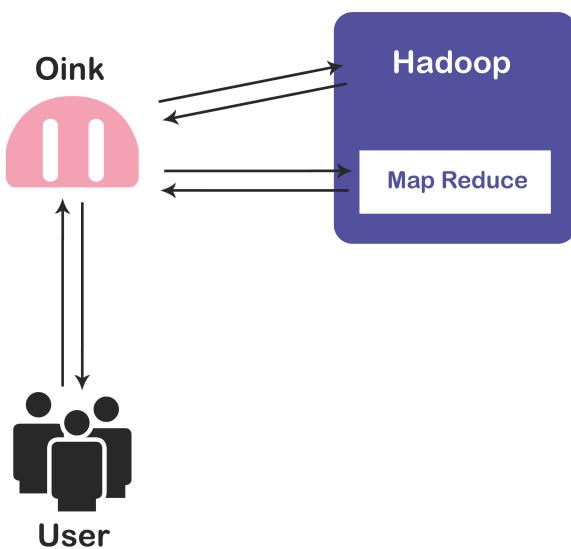
User needs to get access to Hadoop cluster. On-boarding process takes time and is cumbersome.

No way to control the number of jobs being submitted.

User needs to be exposed to the cluster configuration.

No easy process when common libraries need to be upgraded/changed.

Architecture



REST APIs

POST	/pig/jar/\$jarName
POST	/pig/script/\$scriptName
POST	/pig/request/\$scriptName
GET	/pig/request/\$requestId/status
GET	/pig/request/\$requestId/output
GET	/pig/request/\$requestId/stats

Perks / Porks :P



6000
PIG
JOBS
DAILY



>100
Unique
PIG scripts
registered



>100
Unique
users



No
on-boarding
time/
process

Serves as single entry point for PIG job execution where rate limiting and QoS can be applied

Provides a multi-tenant environment for different kinds of Pig executions

No manual intervention for getting user access etc

Abstraction of Hadoop cluster configuration from users