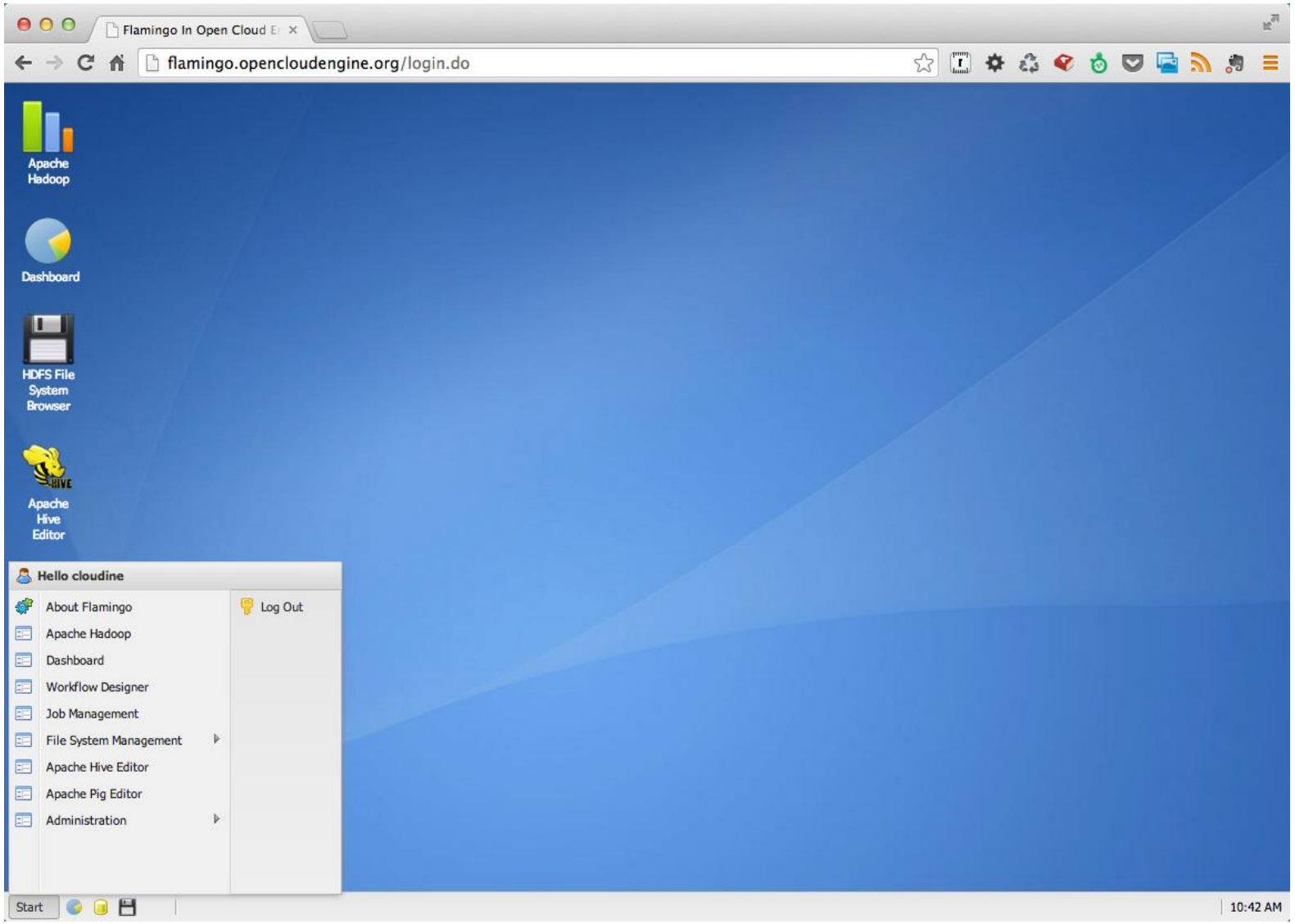
알기 쉬운 Hadoop 기반 빅데이터 플랫폼 아키텍처 및 활용 방안

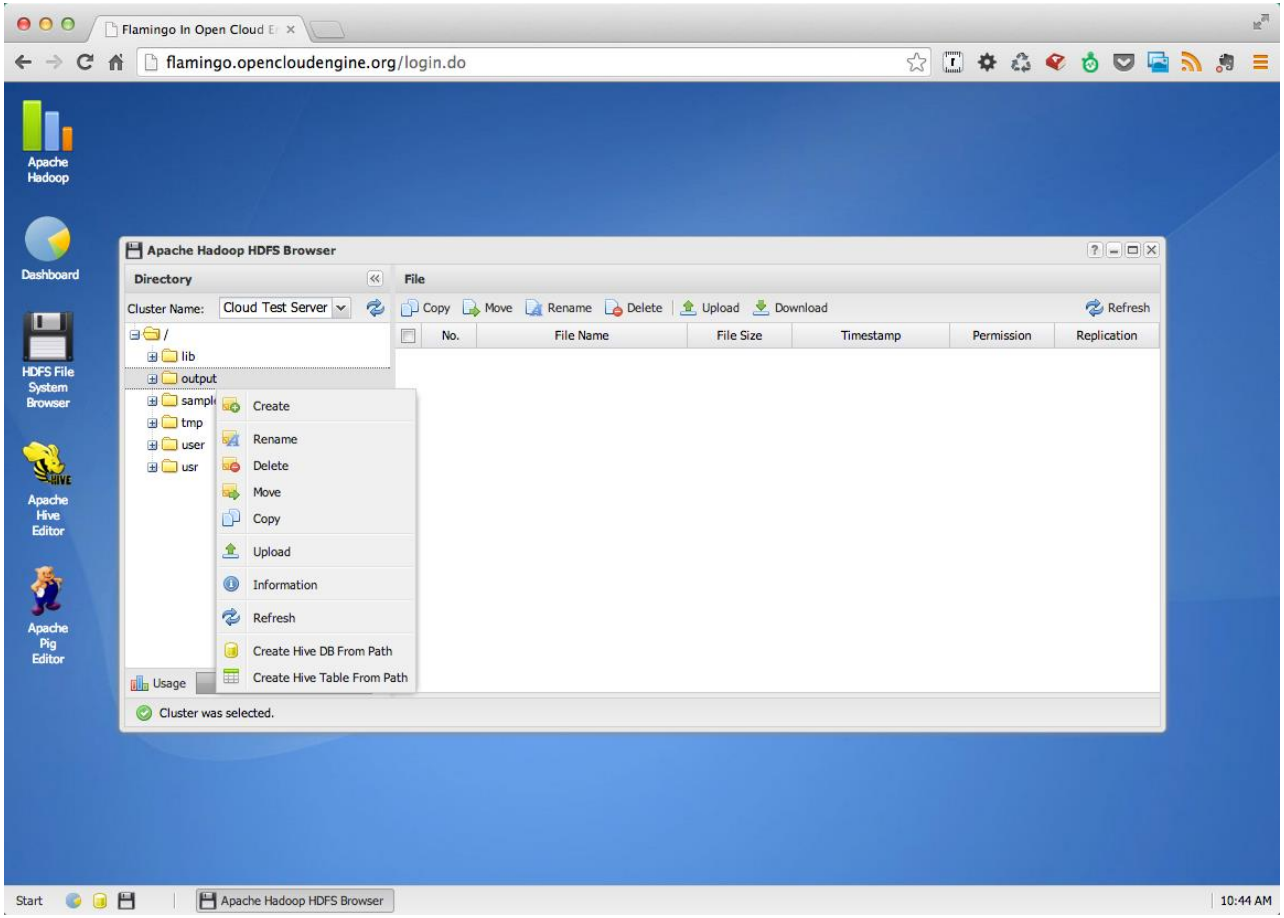
NIPA Open Frontier Lab
박치완
chiwanpark@icloud.com

Flamingo 환경


Flamingo In Open Cloud
flamingo.opencloudengine.org/login.do
Apache Hadoop
Dashboard
HDFS File System Browser
Apache Hive Editor
Hello cloudine
About Flamingo
Apache Hadoop
Dashboard
Workflow Designer
Job Management
File System Management
Apache Hive Editor
Apache Pig Editor
Administration
Log Out
Start
10:42 AM

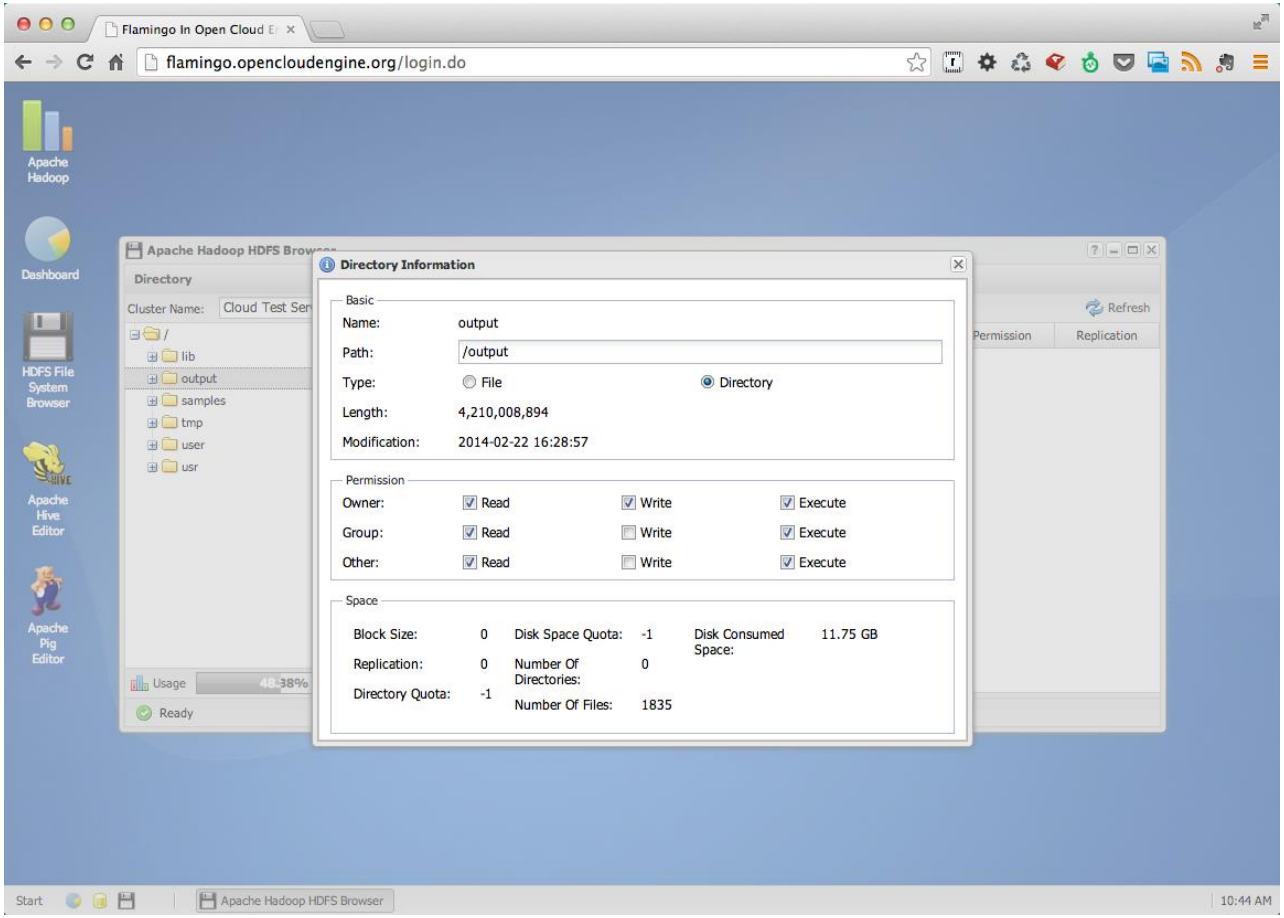
HDFS Browser

HDFS 관리 화면
– 디렉토리 관리, 파일 업로드, 다운로드, Hive Metastore 관리



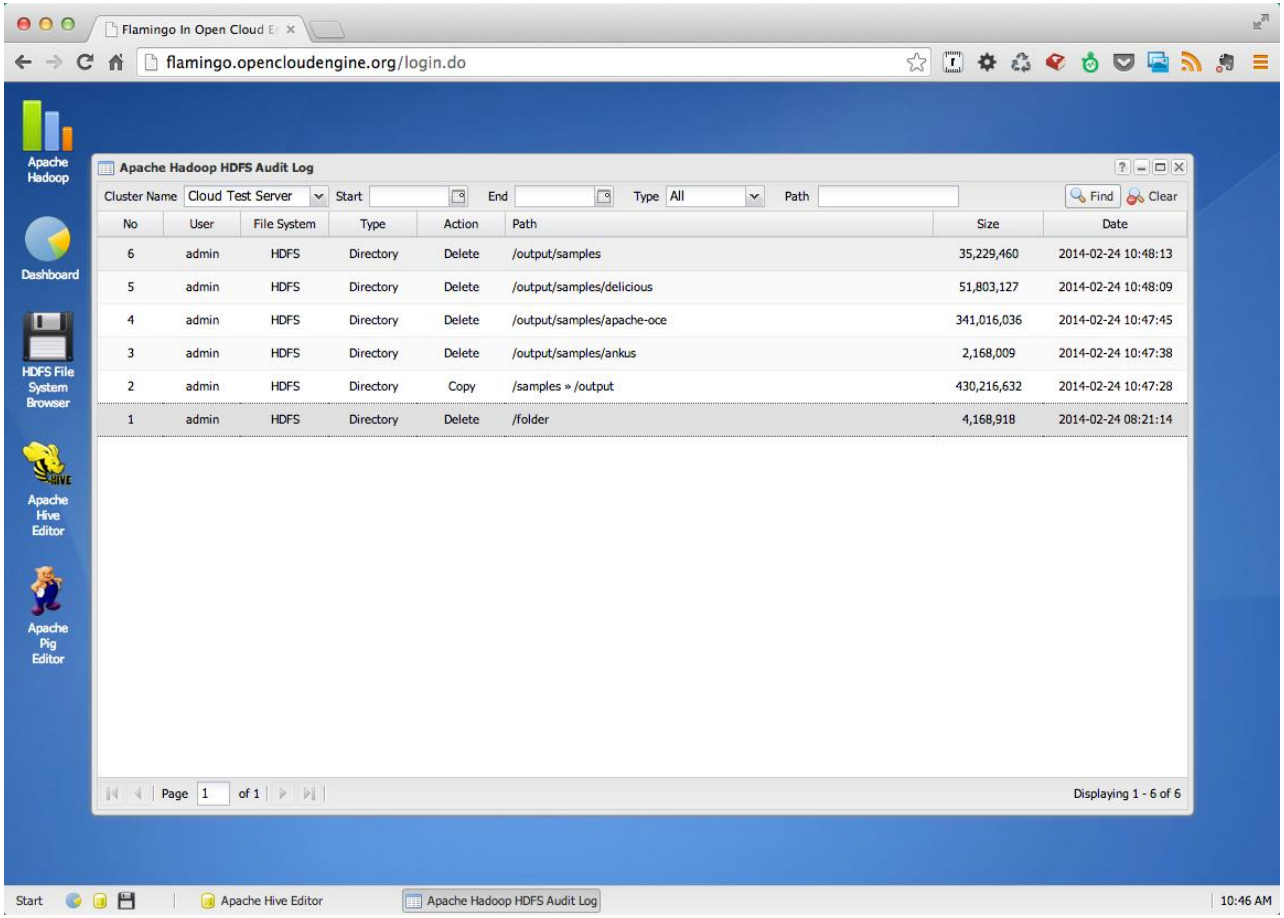
HDFS Browser

HDFS에 저장된 디렉토리 또는 파일의 상세 정보 열람



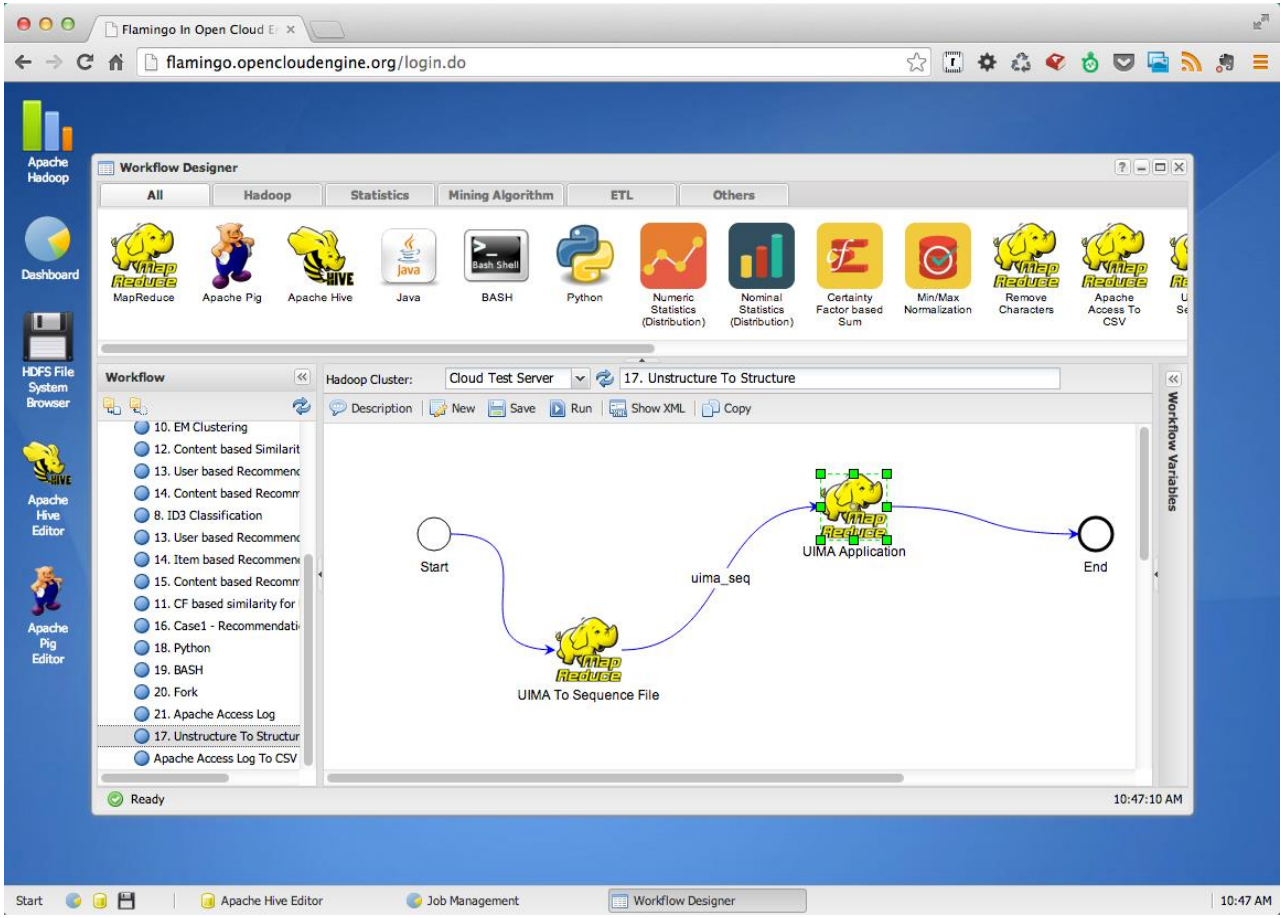
HDFS Browser

HDFS의 파일 처리 이력을 기록해 추후 관리자가 특정 파일의 이력 추적 가능



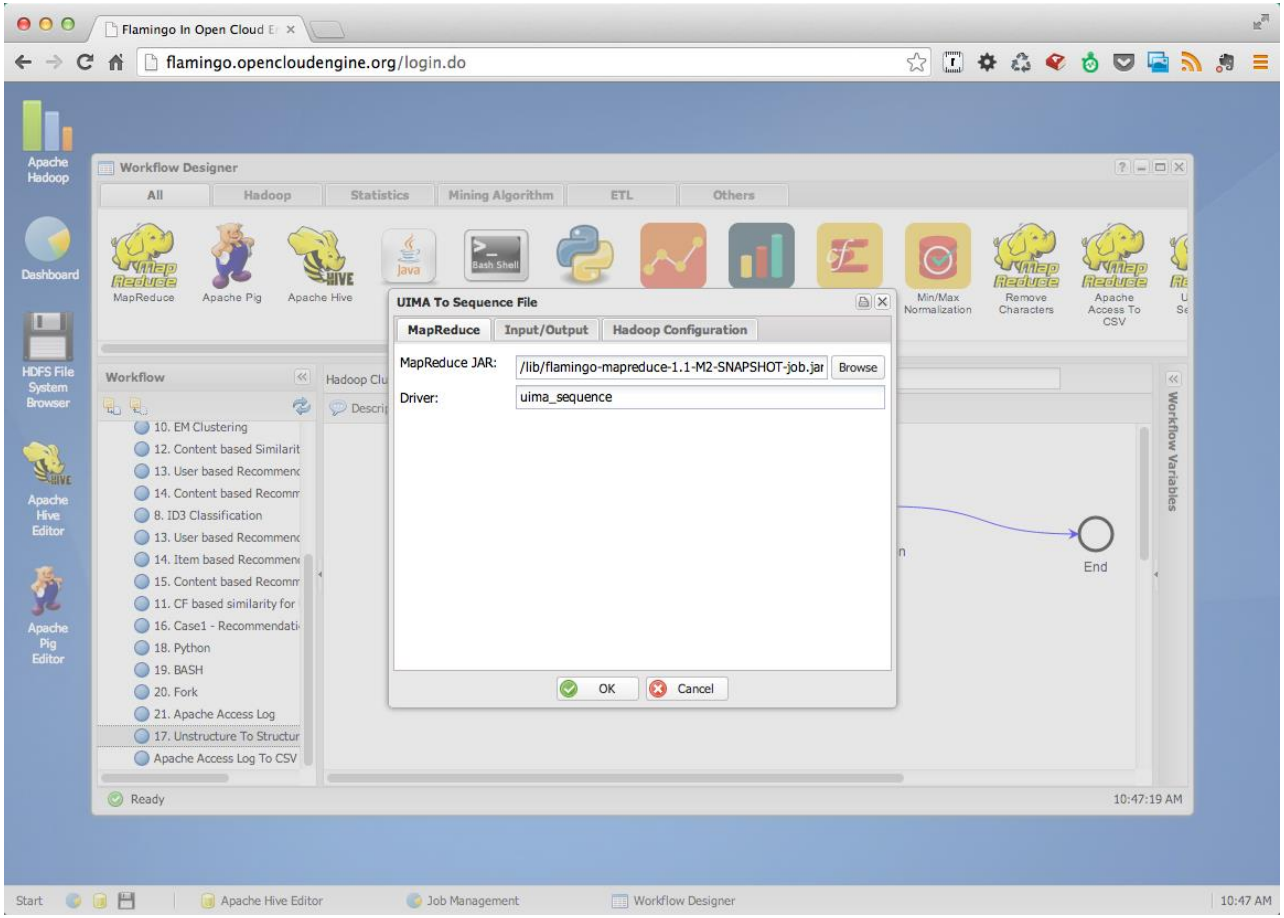
Workflow Designer

데이터 분석 Workflow를 DAG 기반으로 구성하여 작성
- 재사용 가능한 여러 컴포넌트 제공



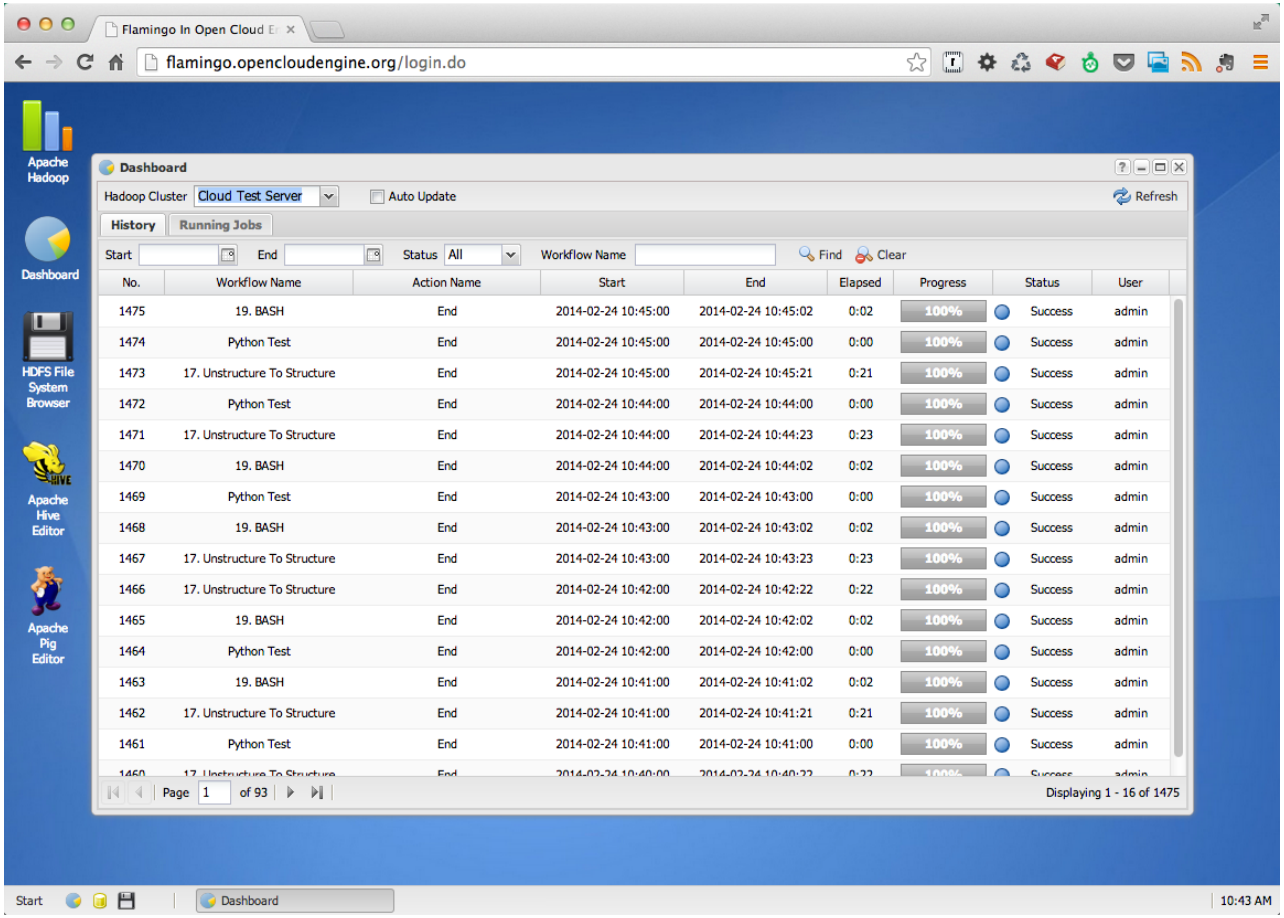
Workflow Designer

기존에 사용하던 MapReduce도 통합하여 Workflow에 적용 가능



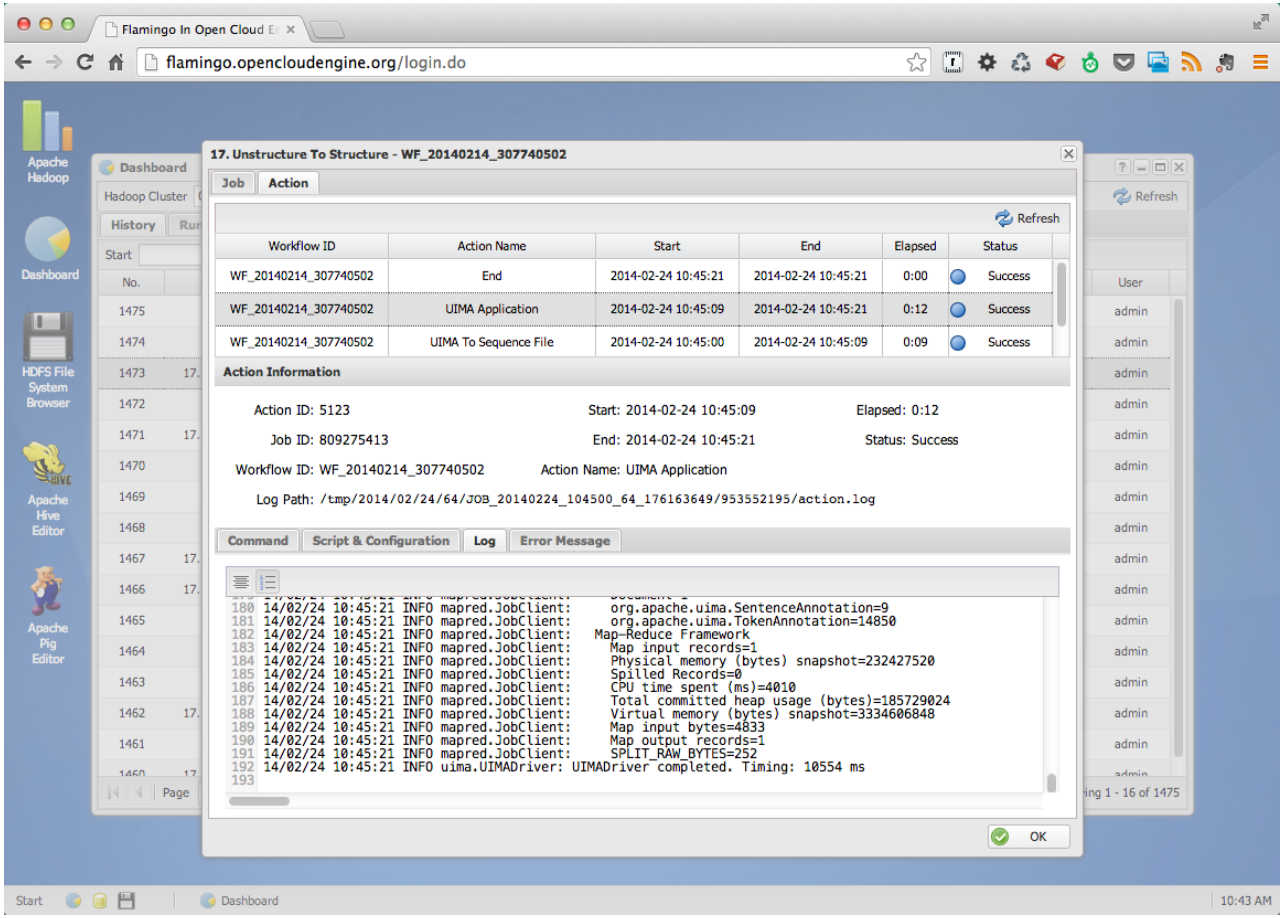
Workflow Dashboard

수행된 Workflow의 기록 열람



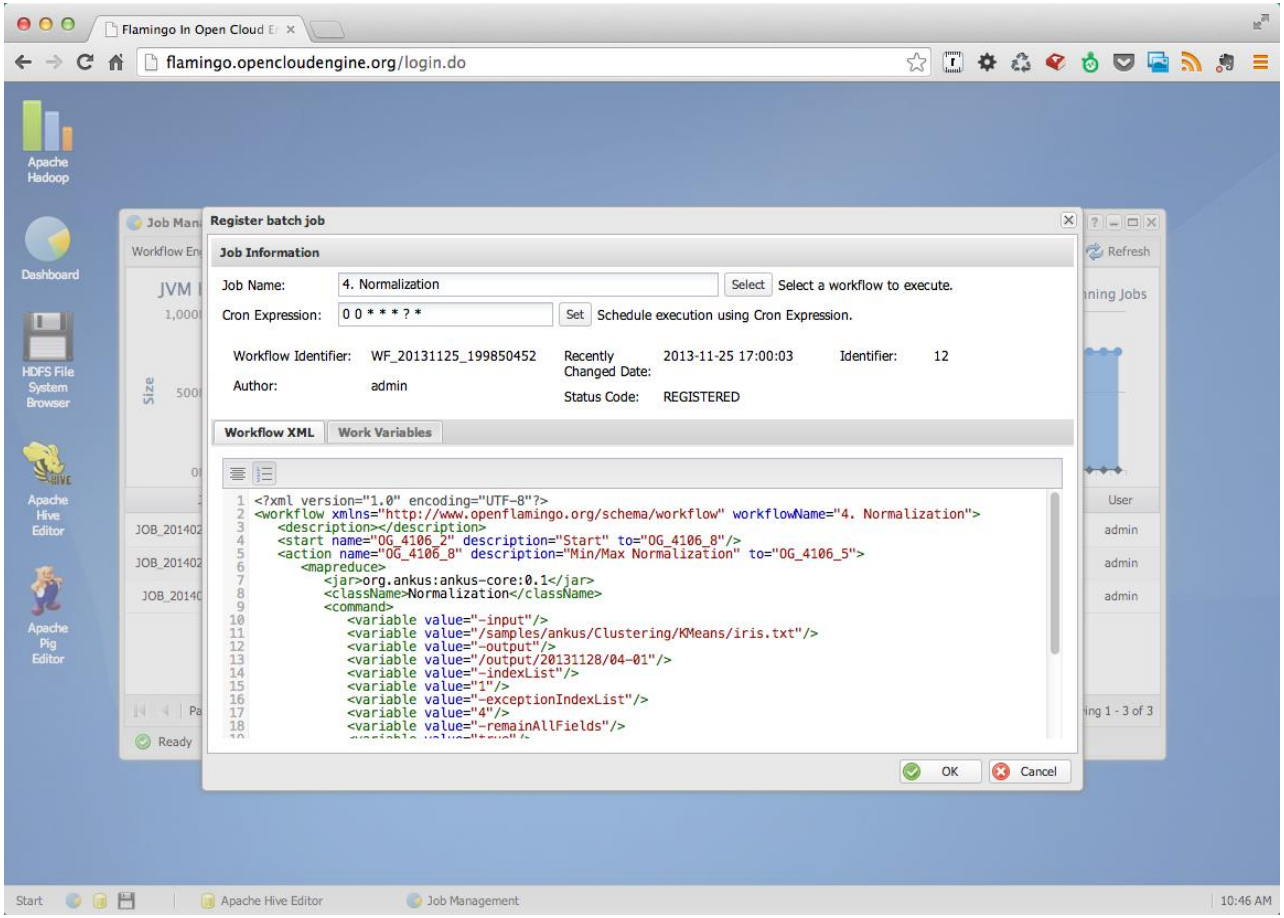
Workflow Dashboard

Workflow내 개별 MapReduce Job 마다 수행 Log, Configuration 확인



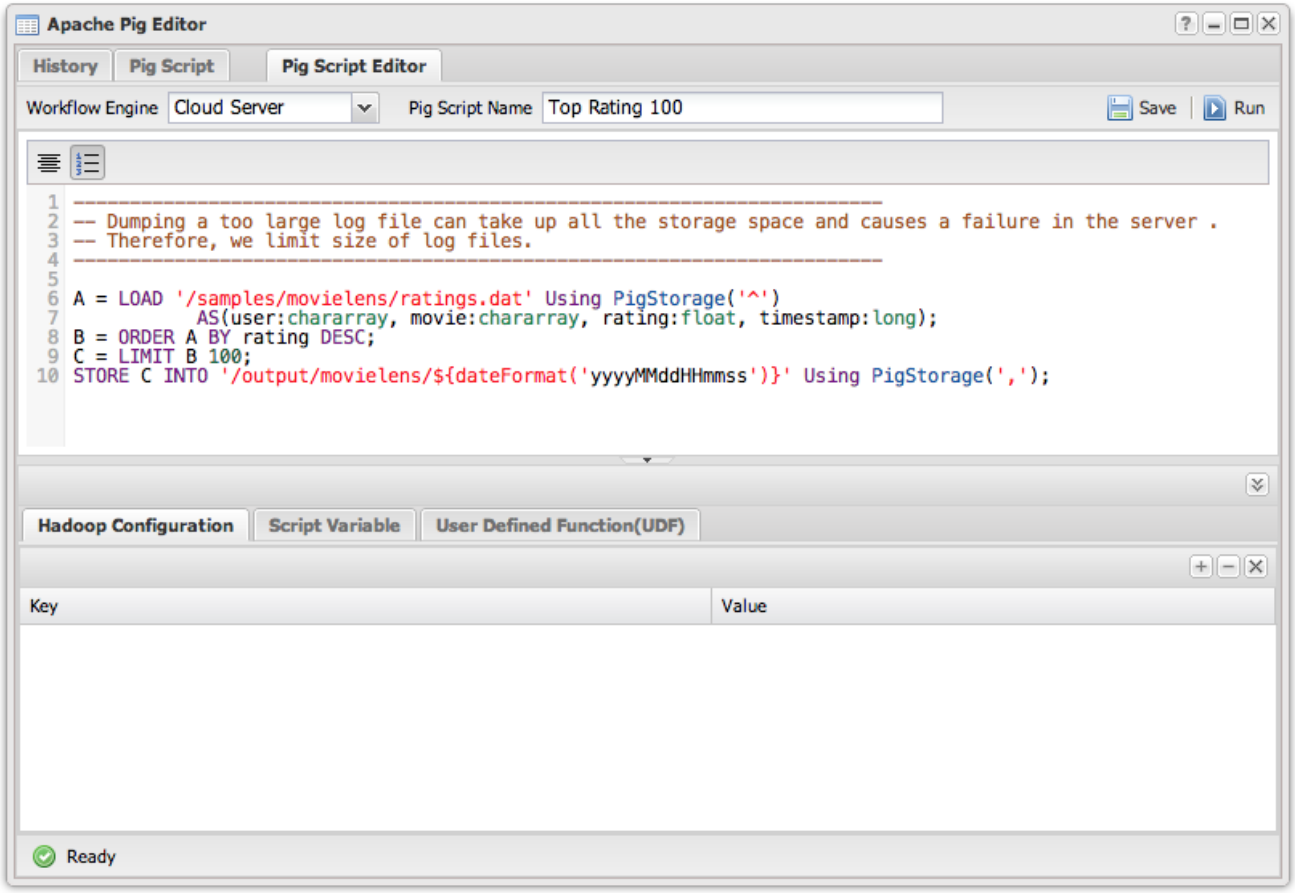
Job Scheduling

구성된 Workflow를 Batch Job으로 등록



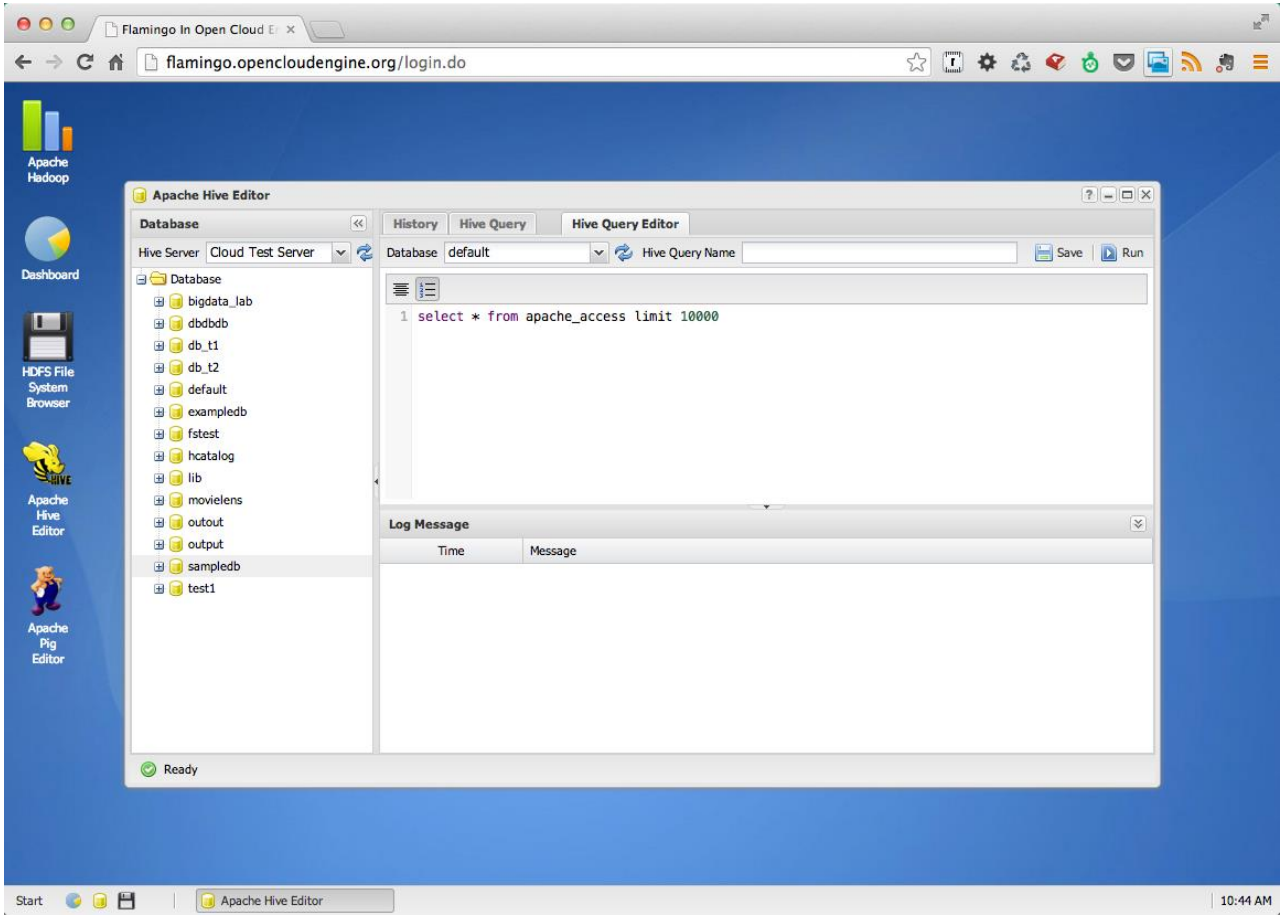
Apache Pig Integration

Apache Pig를 통합해, Pig Latin을 통해 추상화 된 데이터 분석



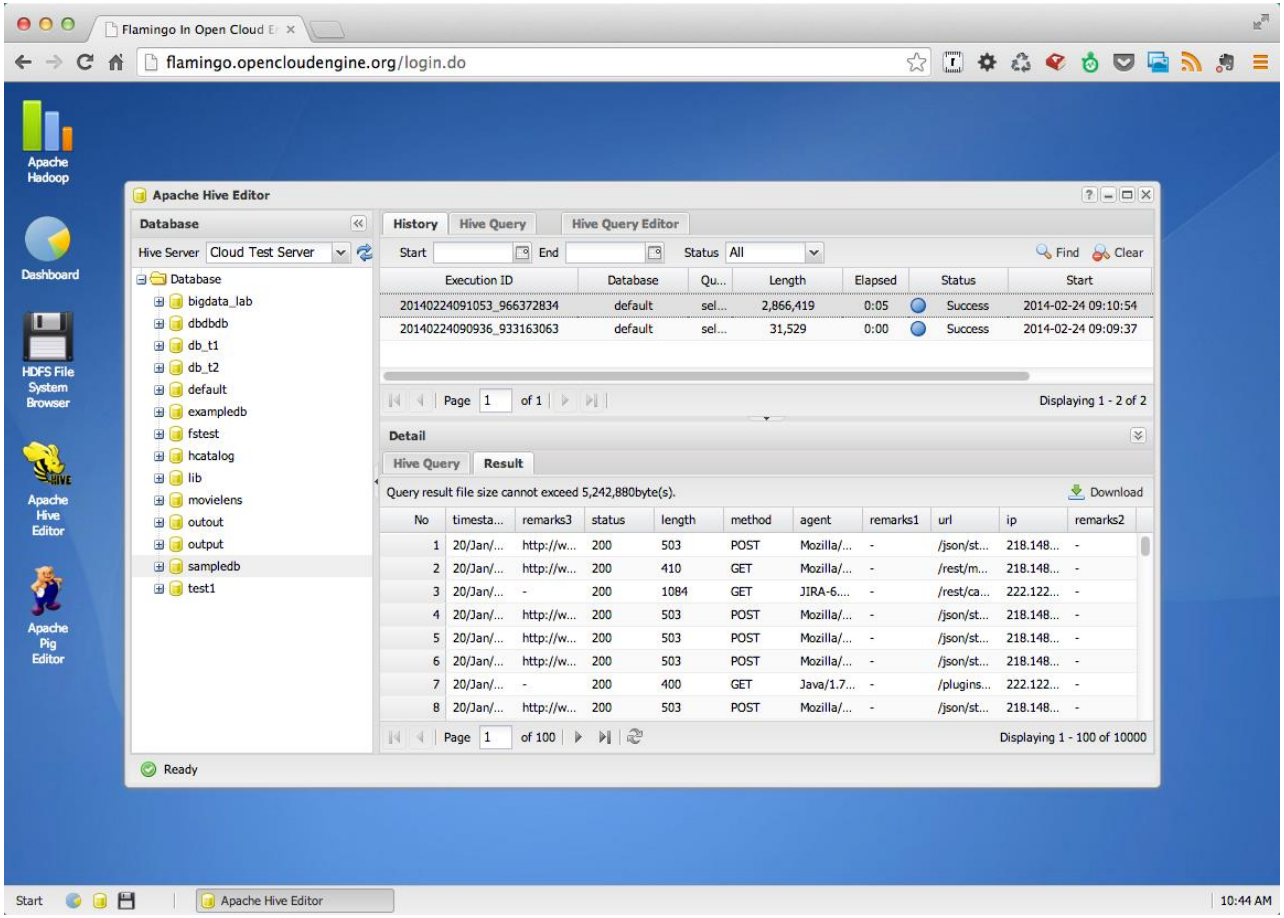
Apache Hive Integration

Apache Hive를 통합해, HiveQL을 통해 SQL과 비슷한 구문으로 데이터 분석 수행



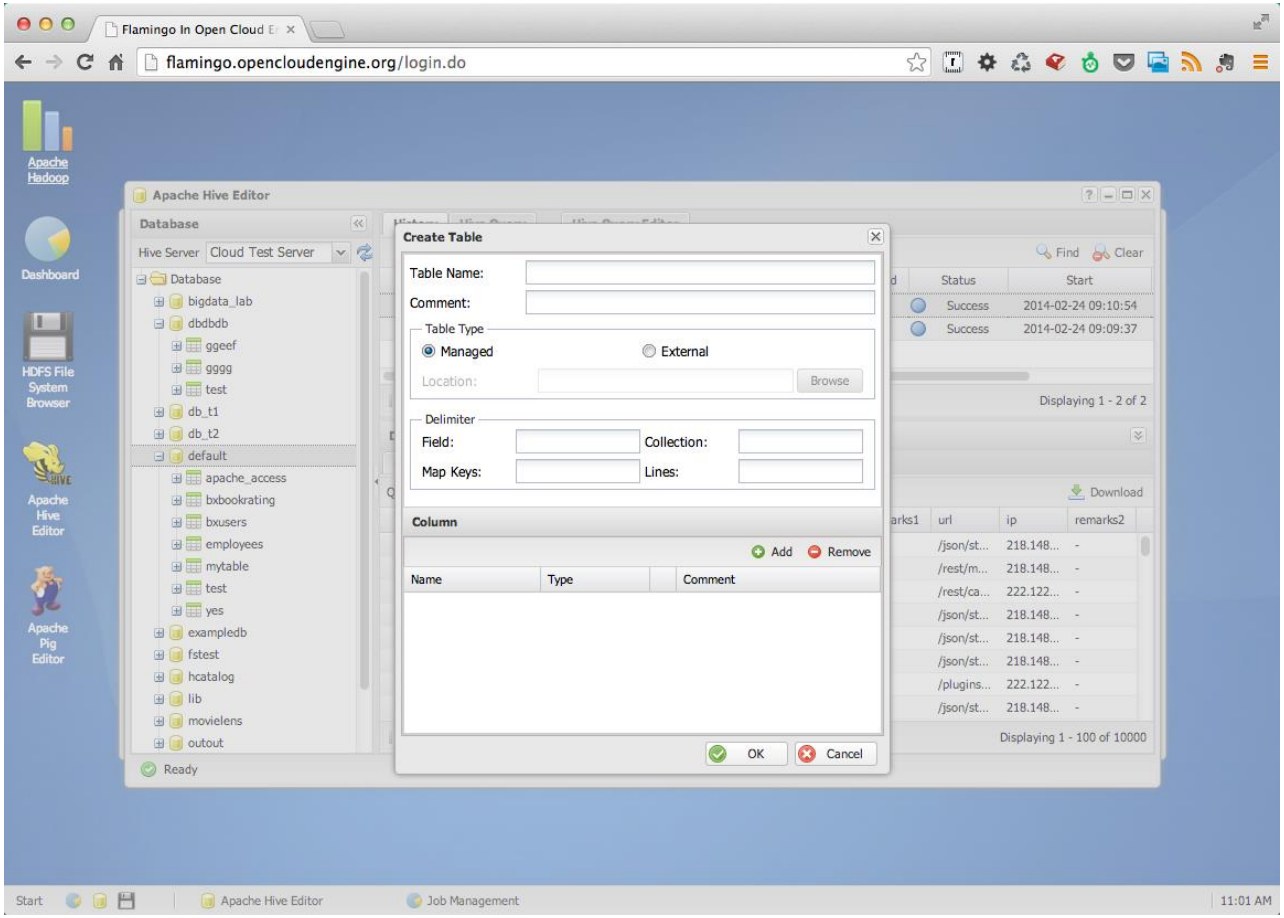
Apache Hive Integration

수행된 Query의 결과를 즉시 열람



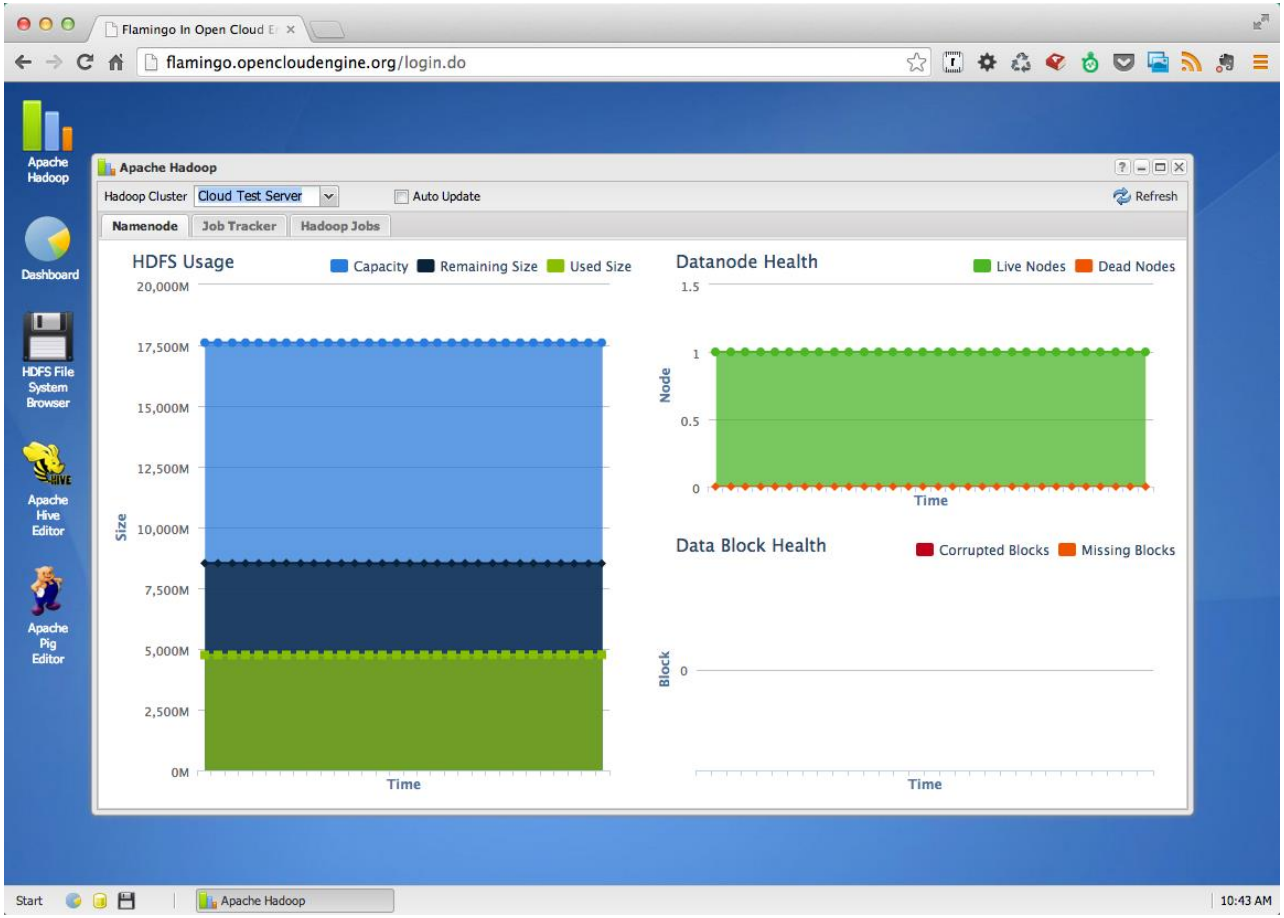
Apache Hive Integration

Hive Table를 Query문 없이 UI 만으로 생성



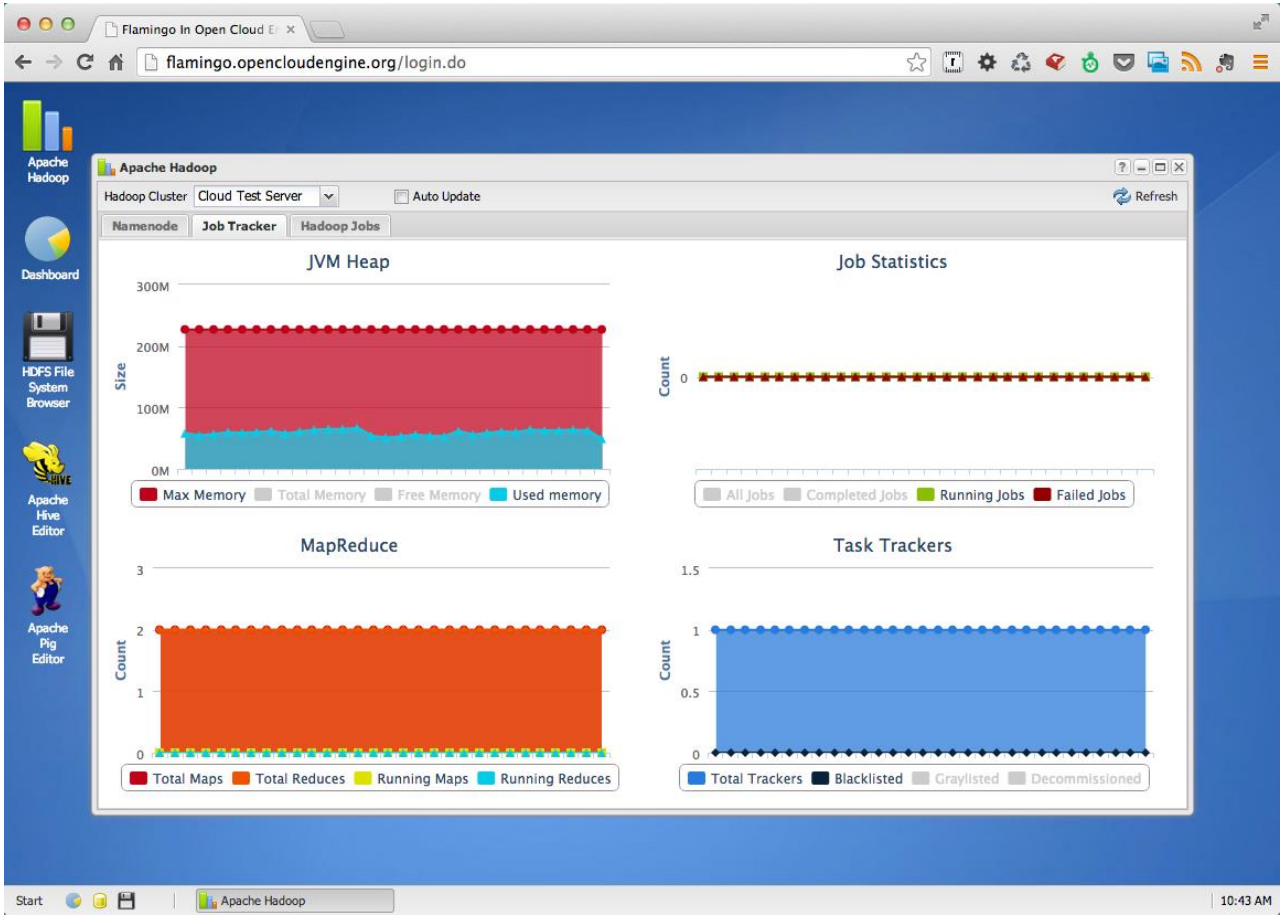
Monitoring

다양한 Monitoring 기능 제공
- HDFS, Datanode Monitoring



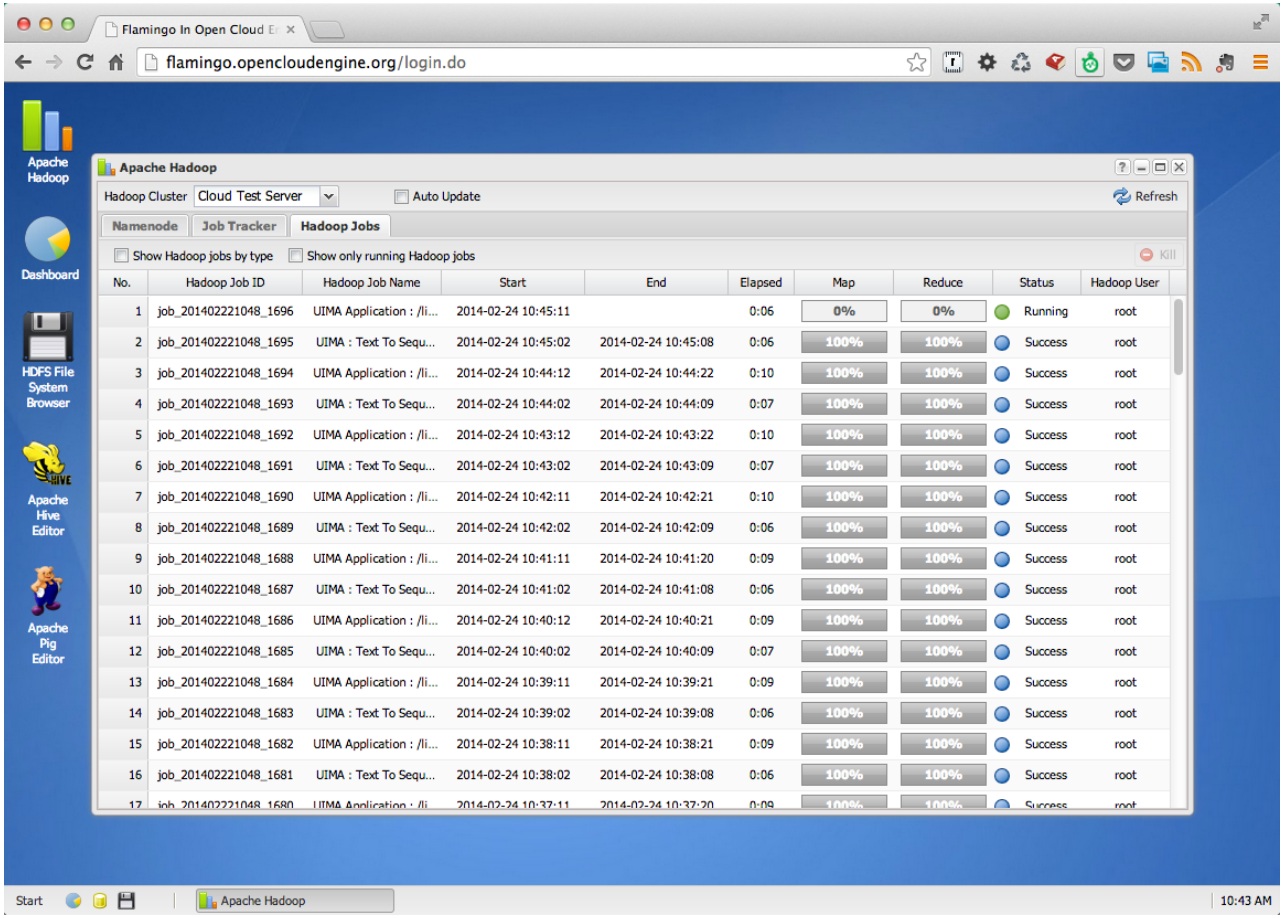
Monitoring

JobTracker, Namenode 등 주요 Hadoop Cluster의 Metrics 제공



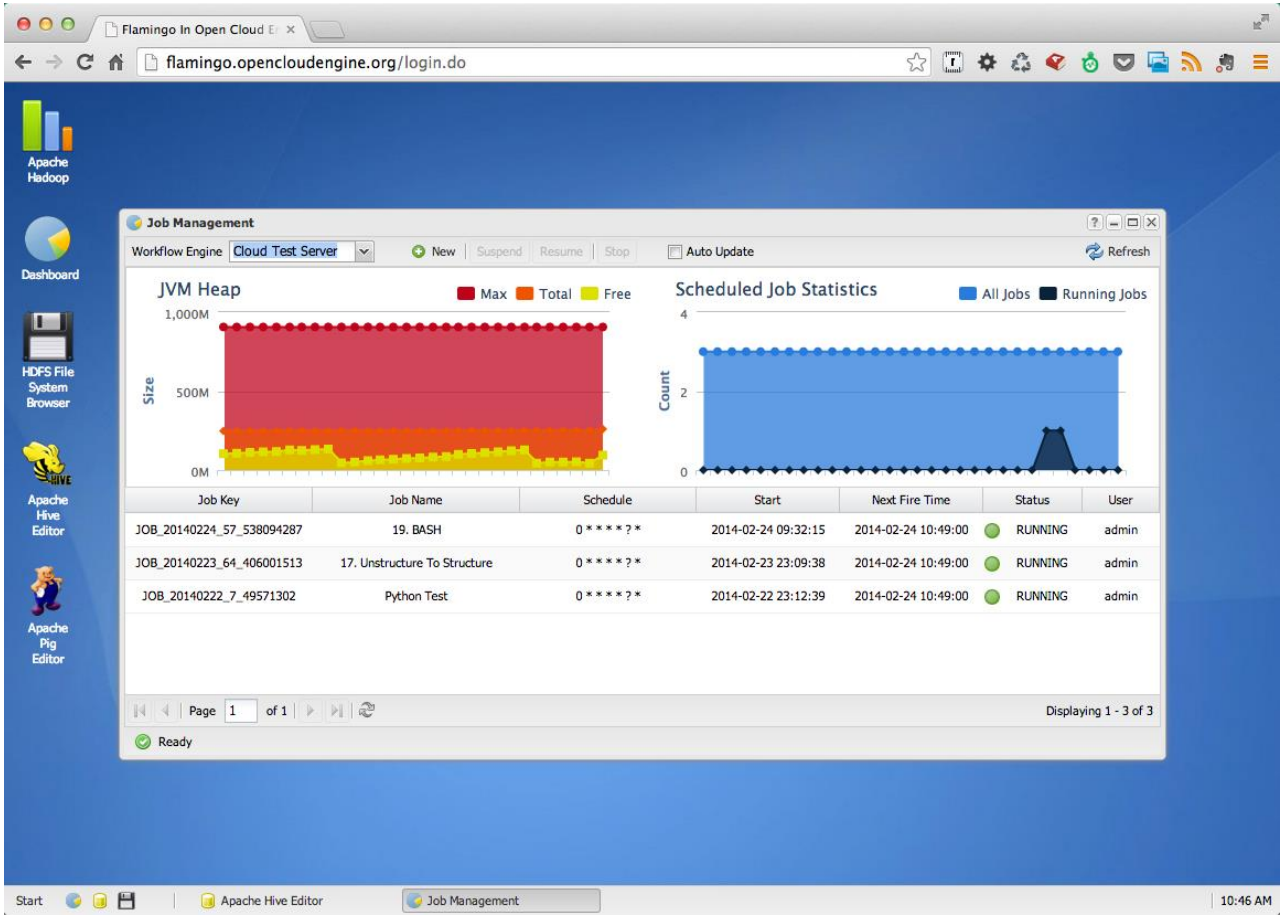
Monitoring

수행한 MapReduce Job에 대한 Monitoring 정보 제공

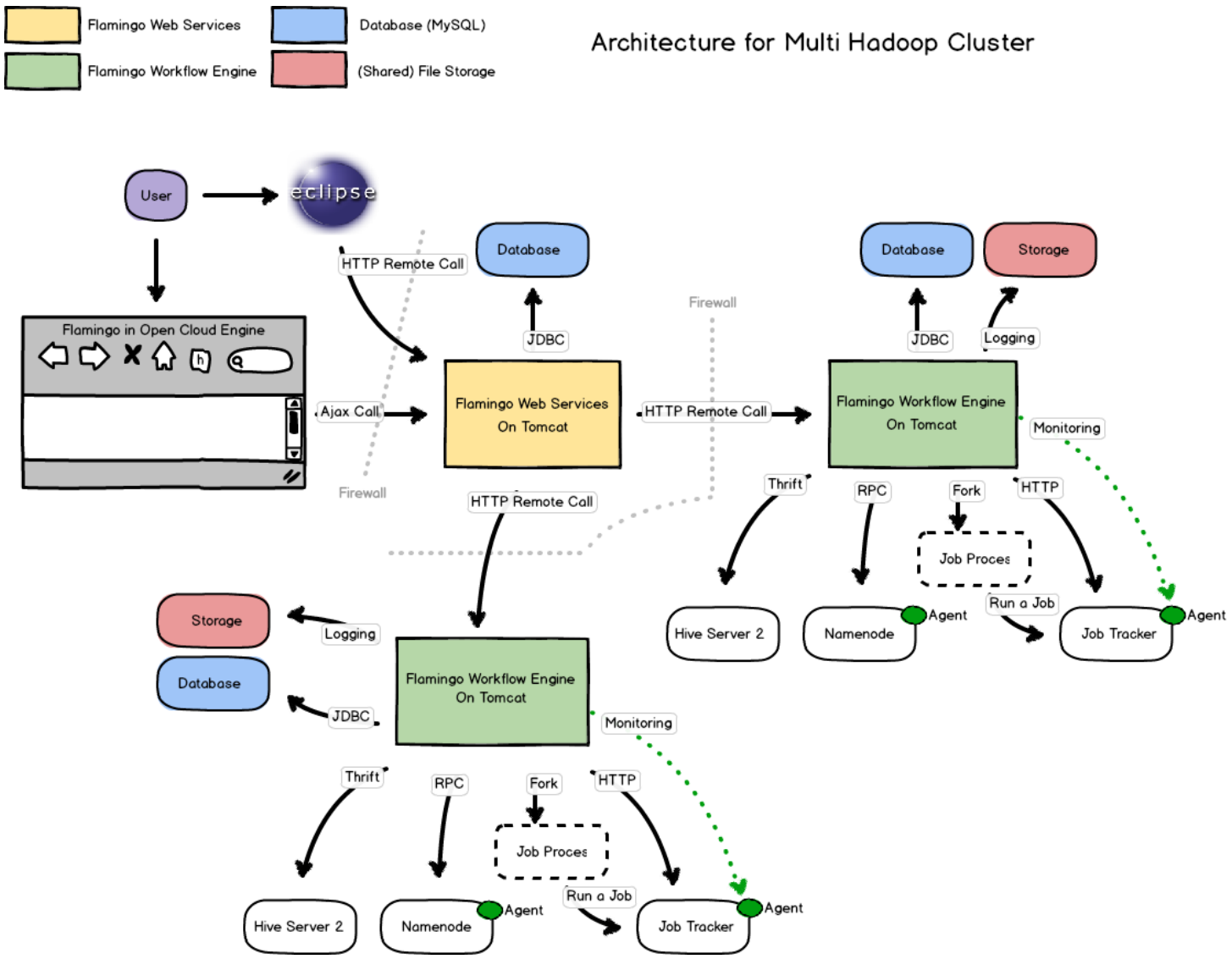


Monitoring

JobTracker, Namenode 등 주요 Hadoop Cluster의 Metrics 제공



Architecture



Future

- 기본적으로 제공되는 Component 보강
  - 데이터 전처리 모듈, 추론 모듈 등
- 사용자 별 Quota 설정 등 개별 제한 기능
- Hadoop 2 지원
- Amazon EMR, Rackspace Hadoop Platform 등 Hadoop 기반 엔터프라이즈 플랫폼 지원

Project Information

- 프로젝트 홈페이지
  - http://wiki.opencloudengine.org/display/IN/Flamingo
- Issue Tracker
  - http://jira.opencloudengine.org
- Build Server
  - http://build.opencloudengine.org
- License
  - Web UI (Ext.JS를 활용해 GPLv3)
  - Engine (Apache License)

DEMO

Q&A