



빅 데이터 분석 및 처리를 위한 웹 기반 관리 도구

ONYCOM

이 과제의 Mission

- 누구나 쉽게 Big Data OpenSource를 사용할 수 있어야 함
 - 웹을 기반으로 각종 오픈소스를 통합
 - 예) 분산 파일 시스템 브라우저(HDFS)
- 기존 데이터 분석용 솔루션 사용자를 흡수할 수 있어야 함
 - 기존 데이터 분석용 솔루션의 기능 제공 필요
 - 스케줄러
 - 디자이너
 - 모니터링
- Big Data 기반 개발 작업에 있어서 편의성을 제공해야 함
 - 개발 도구로서 웹 기반 통합 개발 환경 (향후 개발 예정)
 - Pig Script & Runner (Pig Latin Script)
 - Hive Script & Runner (SQL)

- ❑ 프로젝트명 : Flamingo
- ❑ 목표 : 통합 개발 및 관리 도구 for OpenSource Big Data
- ❑ 라이선스 : GPL v3
- ❑ 기능
 - Web Rich Rich User Interface (과제 범위)
 - MapReduce Designer
 - Dashboard(Job Monitoring)
 - HDFS Browser
 - Pig, Hive IDE (계획)
 - Job Scheduler
 - Reusable MapReduce Module
 - System Management & Monitoring
- ❑ 릴리즈 : 1개월에 1회 예정(찾은 릴리즈를 통한 빠른 피드백)

주관/참여기관 역할

ONYCOM(주관)

- 온라인 커뮤니티 구축(웹 사이트, 위키, 빌드서버 등등)
- 프로젝트 관리
- 오프라인 세미나 및 교육자료
- 시스템 구축 및 공동 개발
- 산출물 관리
- 마케팅 개발

Clouline(참여)

- 아키텍처 수립
- 시스템 구축 및 공동 개발
- 산출물 작성
- 시스템 개발
 - MapReduce Designer
 - HDFS 브라우저
 - 모니터링 개발

커뮤니티 활성화 요구조건



커뮤니티 운영 계획

Dashboard > Workflow Engine > ... > Clustered Job Scheduler

Browse Log In Search Confluence

Clustered Job Scheduler

Added by Edward KIM, last edited by Edward KIM on 1월 22, 2011 (view change) show comment

Workflow Engine은 REST Interface 또는 CLI를 이용하여 Workflow를 스케줄링 할 수 있습니다. 다만 확장성으로 고려했을 때 Job Scheduler를 반드시 Clustering 되어야 합니다. Open Flamingo는 Quartz Job Scheduler를 기반으로 동작하며 Quartz는 Clustering을 지원합니다. Quartz Job Scheduler를 사용하기 위해서 우선 pom.xml 파일에 다음과 같이 dependency를 추가합니다.

```
<dependency>
  <groupId>org.quartz-scheduler</groupId>
  <artifactId>quartz</artifactId>
  <version>1.8.4</version>
</dependency>
```

이제 Quartz의 설정 파일인 quartz.properties 파일을 /WEB-INF/quartz.properties로 합니다.

```
=====
# Configure Main Scheduler Properties
=====
```

WIKI

JIRA

Dashboards Projects Issues Create Issue Quick Search

Login - Spring Security 적용

Comment Voters Watch Issue More Actions Reopen Issue Workflow Share Views

Details

Type:	Task	Status:	Closed (View Workflow)
Priority:	Minor	Resolution:	Fixed
Affects Version/s:	None	Fix Version/s:	None
Component/s:	Workflow Web Application		
Labels:	None		

People

Assignee: 원종석

Reporter: 원종석

Vote (0)

Watch (0)

Dates

Due: 02/4월/12

Created: 31/3월/12 2:07 오후

Updated: 06/4월/12 9:45 오전

Resolved: 02/4월/12 3:01 오전

Description

Spring Security 적용

Issue Tracker

Dashboard Source Projects People Reviews

Flamingo-Workflow-Manager

All trunk Select branch or tag

- src/ul>- argoum/
- assembly/
- javado/
- main/ul>- java/org/openflamingo/ul>- controller/
- core/
- datasource/
- domain/
- engine/
- exception/
- hadoop/
- jaxb/
- scheduler/
- service/
- test/
- util/
- resources/
- webapp/

- site/
- test/
- xref/
- build.xml
- pom.xml

Activity Commit Graph Files Users Reports

All Commits Reviews Filter commits

일요일 29 1월

tharenheit 386 changed 1 file root: View Resolver

금요일 27 1월

soryoung 384 changed 1 file root: html test

화요일 24 1월

tharenheit 371 changed 10 files root: JobConf 화면 예제를 만들어봄.

370 changed 1 file root: POM 분리.

월요일 23 1월

tharenheit

Version Control

□ 구성원 : Committer와 Contributor로 구분하여 운영

- Committer는 일정 수준 이상 Contributor로 기여한 개발자에게 권한을 부여

□ 기존 커뮤니티와 협업

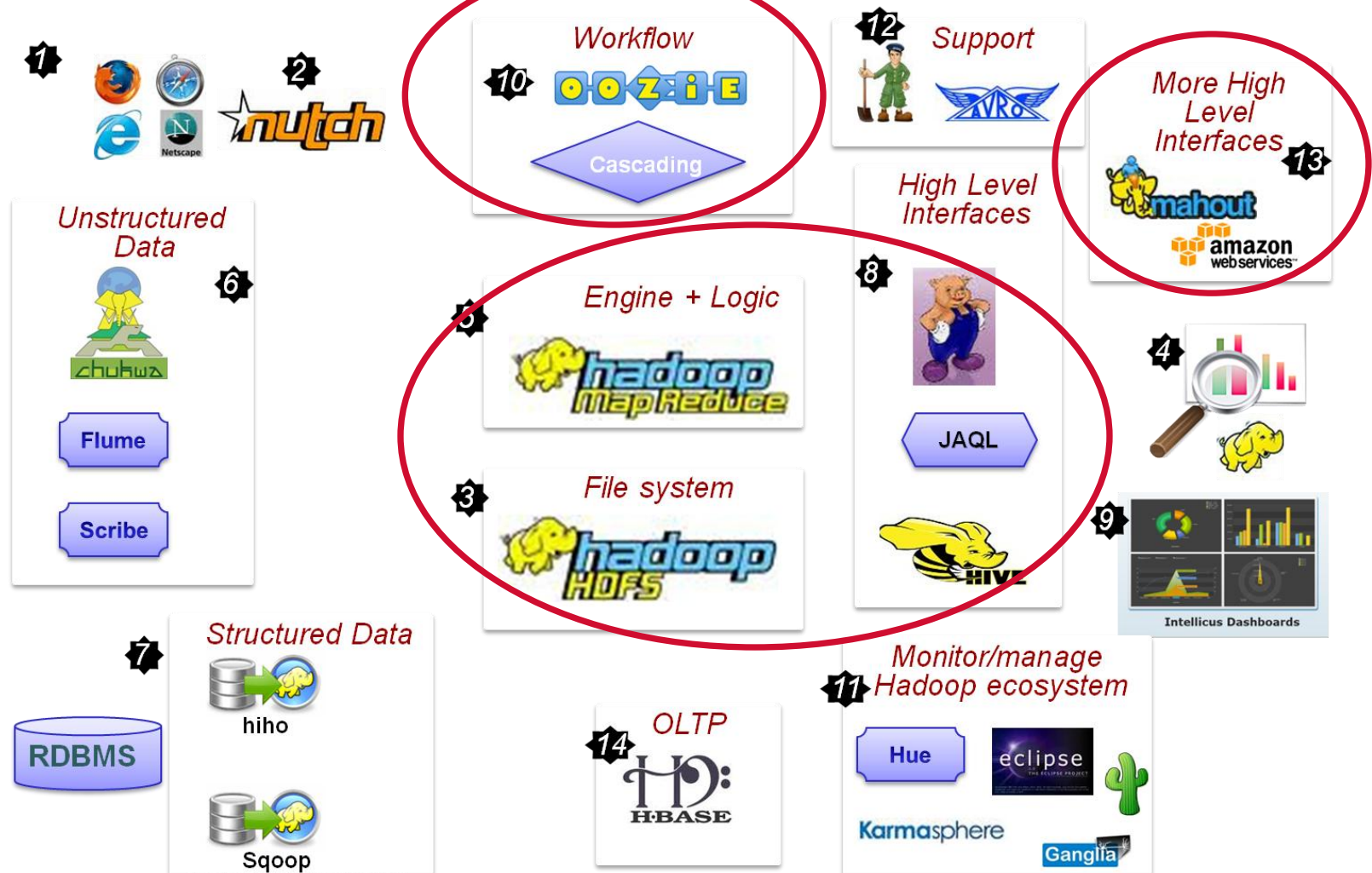
- 클라우드 컴퓨팅 구현 기술 모임과 Hadoop Community와 연계하여 본 과제에 대한 기능 검증 및 테스트 그리고 피드백
- JCO 소속 커뮤니티의 현업 전문가들로 부터 멘토링

□ 커뮤니티의 활성화를 위한 노력

- **현업에서도 쓰일 수 있을만한 기능과 품질을 보장하는 것이 매우 중요**
- 제품 릴리즈는 일정한 주기에 따라서 신속하게 릴리즈(오픈소스의 특징)
- 누구나 접근할 수 있는 WIKI, Issue Tracker, Version Control 웹 화면 제공
- 교육 및 세미나를 통해서 직접 사용하고 시연을 통해 홍보

Hadoop Ecosystem

Hadoop Ecosystem Map



Appendix :: Feature :: HDFS Browser



- Big Data 처리시 가장 많이 개발자들이 수행하는 작업이 파일을 탐색하고 처리하는 작업

The screenshot displays the HDFS Browser application. On the left, a '디렉토리' (Directory) pane shows a tree structure starting from the root '/', listing various folders like 'app', 'button', 'chart', etc. The main area on the right, titled '파일' (File), shows a table of files. The table has columns for 'Company', 'Price', 'Change', '% Change', and 'Last Updated'. The bottom status bar indicates 'HDFS 사용량' (HDFS Usage) as '100M/199M' and the 'Path' as '/'. Navigation buttons like '복사' (Copy), '이동' (Move), '삭제' (Delete), etc., are visible above the table.

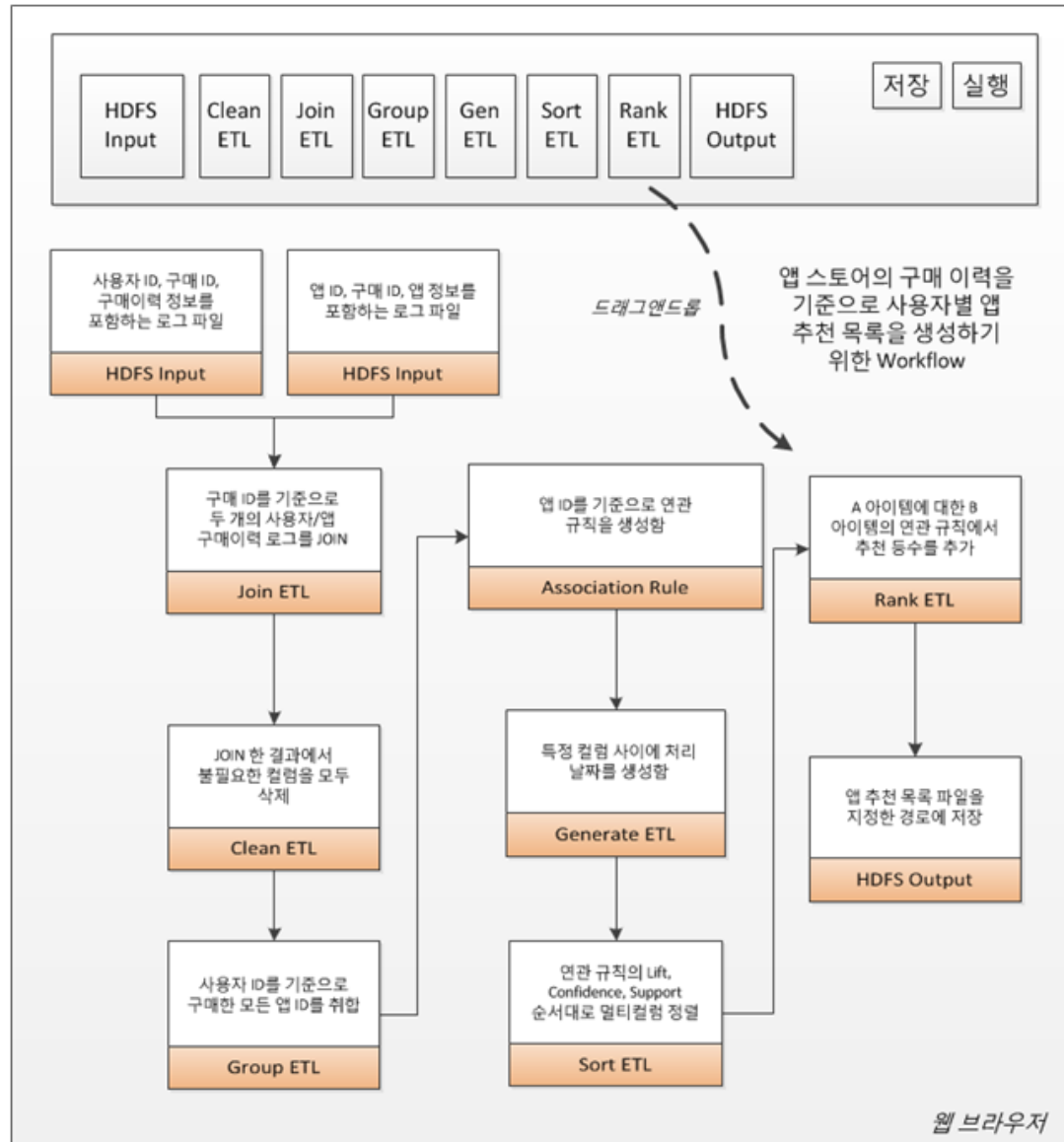
Company	Price	Change	% Change	Last Updated
3m Co	\$71.72	0.02	0.03	09/01/2011
Alcoa Inc	\$29.01	0.42	1.47	09/01/2011
Altria Group Inc	\$83.81	0.28	0.34	09/01/2011
American Express Company	\$52.55	0.01	0.02	09/01/2011
American International Group, I...	\$64.13	0.31	0.49	09/01/2011
AT&T Inc.	\$31.61	-0.48	-1.54	09/01/2011
Boeing Co.	\$75.43	0.53	0.71	09/01/2011
Caterpillar Inc.	\$67.27	0.92	1.39	09/01/2011
Citigroup, Inc.	\$49.37	0.02	0.04	09/01/2011
E.I. du Pont de Nemours and C...	\$40.48	0.51	1.28	09/01/2011
Exxon Mobil Corp	\$68.10	-0.43	-0.64	09/01/2011
General Electric Company	\$34.14	-0.08	-0.23	09/01/2011
General Motors Corporation	\$30.27	1.09	3.74	09/01/2011
Hewlett-Packard Co.	\$36.53	-0.03	-0.08	09/01/2011
Honeywell Intl Inc	\$38.77	0.05	0.13	09/01/2011
Intel Corporation	\$19.88	0.31	1.58	09/01/2011
International Business Machines	\$81.41	0.44	0.54	09/01/2011
Johnson & Johnson	\$64.72	0.06	0.09	09/01/2011
JP Morgan & Chase & Co	\$45.73	0.07	0.15	09/01/2011

Appendix :: Feature :: Dashboard

- MapReduce Designer에서 그려진 플로우를 실행하는 경우 실행 이력을 모니터링하는 기능

Hadoop Job Tracker Summary							
클러스터: <input type="text"/>   정보							
Job Id	Job Name	Map %	Reduce %	Map Total	Map Completed	Reduce Total	Reduce Completed
JOB_20111023_0001	Sample Job	100%	100%	200	50	200	50
JOB_20111023_0002	Sample Job	100%	100%	200	40	200	40
JOB_20111023_0003	Sample Job	100%	100%	200	30	200	30
JOB_20111023_0004	Sample Job	100%	100%	200	20	200	20
JOB_20111023_0005	Sample Job	100%	100%	200	10	200	10
JOB_20111023_0006	Sample Job	100%	100%	200	50	200	50
JOB_20111023_0007	Sample Job	100%	100%	200	40	200	40
JOB_20111023_0008	Sample Job	100%	100%	200	30	200	30
JOB_20111023_0009	Sample Job	100%	100%	200	20	200	20
JOB_20111023_0010	Sample Job	100%	100%	200	10	200	10
JOB_20111023_0011	Sample Job	100%	100%	200	40	200	40
JOB_20111023_0012	Sample Job	100%	100%	200	30	200	30
JOB_20111023_0013	Sample Job	100%	100%	200	20	200	20

Appendix :: Feature :: MapReduce Designer



Appendix :: 개발 목표

□ 기능적 측면

- MapReduce의 실행 이력을 모니터링하는 Hadoop Job 모니터링
- Apache Hadoop Cluster의 분산 파일 시스템을 제어할 수 있는 윈도 익스플로러 스타일의 HDFS 브라우저
- 웹 기반으로 MapReduce의 처리 흐름을 설계 및 실행할 수 있는 MapReduce Designer
- MapReduce Designer 통합을 위한 MapReduce 기반 코드 표준 정립

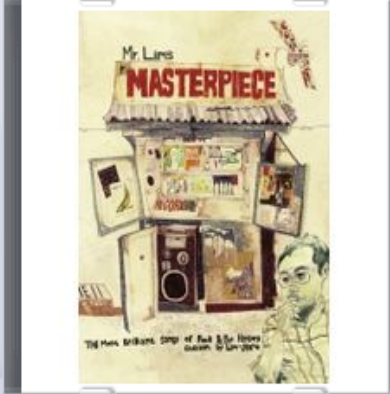
□ 환경적 측면

- 위키, 이슈 트래커 구축
- Mavenization & Maven Repository 구축
- Continuous Integration 구축
- 설계, 설치 문서

Appendix :: 멜론의 연관곡 추천

Melon은 연관 규칙을 통해 곡을 추천한다.

🔴 임진모가 추천하는 '명작 중의 명작'



아티스트 Various Artists
발매일 2009. 02. 17
발매사 Universal Music
기획사 Universal Music
평점 ★★★★★ 4.4 / 참여 56명 [점수주기](#)
앨범평 7개 [앨범평쓰기](#)

[앨범듣기](#) [앨범스크랩](#) [CD구매](#)

활동장르

Pop, Rock, Blues, Rap /
Hip-hop, R&B / Soul

스타일

Soul, Blues Rock, Hard
Rock, Soft Pop/Rock,
Progressive / Art Rock

Album Info 앨범 소개

음악평론가 임진모가 추천하는 [명작 중의 명작]

'임진모가 추천하는 명작 시리즈'에서 엄선된 26곡의 불멸의 팝 명곡! 국내 최고의 음악 평론가가 강력 추천하는 곡들만 모았다!
에릭 클랩튼, 엘튼 존, 스...

MORE >

이 앨범의 연관곡 추천

- ☐ Hopeless KT Tun...
- ☐ Old And Wise Alan Par...
- ☐ Again And Ag...Jewel
- ☐ The Servant Cocosu...
- ☐ Born To Try Delta Go...
- ☐ Easy Commo...
- ☐ Misread Kings Of...
- ☐ Lea Toto
- ☐ Babe I'm Gon...Led Zep...
- ☐ Scars (Acous...Papa Ro...

[듣기](#) [다운](#) [담기](#)