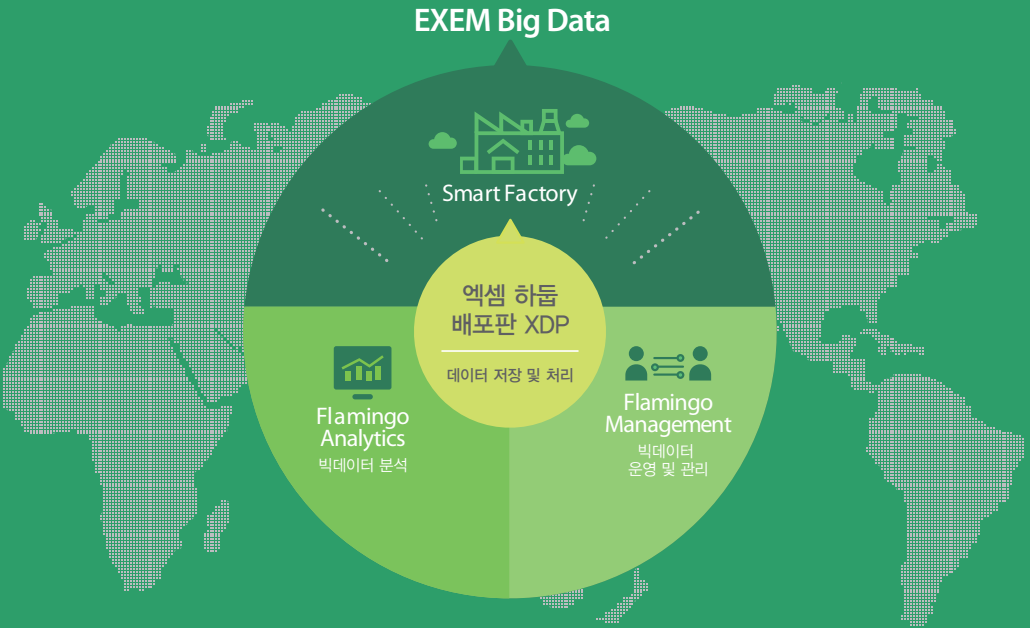


누구나 **기술의 장벽 없이** 세상의 모든
빅데이터를 다룰 수 있도록



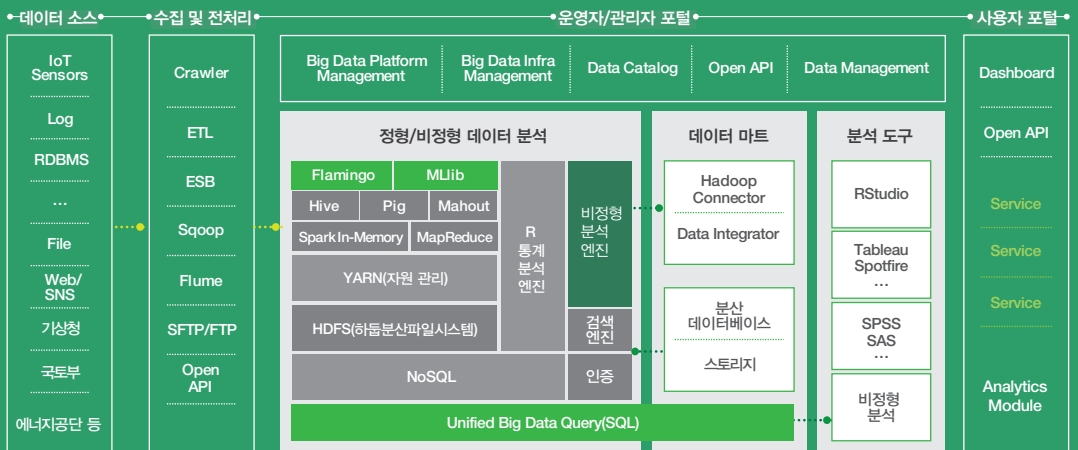
Hadoop 기반 빅데이터 운영·관리·분석 플랫폼 Flamingo Analytics | Management

Flamingo는 Hadoop EcoSystem과 결합하여 빅데이터의 수집·저장·운영·분석에 이르는 생명주기를 효과적으로 관리할 수 있도록 지원하는 플랫폼 소프트웨어입니다.

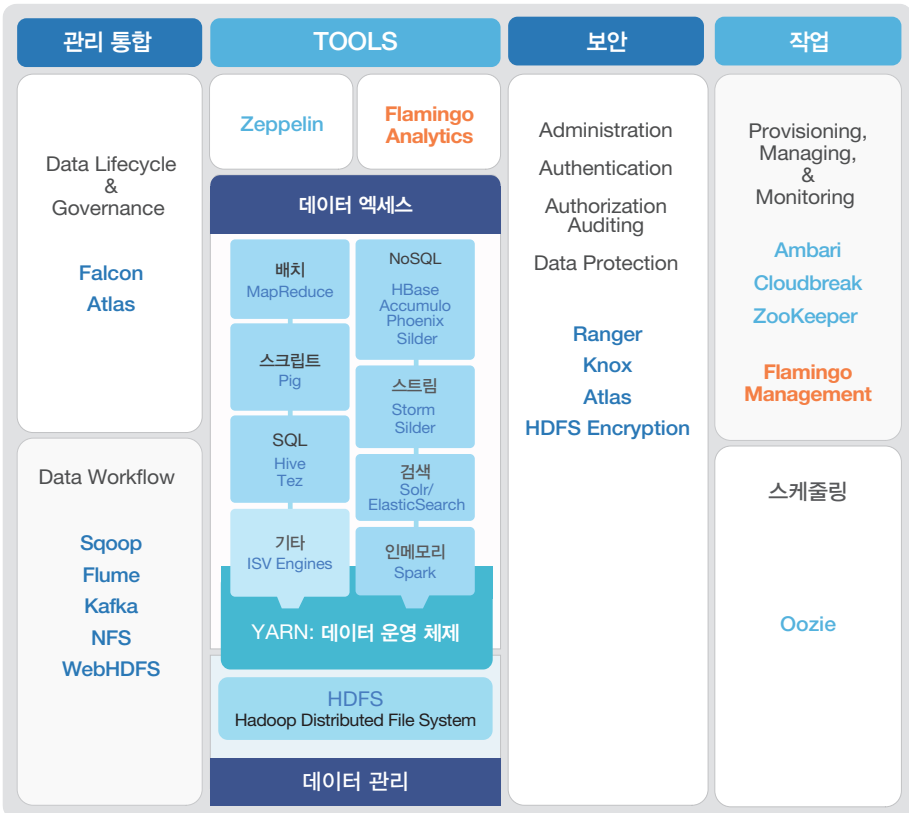


엑셈 빅데이터 아키텍처

엑셈은 다년간의 빅데이터 사업 경험과 축적된 기술력을 바탕으로 고객에게 최적의 빅데이터 아키텍처를 제공합니다.



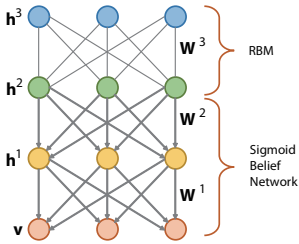
빅데이터 구축을 위한 엑셈의 하둡 배포판인 XDP는 Hortonworks HDP와 Flamingo를 최적화한 하둡 플랫폼입니다. Flamingo를 개발하는 오픈소스 커미티의 전문 기술지원이 가능한 배포판입니다. 사용자는 XDP를 통해 빅데이터 플랫폼의 관리에 있어 수집·저장·운영·분석 등의 전 생명주기를 관리할 수 있습니다.



이미지 인식, 음성 검색, 전자상거래 사기탐지, 추천 시스템, 회귀분석 등 딥러닝을 손쉽게 사용 가능하도록 플랫폼에 알고리즘을 탑재하여 분석할 수 있는 기능을 제공합니다.

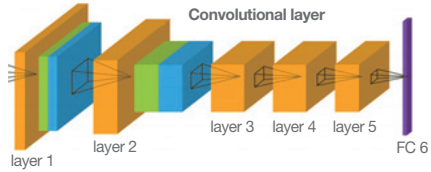
· 심층 신뢰 신경망(DBN)

영상처리, 음성인식, 이미지 인식, 자연어 처리 등에 활용



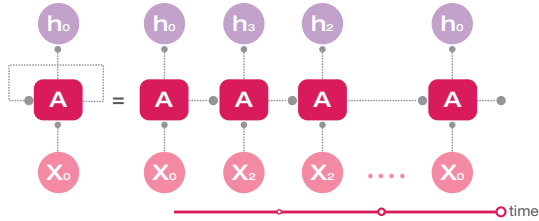
· 컨볼루션 신경망(CNN)

변화 객체 탐지, 객체 정보 추출, 지형 정보 추출에 활용



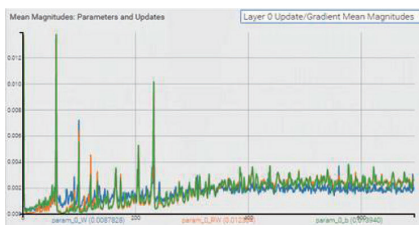
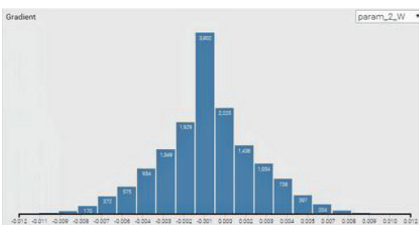
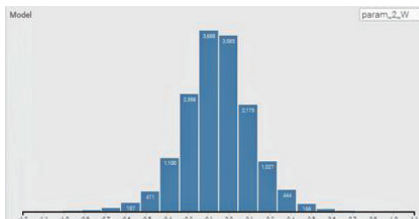
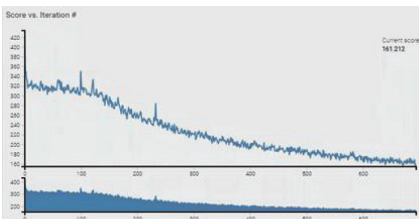
· 순환 신경망(RNN)

시간 변화에 따른 특이 객체 탐지 및 다음에 일어날 상태 또는 이벤트를 예측하는데 활용



딥러닝 알고리즘 최적화

딥러닝의 핵심은 학습 결과의 수준을 최대한 끌어올리는 최적화 과정입니다. 학습하는 동안 시각화를 통해 최적화된 상황을 모니터링합니다.

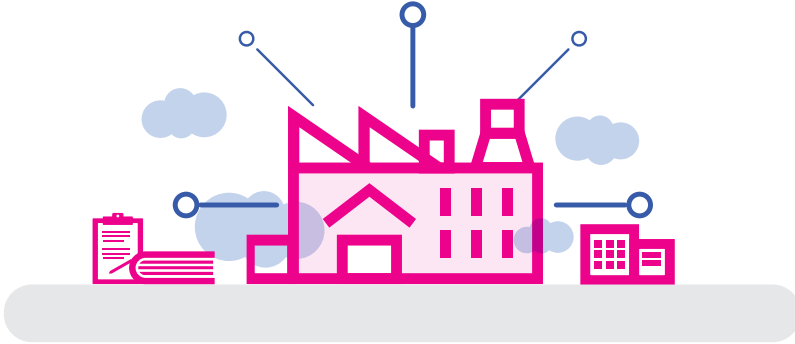


Smart Factory

제조 현장의 최적 의사결정 지원

엑셀은 Smart Factory를 통해 MES 및 공장설비에서 수집한 데이터를 클라우드 및 서버에 저장하고, 이를 분석하여 제조 현장에서 최적의 의사결정을 지원하는 플랫폼을 제공합니다.

빅데이터 수집 및 분석 | 공정 데이터의 시계열 처리 | 의사결정 시나리오 및 서비스 | OPEN API를 이용하여 MES와 연동



아키텍처

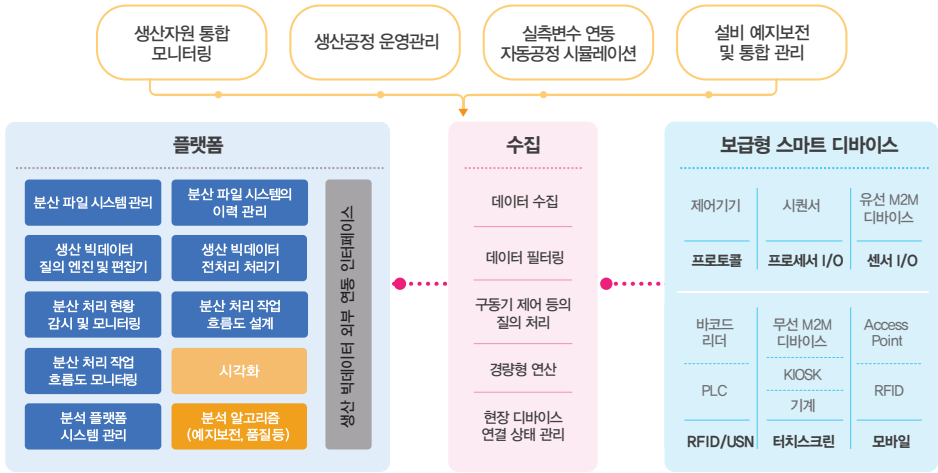
Smart Factory의 기본 아키텍처는 Flamingo Analytics를 중심으로 한 데이터 수집 및 분석 레이어와 이를 활용한 의사결정 시나리오 및 서비스 모델 레이어로 구성됩니다.



구축 사례

한국생산기술연구원 Smart Factory

한국생산기술연구원은 다이캐스팅 분야에 Flamingo를 이용하여 공정 데이터를 바탕으로 불량 원인과 센서 간의 상관관계를 분석하였습니다.

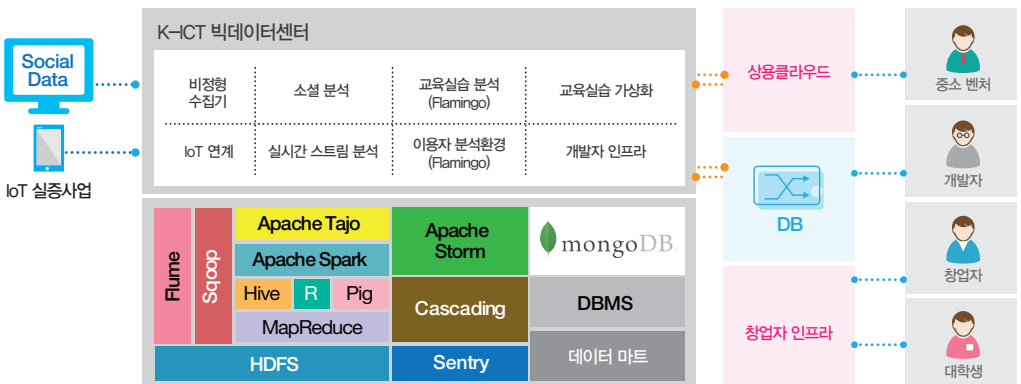


! 생산 빅데이터 분석 및 운영 환경 !



한국정보화진흥원 K-ICT 빅데이터 센터

한국정보화진흥원의 K-ICT 빅데이터 센터는 대규모 인프라 및 분석환경을 제공하며 이를 위해 Flamingo를 이용하고 있습니다. 특히 가상화를 통해, 다양한 기관에서 가상머신을 이용하여 실습할 수 있는 환경을 제공합니다.



! K-ICT 빅데이터 센터 분석 환경 !



엑셈 빅데이터 사업

엑셈은 전문화된 빅데이터 기술과 경험을 바탕으로 빅데이터 구축·운영·분석·교육에 이르는 다양한 서비스 및 제품을 보유하고 있습니다.

- 빅데이터 분석 컨설팅 : 고객의 인사이트 및 차별화된 서비스를 위한 데이터 분석 컨설팅
- 빅데이터 플랫폼 S/W 제공 : 빅데이터 수집·저장·운영·관리·분석 등을 위한 Flamingo, XDP S/W 제공
- 빅데이터 플랫폼 운영 및 기술지원 : 하둡 배포판 및 플랫폼 운영을 위한 전문 기술 지원
- 빅데이터 아카데미 : 하둡 및 빅데이터 기술에 대한 초보자 및 전문가를 위한 4일 교육 프로그램

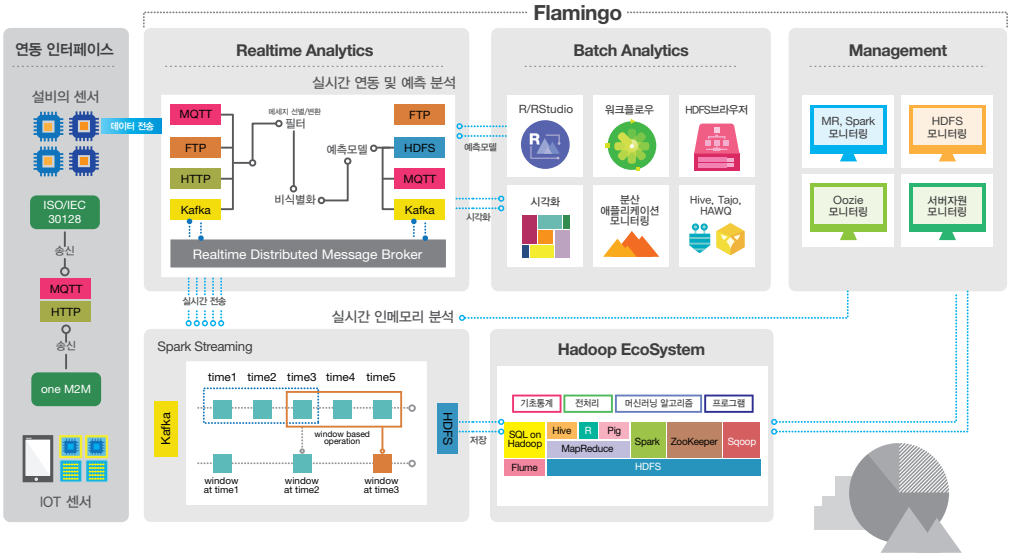
사업 연혁



- 삼성디스플레이 하둡 유지보수 및 데이터 처리 사업
- CITIBANK EAP 빅데이터 플랫폼 구축
- 한국정보화진흥원 K-ICT 빅데이터 센터 교육 사업
- 한국정보화진흥원 K-ICT 빅데이터 센터 분석 인프라 고도화 사업
- 도로공사 C-ITS 빅데이터 센터 구축 사업
- 식품의약품안전처 데이터마이닝 시스템 구축 사업
- 인공지능 Exobrain WiseKB 빅데이터 이해 기반 자가학습형 지식베이스 및 추론기술 개발
- 국토부 공간빅데이터 사업
- 삼성전자 빅데이터 센터 BDP 1.0(Flamingo) 플랫폼 개발 및 구축 사업
- 삼성전자 빅데이터 센터 BDP 2.0 고도화 사업
- ETRI 테라바이트급 NoSQL 기반 동영상 유사 식별 시스템 개발 사업
- 국방부 빅데이터 플랫폼 개발 및 구축 사업
- 한국정보화진흥원 빅데이터 전략센터 고도화 사업
- 삼성반도체 YMS 빅데이터 공정 분석 플랫폼 구축 사업
- 삼성SDI 빅데이터 기반 품질 예측 인프라 구축 및 유지보수 사업
- 삼성전기 빅데이터 기반 품질 예측 인프라 구축 사업
- KT 자사경영정보관리 시스템 구축 사업

빅데이터의 실시간·배치 분석, 운영을 위한 Flamingo

Flamingo는 빅데이터 기술을 기반으로 데이터 운영·관리·분석·IoT에 이르는, 데이터를 활용할 수 있는 모든 분야에서 사용 가능한 통합 플랫폼입니다. IoT 센서 데이터는 배치·분석용 예측 모델에 실시간으로 빠르게 생성되는 데이터를 적용하여 예측합니다.

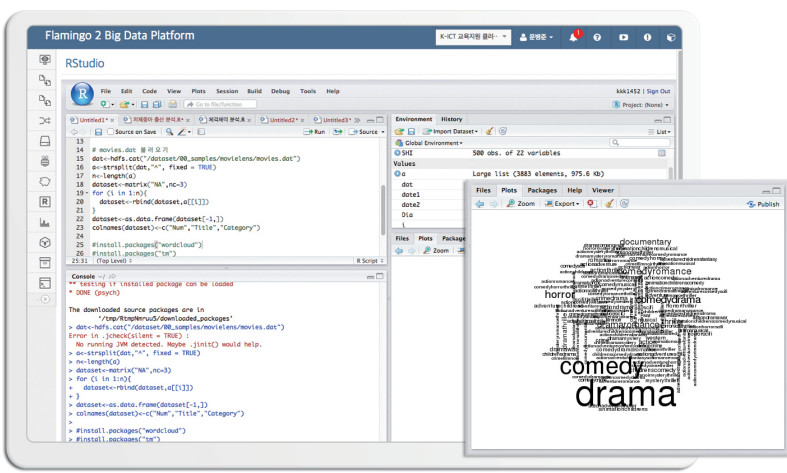


RStudio Server

R 통계 분석을 위한 통합 환경



R은 소규모 데이터를 이용하여 분석하고 모델을 만드는 최고의 도구입니다. Flamingo는 RStudio Server와 연계하여 Flamingo의 단일 계정으로 R 분석을 수행하고 이를 Hadoop의 병렬 처리로 쉽게 이관할 수 있는 환경을 제공합니다.

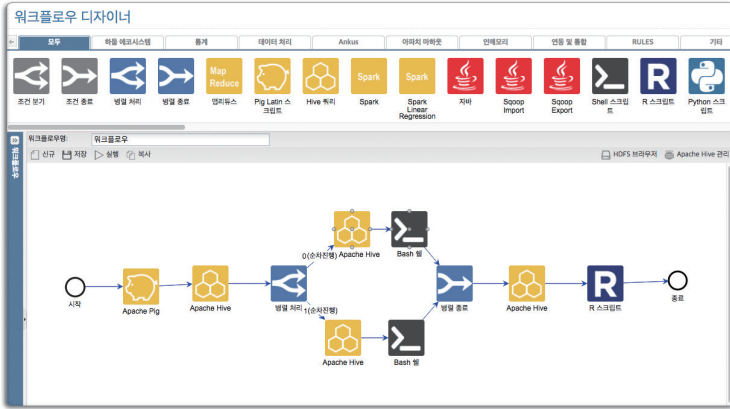


Workflow Designer

데이터 분석을 위한 워크플로우 설계



Hadoop EcoSystem의 설치 후 병렬 처리 및 분석을 위해 MapReduce, Hive, R, Pig, Spark 등의 다양한 서비스를 활용합니다. 데이터 처리 및 분석은 여러 단계의 순서로 동작합니다. Flamingo는 이러한 데이터 분석 과정을 시각화하고 초보자도 손쉽게 데이터 처리와 분석을 수행할 수 있도록 통합된 사용자 인터페이스를 제공합니다.



HDFS Browser

분산 파일 시스템 관리



HDFS(하둡 분산 파일 시스템)에는 수많은 양의 로그 데이터가 저장되어 있습니다. 운영자와 개발자 관점에서 디렉토리와 파일 관리 작업은 분석 업무 만큼이나 중요하지만 번거롭다는 문제점이 존재합니다. Flamingo의 HDFS Browser는 이러한 번거로움을 해소하는 최고의 편의성과 기능성을 제공합니다.

The screenshot shows the HDFS Browser interface. On the left, there's a navigation pane with 'HDFS 디렉토리' and 'HDFS 파일'. The main area displays a table of files with columns for name, size, and modification time. A circular gauge indicates '남아 있는 양' (Remaining amount). On the right, a '파일 내용 보기' (View file content) window is open, showing a list of file paths and their corresponding content. The content is displayed in a monospaced font, showing a list of file paths and their corresponding content.

업로드
HDFS Browser를 통해 대용량 파일을 직접 업로드 가능

내용보기
Hadoop HDFS에 있는 파일 내용을 바로 확인 가능

YARN, MapReduce Monitoring

분산 애플리케이션 모니터링



Hadoop은 클러스터를 구성하는 서버의 자원을 효율적으로 관리하기 위해 YARN을 사용합니다. YARN은 Hadoop 서버에 분석 작업을 실행할 때 어떻게 자원(vCore, RAM)을 분배할지 결정합니다. Hive, Spark, MapReduce 등의 작업을 실행하면 YARN이 자원을 배정하고, Flamingo는 이를 단일 웹 환경에서 모니터링 가능하도록 지원합니다.

YARN 애플리케이션

YARN 기반으로 작성된 모든 애플리케이션(MR, Spark, Tez 등)에 대한 모니터링

애플리케이션 요약
애플리케이션의 요약 정보를 제공하여 실행 내역을 빠르게 확인

애플리케이션 로그
상세 애플리케이션의 실행 로그를 모니터링에서 바로 확인

Apache Hive Management

Hadoop Data Warehouse 관리 및 개발



Apache Hive는 Hadoop EcoSystem에서 가장 많이 사용되는 Query Engine입니다. Hive의 메타정보를 보관하고 있는 Metastore는 데이터베이스와 테이블 정보를 저장하고 있으며, 개발자는 커맨드 라인으로 Query를 실행하여 테이블과 데이터베이스를 관리합니다. Flamingo는 단일 웹 기반 사용자 인터페이스에서 이 모든 것을 수행할 수 있는 통합 환경을 제공합니다.

Apache Hive 관리

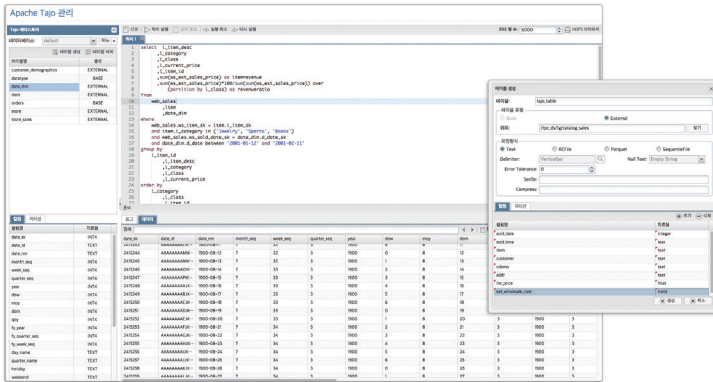
Hive의 DB, Table, Partition 관리를 커맨드에서 스크립트로 작성하지 않고 이를 통해 관리

웹 편집기에서 작성한 쿼리를 바로 수행하며, 수행 결과를 하단의 그리드를 통해 확인 가능

SQL on Hadoop

Hadoop 기반 고성능 DW 엔진의 관리 및 개발

SQL on Hadoop은 Hive 대비 더욱 빠른 고성능 Data Warehouse 엔진입니다. 메타정보를 관리하고 쿼리를 실행하는 것은 중요한 업무이지만, 통합 환경을 제공하는 도구는 이제껏 없었습니다. Flamingo는 대표적인 SQL on Hadoop 엔진인 Apache/Pivotal HAWQ, Apache Tajo를 지원합니다. Metastore를 통해 데이터 베이스와 테이블 관리 및 쿼리 실행 작업을 단일 웹 기반 사용자 인터페이스에서 가능하도록 합니다.

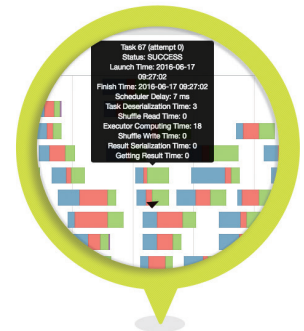
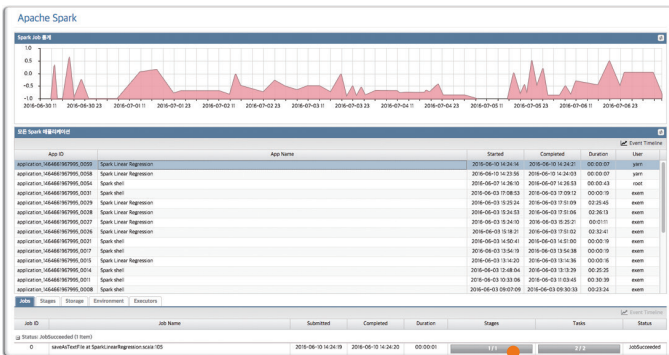


Spark Monitoring

In-Memory 분석 모니터링



Spark는 MapReduce를 대체할 차세대 분석 엔진입니다. Spark로 작성된 분석 애플리케이션은 Hadoop Cluster를 구성하는 서버의 메모리에 데이터를 적재하고 이를 병렬 처리합니다. Flamingo는 Spark 분석 애플리케이션이 메모리상에서 어떠한 추이로 동작하는지 YARN과 Spark Master 정보를 토대로 시각화합니다.



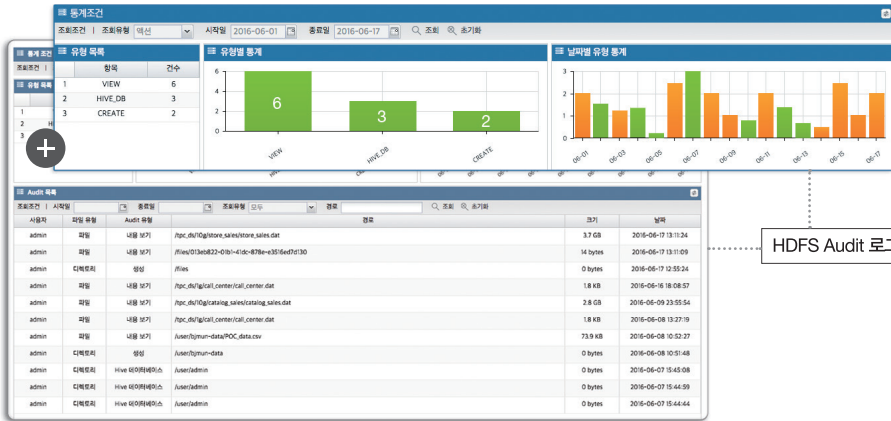
Event Timeline의 시각화를 통해 Spark 애플리케이션이 수행된 정보를 한눈에 파악 가능

HDFS Audit

분산 파일 시스템 관리



HDFS는 대량의 데이터가 저장되고 다수의 사용자들이 작업을 하는 중요한 공간입니다. 해당 공간에 있는 데이터를 다운로드하거나 삭제할 경우 이에 대한 사용자 행위는 기록되어야 합니다. Flamingo는 HDFS Browser 상에서 일어나는 모든 행위를 사용자와 행위 별로 기록하고 이를 시각화합니다.



HDFS Audit 로그

Web CLI

개발 및 운영의 편의를 위한 작업 환경



Hadoop Cluster는 개발자 및 운영자 PC와는 별도로 구축된 IP 대역을 사용하는 것이 일반적입니다. 또한 대용량 데이터를 보관하고 있으므로 보안에 많은 주의를 기울여야 하는 서버이기도 합니다. 여러 단계를 거쳐 SSH로 접근하는 시스템의 경우 커맨드 작업에 있어서 접근의 어려움이 다수 발생합니다. Flamingo는 웹브라우저 상에서 제한된 서버에 접근하여 작업할 수 있도록 지원합니다.

