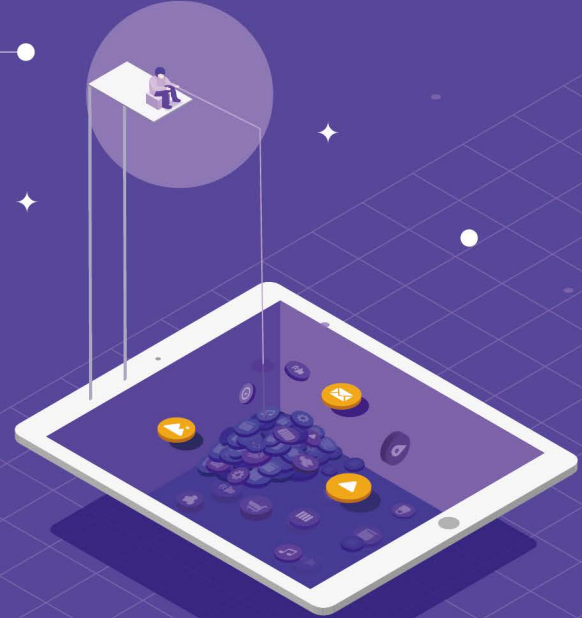


flamingo for Hadoop

Indispensable for Hadoop Management



exem은 대한민국 기업정보시스템의 안정적인 운용과 발전을 견인하였으며 데이터베이스를 비롯한 전 구간의 시스템에서 빅데이터, 인공지능 분야까지 대한민국을 대표하는 글로벌 IT 성능관리 · 빅데이터 · 인공지능 전문기업으로 성장하였습니다.



하둡 모니터링, 워크플로우, 데이터분석 지원을 위한 빅데이터 통합 관리 솔루션

Flamingo는 빅데이터 플랫폼인 하둡 클러스터의 통합 관리 솔루션으로, 하둡과 에코 시스템의 실시간 서비스 감시 및 주요 성능 지표의 수집, 진단 및 모니터링, 데이터 처리를 위한 워크플로우 작성, 데이터 분석 지원까지 이르는 하둡 시스템의 가용성 및 성능의 관리를 효율적으로 수행할 수 있도록 지원합니다.



Realtime Monitoring

Hadoop EcoSystem에 최적화된 다양한 성능지표의 실시간 감시

- Hadoop Core Server
- Apache Spark
- Apache Hive
- HDFS
- Apache Oozie
- Zookeeper
- HBase
- Cluster Servers



Workflow

작성하기 어려운 워크플로우를 간단히 작성하고 손쉽게 테스트할 수 있는 환경 지원

- Workflow (Designer & Monitoring)
- Apache Oozie workflow (Designer & Monitoring)

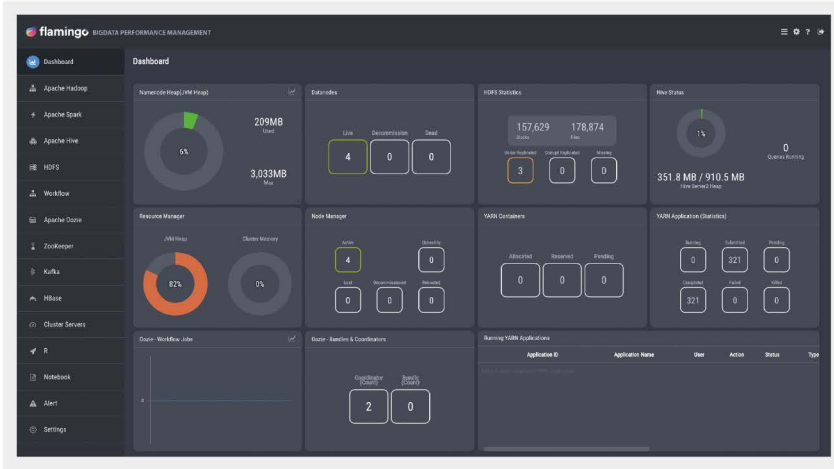


Data Analysis

다양한 빅데이터 환경에서 필요한 정보 도출을 위한 분석 환경 지원

- Notebook (R/Python 등)
- Hive, HBase Editor
- R-Studio 연계 지원

Dashboard



Dashboard View를 통해 Hadoop Cluster의 핵심 지표(Resource Manager, Namenode, HDFS, Hive, YARN, Node Manager, Oozie 등)를 쉽고 직관적으로 파악할 수 있습니다.

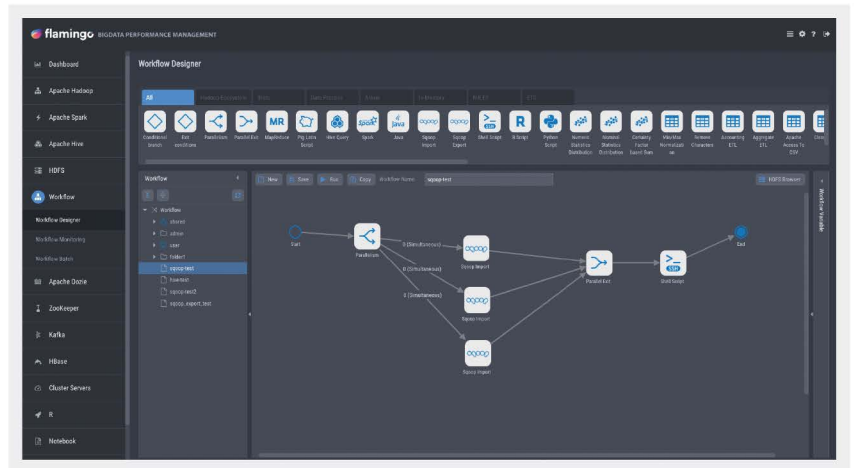
- 10초 주기 실시간 모니터링 정보 수집(Polling이 아닌 Pushing 데이터 수집을 통한 정확성 확보 및 누락 데이터 방지)
- Namenode 메모리, Datanode/HDFS 상태
- Hive Server 상태
- Resource Manager / Node Manager 상태
- YARN Container / YARN Application 상태
- Oozie Workflow / Coordinator 상태

Workflow

Workflow Designer View

워크플로우 디자이너는 GUI 기반으로 손쉬운 드래그 앤 드롭 작업으로 워크플로우를 작성하고 테스트까지 진행 가능한 기능을 제공합니다.

- 드래그 앤 드롭 지원
- 분석 워크플로우를 UI 기반으로 빠르게 드로잉 및 실행
- 다양한 알고리즘 (머신러닝 알고리즘, ETL, 통계 등) 및 모듈 (Java, Spark, Python 등) 탑재를 통해 빠르게 Flow를 구성하여 실행 가능

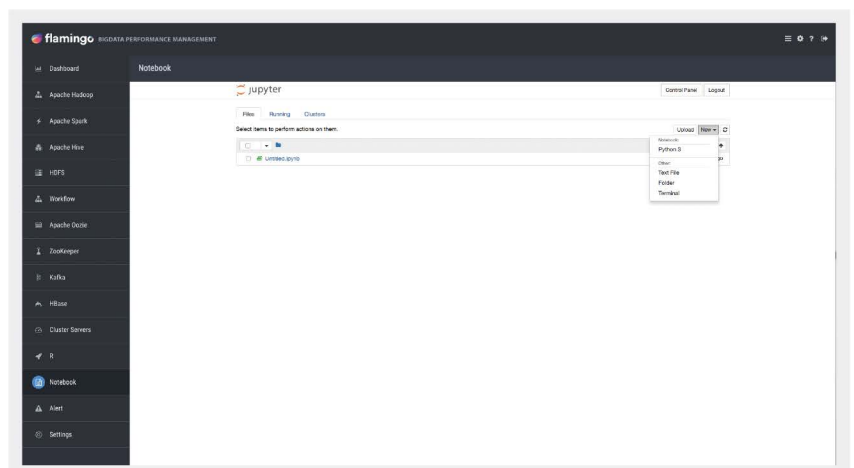


Data Analysis

Notebook View

Flamingo는 멀티 사용자에게 대한 노트북 스타일의 분석 도구를 제공함으로써 Python, Scala, Spark, R 등을 이용하여 빠르게 빅데이터 분석 작업이 가능합니다.

- 멀티 사용자에게 대한 분석 환경 지원
- Python 2/3, R, Scala, Spark Kernel 지원을 통해 다양한 언어 동시분석
- 노트북 스타일의 분석 화면 제공
- R-studio와 연동한 데이터분석 환경 지원

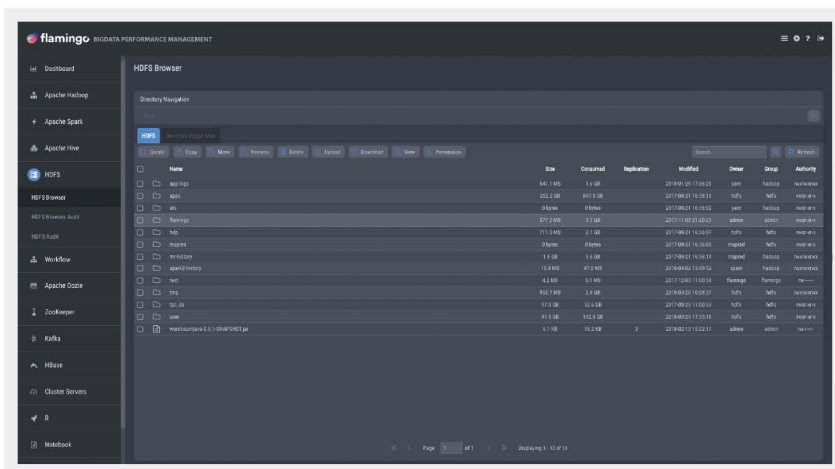


Realtime Monitoring

HDFS Browser

HDFS 브라우저는 HDFS 상의 디렉토리 및 파일을 관리하는 Flamingo의 핵심 기능입니다. GUI 환경 상에서 손쉽게 파일을 업로드하고 다운로드할 수 있으며, 수십만 개의 파일도 거뜬히 표시합니다.

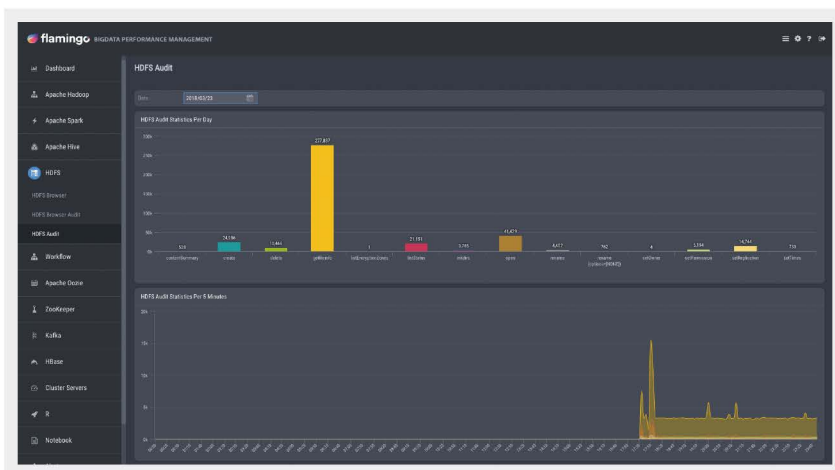
- HDFS의 파일 및 디렉토리 관리
- 하나의 디렉토리에 수십만 개의 파일이 있더라도 UI 문제없이 페이지
- 디렉토리 및 파일의 권한 관리
- 파일 내용 보기 및 파일 업로드/다운로드
- 문자열 기반 디렉토리 경로 입력 시 해당 디렉토리로 이동
- 디렉토리 별 사용량 표시법



HDFS Filesystem Audit

Flamingo는 Namenode에서 발생하는 모든 이벤트를 수집하여 HDFS 상에서 발생하는 이벤트를 모니터링합니다.

- HDFS 상의 모든 이벤트 수집 후 통계 생성
- 시간대별 발생 이벤트 추이 그래프



Spark View

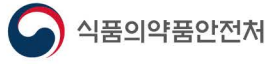
Spark는 In-Memory 분석 프레임워크로서 Hadoop EcoSystem의 핵심 서비스가 되어가고 있습니다. Flamingo에서는 Spark Job의 실행 이력부터 모니터링 정보를 상세하게 정보를 표시하는 모니터링을 제공합니다.

- Spark Job의 일별 통계/실행 이력
- Spark Job의 Job ID 별 실행 이력 및 Event Timeline
- Spark Job의 Stage 별 실행 상태
- Spark Job의 Storage 상태/Executor의 동작 상태 및 Timeline
- Spark의 EcoSystem 지원 (Hive, Hbase, Kafka, Oozie, Zookeeper)

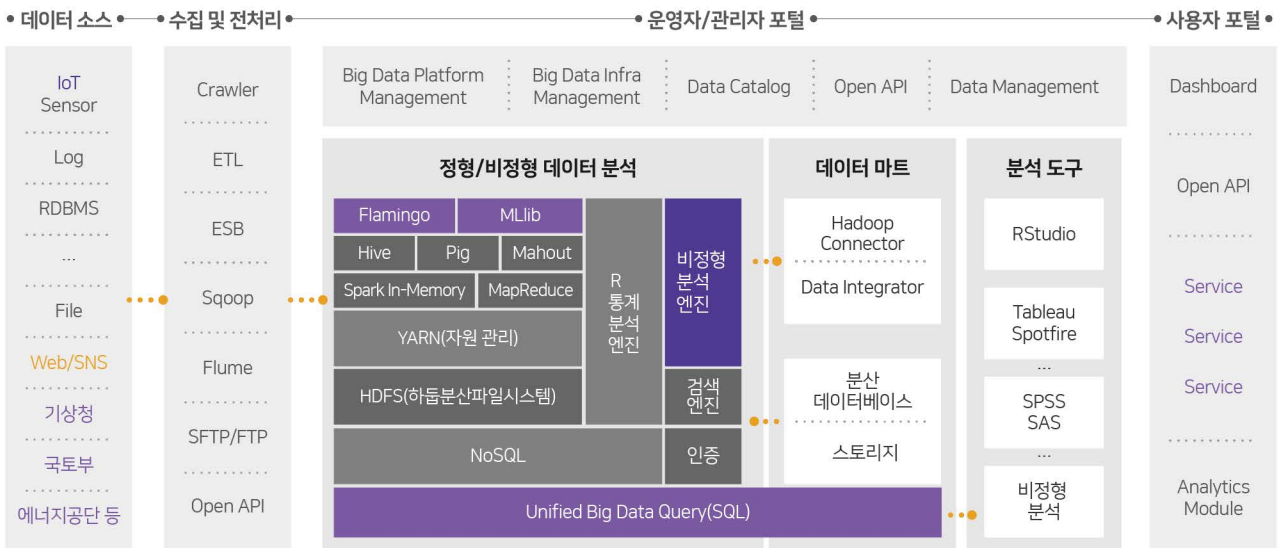
Application ID	Application Name	Started	Completed	Drivers	User
local-10226887982	PySparkShell	2018-04-02 11:27:30	2018-04-02 11:28:59	0/0/0/38/384	flamingo
local-10226888154	PySparkShell	2018-04-02 11:28:30	2018-04-02 11:29:15	0/0/0/21/364	flamingo
application_151788494181_0	sparkof-sparkching	2018-04-02 11:21:54	2018-04-02 11:27:04	0/0/4/2/282	flamingo
application_151788494181_0	sparkof-sparkching	2018-04-02 11:21:54	2018-04-02 11:27:04	0/0/0/0/173	flamingo
application_151788494181_0	sparkof-sparkching	2018-04-02 11:21:54	2018-04-02 10:56:34	0/0/0/0/168	flamingo
local-10226888155	PySparkShell	2018-04-02 05:27:13	2018-04-02 05:27:52	0/0/0/0/115	flamingo

Job ID	Job Name	Submitted	Completed	Duration	Stages	Tasks	Status
1	sharding at hbase-hadoop-connection-test-2	2018-04-02 11:27:13	2018-04-02 11:28:18	00:02:23.75	3/2	184/1768	Job finished
6	sparkof-sparkching	2018-04-02 11:27:04	2018-04-02 11:27:18	00:00:05.08	1/1	1/1	Job finished

Our Customers



Architecture



한국전력(KEPCO)의 'KEPCO 빅데이터 통합 플랫폼 구축사업' 사례

- 수행기간 : 2017.01 ~ 2018.01
- 수행사 : 엑셈, 한전KDN 컨소시엄
- 처리 데이터 : 원시데이터, 가동데이터, 외부 공개 데이터 및 소셜 데이터
- 처리 데이터량 : 2017년 12월 기준 2천 500억 건, 120 TB

