

Flamingo

Big Data Performance Management

 Product Documentation

It's the Best Big Data Performance Management Solution.
Maximize Your Hadoop Cluster with Flamingo. Monitoring, Analyzing, and Visualizing.

CFlamingo. Product DocumentationONTENTS

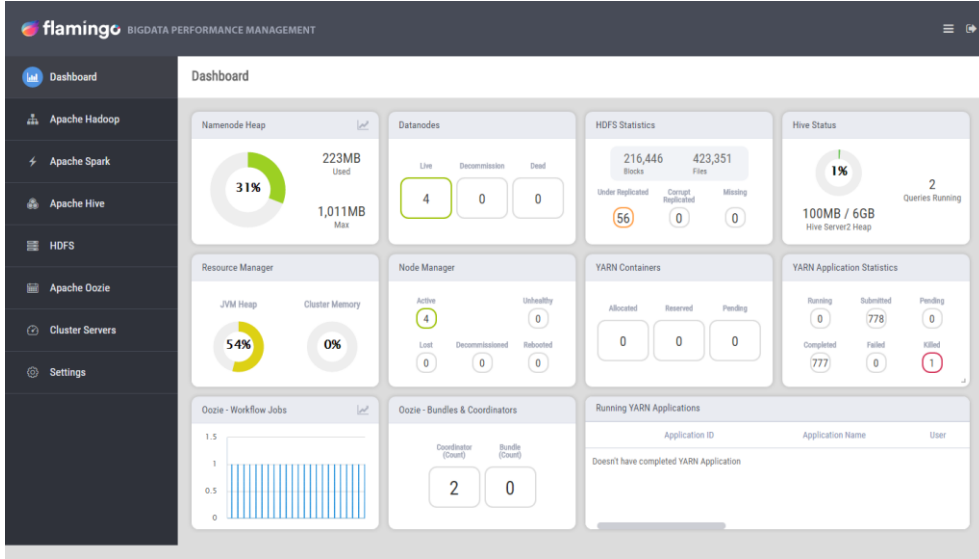
01 제품 개요

Overview

Architecture

Feature

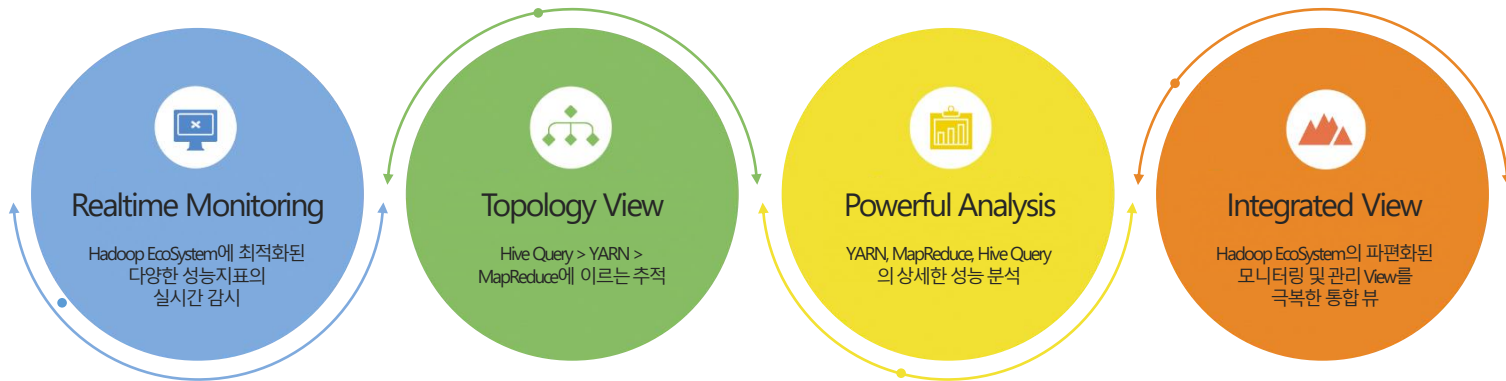
02 주요 기능



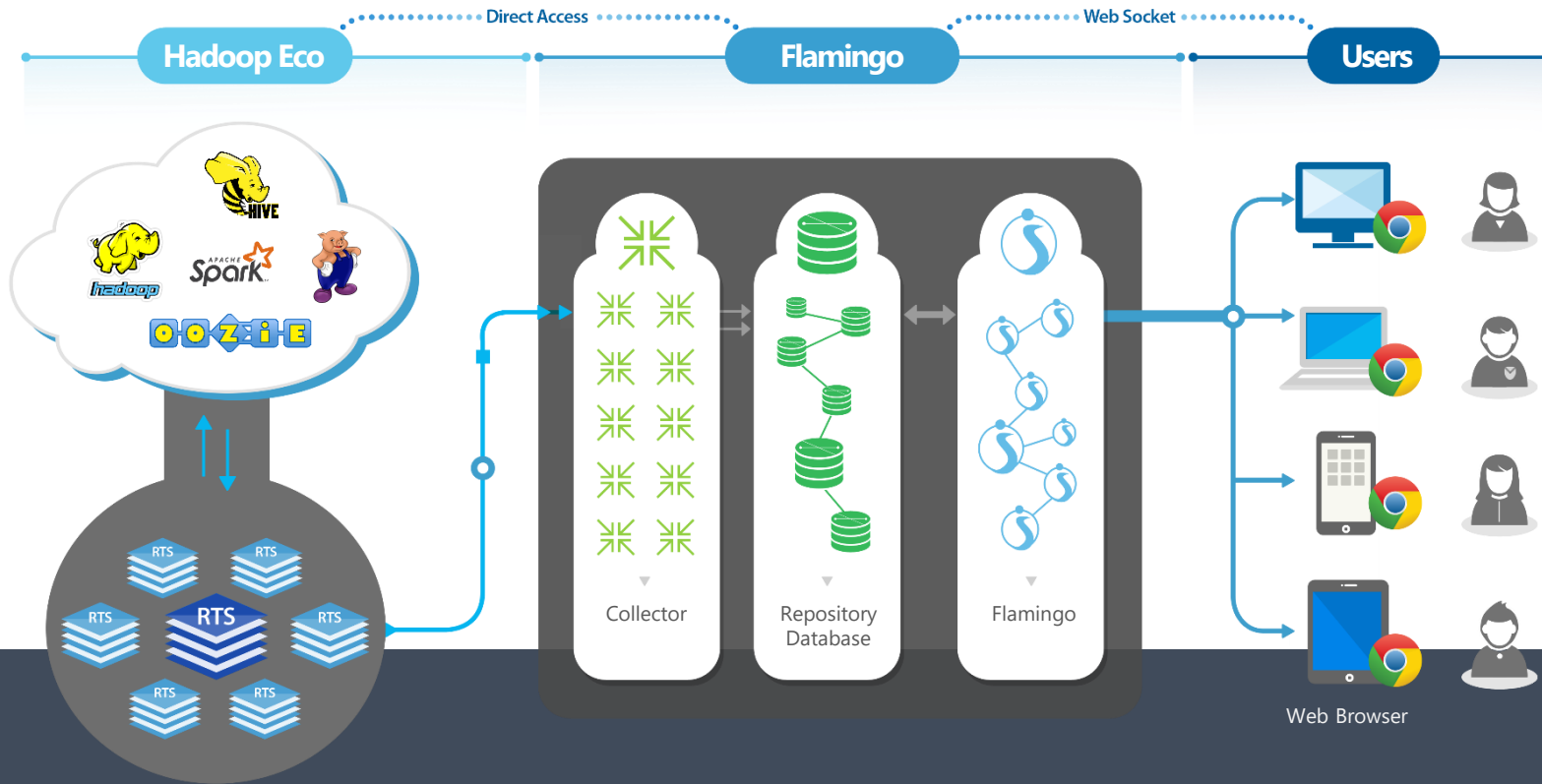
보기 좋게, 쉽게, 빠르게 새로워진 Flamingo

- 쉽고 강력한 UX로 사용자의 편의를 가장 먼저 생각합니다.
- 통합적인 관점에서 실시간으로 모니터링 합니다.
- 다양한 분석 애플리케이션의 성능 데이터를 통해 가장 중요하고 상세한 정보를 제공합니다.

Flamingo는 Big Data 성능 관리를 위한 전문 소프트웨어 솔루션으로 Hadoop EcoSystem의 가용성 및 성능을 효율적으로 관리합니다. Pro-Active한 모니터링 기법을 제공함으로써 문제 상황을 빠르게 인지하고, MapReduce, Hive, Spark 등으로 작성한 분석 애플리케이션의 상세한 정보를 확인하고 분석할 수 있도록 합니다.



“Hadoop EcoSystem의 성능데이터를 최소한의 부하로 수집하여 모니터링 및 저장”



“Apache Hadoop의 표준에 따라 개발되어 다양한 Hadoop 배포판 지원”

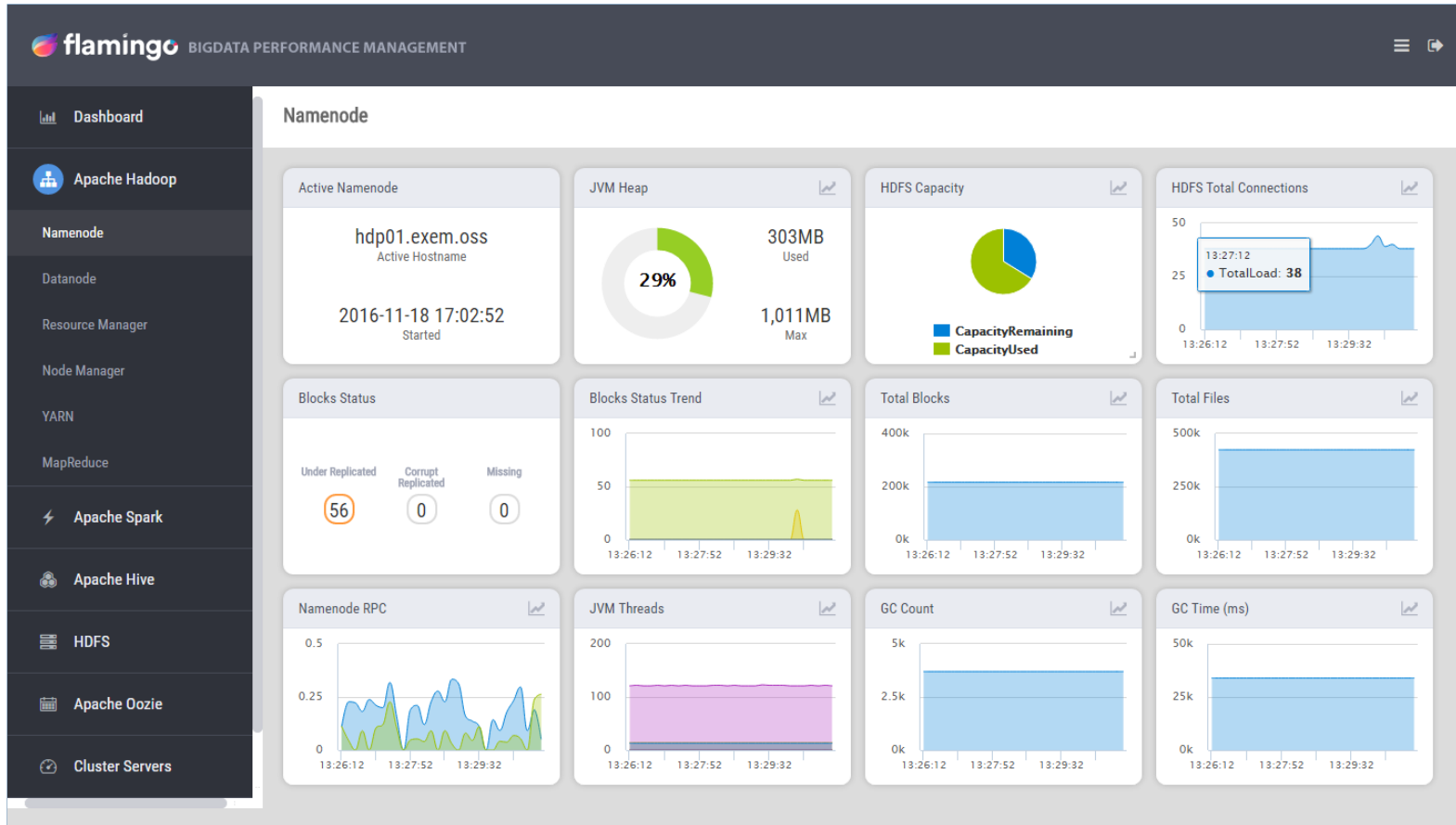


cloudera

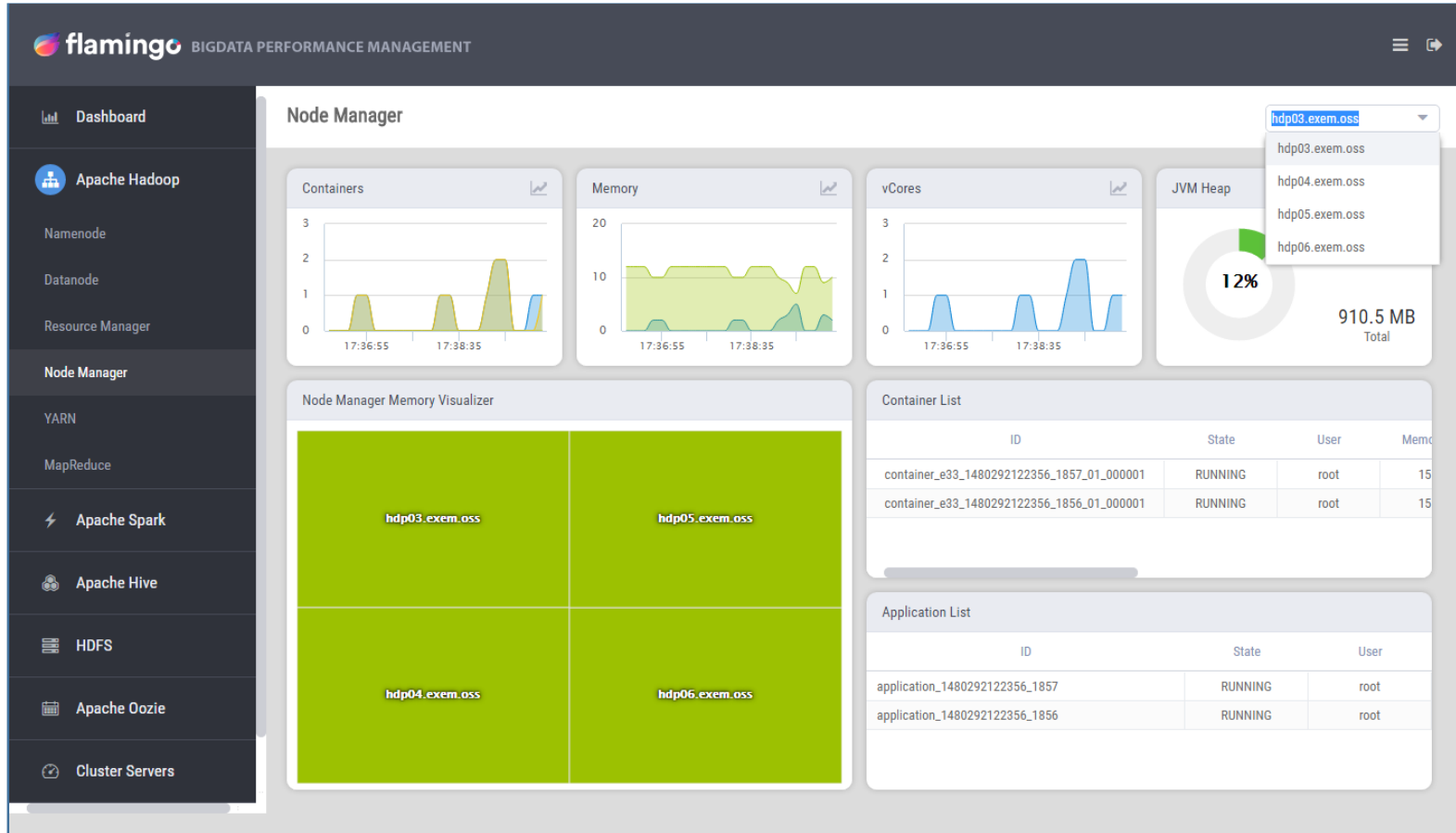

Hortonworks

Pivotal™

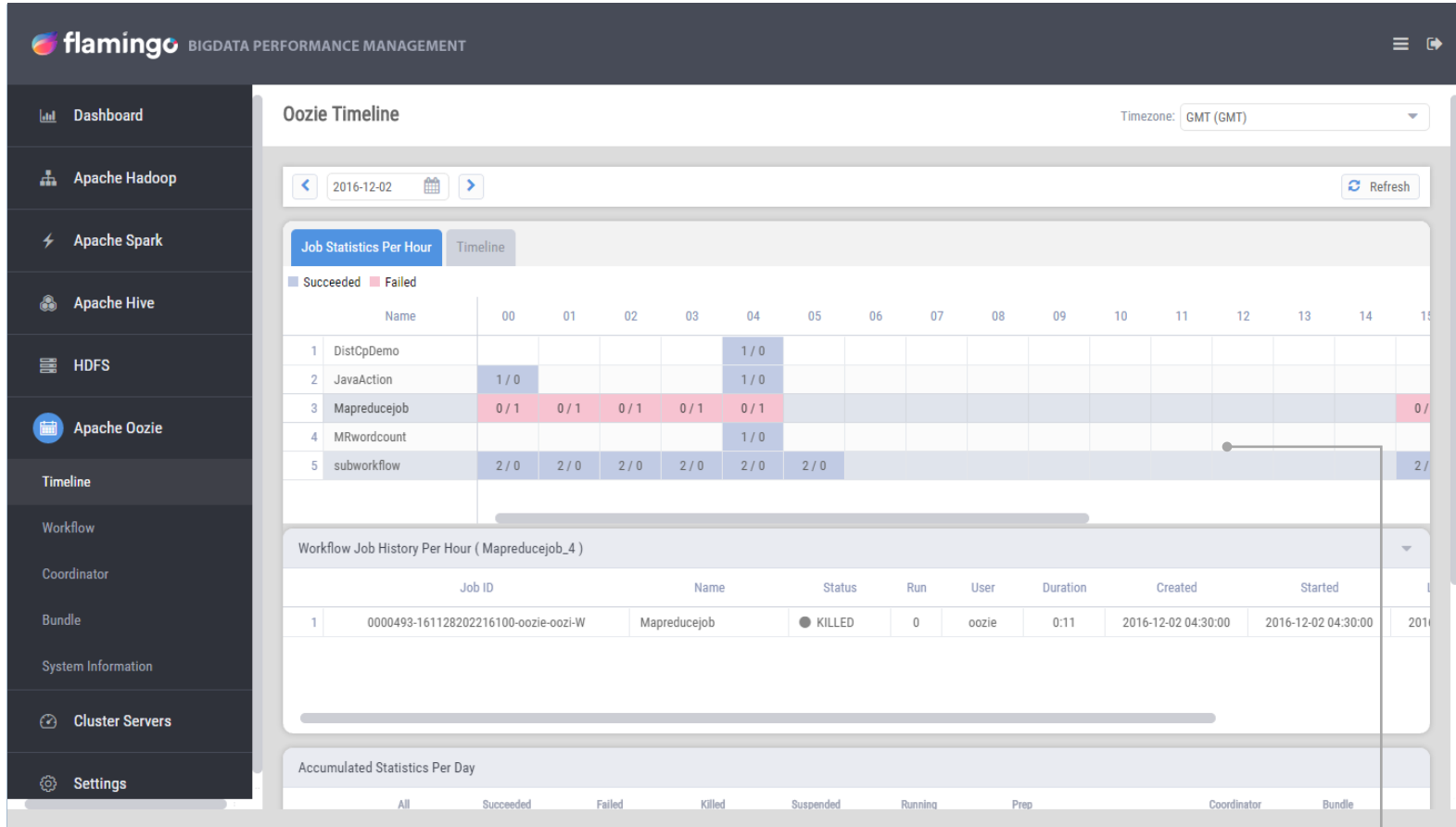
"직관적이고 쉬운 사용자 인터페이스"



" 노드별 접근방식을 통한 분석 애플리케이션의 분배 현황 "



“배치 작업의 시간단위 Timeline 모니터링을 통한 효율적인 운영 관리”



Oozie Workflow, Coordinator Job의 1시간 단위의 그룹핑

“디렉토리에 수십만개의 파일도 거뜬하게 표현하는 HDFS 브라우저”

The screenshot shows the Flamingo HDFS Browser interface. On the left is a navigation sidebar with options like Dashboard, Apache Hadoop, Apache Spark, Apache Hive, HDFS, HDFS Browser, HDFS Browser Audit, HDFS Audit, Apache Oozie, Cluster Servers, and Settings. The main area is titled 'HDFS Browser' and contains a 'Directory Navigation' section with a 'Root' breadcrumb. Below this is a 'Directory Usage Map' tab and a toolbar with actions like Create, Copy, Move, Rename, Delete, Upload, Download, View, and Permission. The core of the interface is a table listing HDFS directories with columns for Name, Size, Consumed, Replication, Modified, Owner, Group, and Authority. The table lists 17 directories, including 'app-logs', 'apps', 'ats', 'flamingo', 'hdfs', 'hdp', 'many', 'mapred', 'mr-history', 'oozie', 'output', 'spark-history', and 'system'. At the bottom of the table, it indicates 'Page 1 of 1' and 'Displaying 1 - 17 of 17'.

Name	Size	Consumed	Replication	Modified	Owner	Group	Authority
app-logs	321.5 MB	964.4 MB		2016-11-28 21:04:32	yarn	hadoop	rwrxrwx
apps	96.1 GB	290.3 GB		2016-09-13 15:25:56	hdfs	hdfs	rwxf-r-x
ats	0 bytes	0 bytes		2016-09-13 15:22:14	yarn	hadoop	rwxf-r-x
flamingo	1.3 MB	3.8 MB		2016-11-03 12:53:05	admin	admin	rwxf-r-x
hdfs	0 bytes	0 bytes		2016-09-28 09:07:05	hdfs	hdfs	rwxf-r-x
hdp	701.3 MB	2.1 GB		2016-11-03 12:38:04	hdfs	hdfs	rwxf-r-x
many	2.7 MB	8 MB		2016-09-26 17:39:43	hdfs	hdfs	rwxf-r-x
mapred	0 bytes	0 bytes		2016-09-13 15:22:18	mapred	hdfs	rwxf-r-x
mr-history	266.4 MB	799.3 MB		2016-09-13 15:22:25	mapred	hadoop	rwrxrwx
oozie	69.8 KB	209.3 KB		2016-11-30 14:48:26	hdfs	hdfs	rwxf-r-x
output	0 bytes	0 bytes		2016-11-04 15:31:48	hdfs	hdfs	rwxf-r-x
spark-history	185.1 MB	555.3 MB		2016-12-02 13:53:49	spark	hadoop	rwrxrwx
system	0 bytes	0 bytes		2016-10-12 15:38:05	hdfs	hdfs	rwxf-r-x

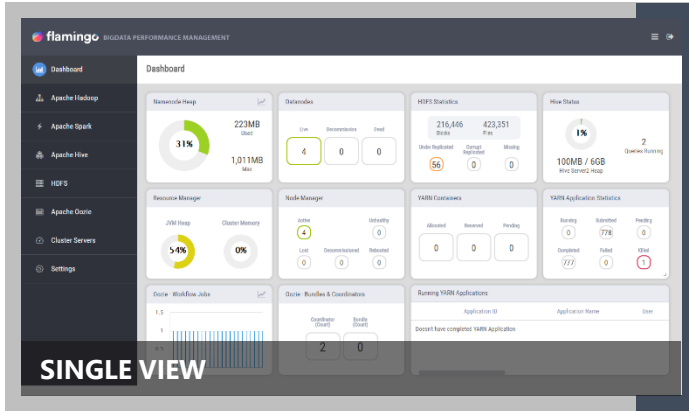
하나의 디렉토리에 수십만개의 파일을 성능 저하없이 관리하는 능력



"Maximize Hadoop Cluster Performance with Flamingo"

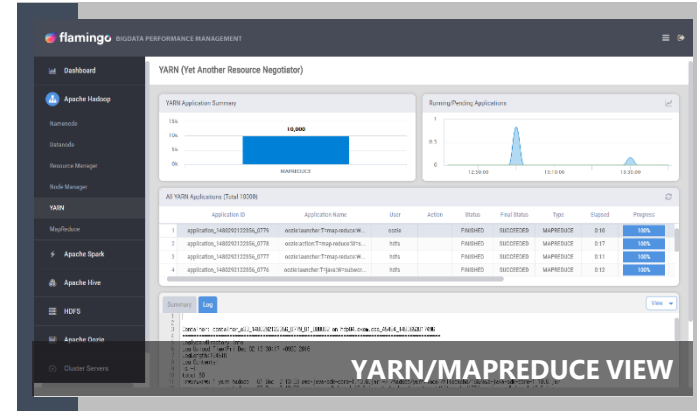
“Hadoop EcoSystem 성능관리에 최적화된 Monitoring View”

HADOOP CLUSTER VIEW / HIVE VIEW / YARN/MAPREDUCE VIEW / OOZIE VIEW



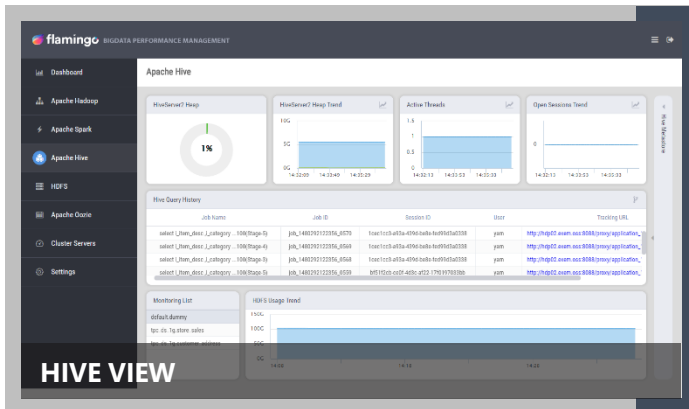
SINGLE VIEW

Hadoop Cluster 핵심 모니터링 지표에 대한 집중적인 모니터링



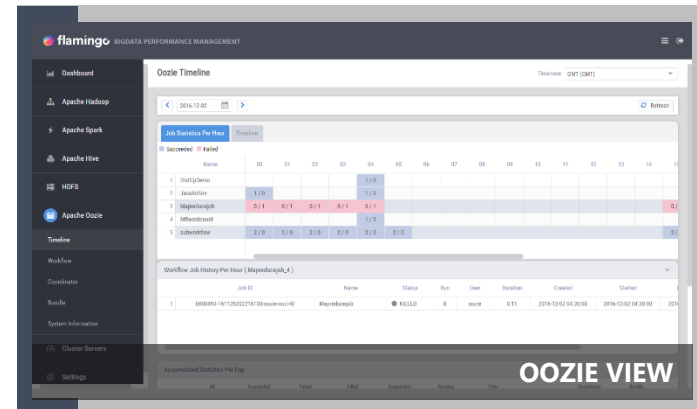
YARN/MAPREDUCE VIEW

YARN/MAPREDUCE 핵심 관리 항목의 모니터링에 최적화



HIVE VIEW

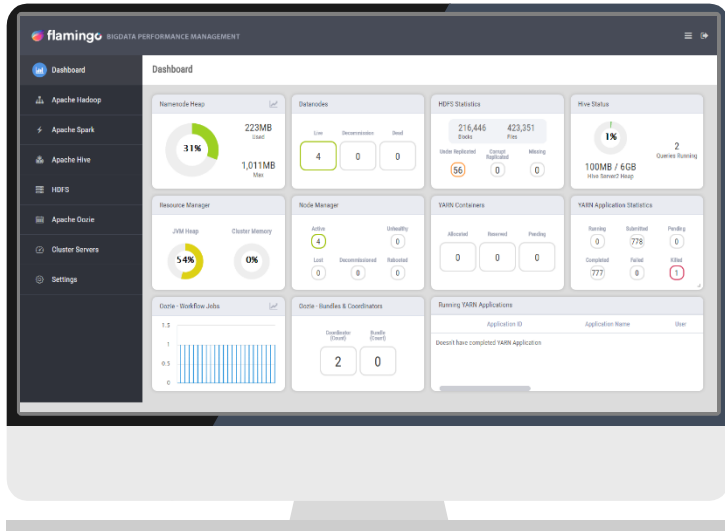
HIVE 핵심 관리 항목의 모니터링에 최적화



OOZIE VIEW

OOZIE 핵심 관리 항목의 모니터링에 최적화

“Hadoop Cluster 종합 감시”를 위한 전용 화면 제공”



■ 제품 화면 상세

Resource Manager

JVM Heap: 54%

Cluster Memory: 0%

Hive Status

1%

2 Queries Running

100MB / 6GB
Hive Server2 Heap

YARN Application Statistics

Running: 0	Submitted: 778	Pending: 0
Completed: 777	Failed: 0	Killed: 1

Namenode Heap

31%

223MB Used / 1,011MB Max

HDFS Statistics

216,446 Blocks / 423,351 Files

Under Replicated: 56 / Corrupt Replicated: 0 / Missing: 0

Oozie - Workflow Jobs

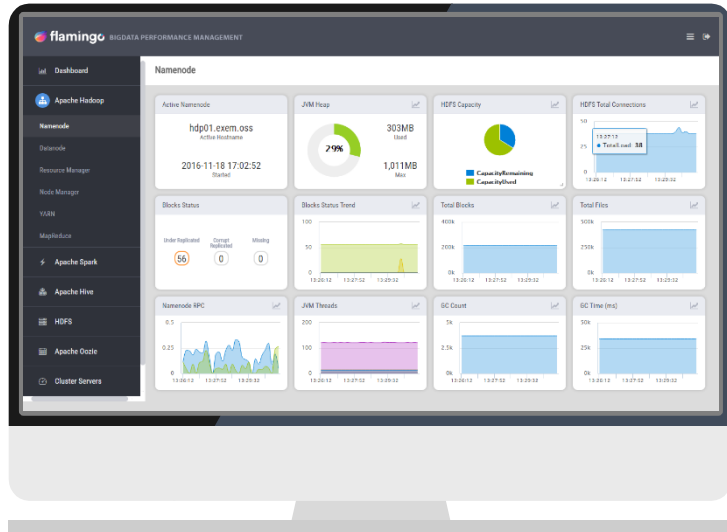
View Type DASHBOARD VIEW

DASHBOARD View를 통해 Hadoop Cluster의 핵심 지표(**Resource Manager**, **Namenode**, **HDFS**, **Hive**, **YARN**, **Node Manager**, **Oozie** 등) 모니터링에 대해서 쉽게 파악할 수 있습니다.

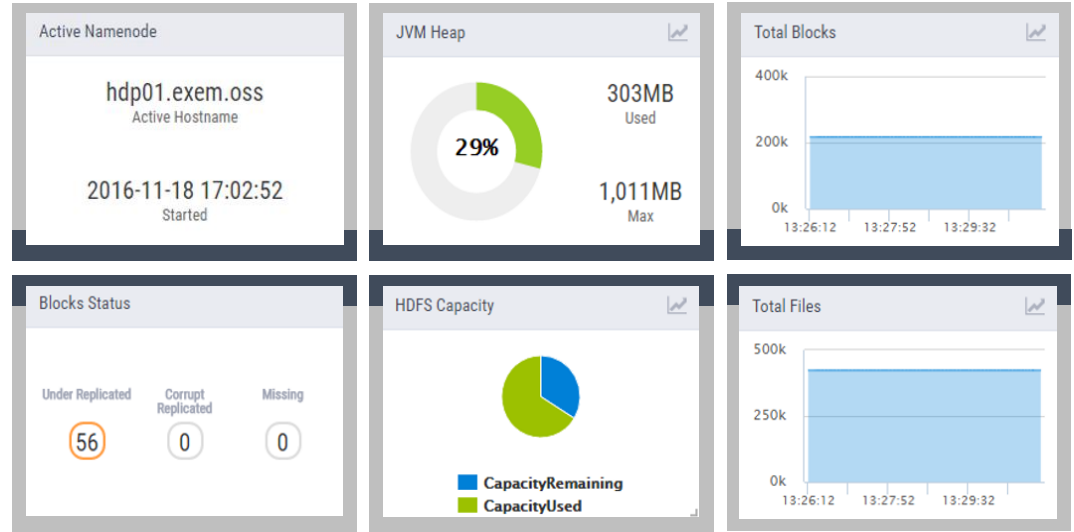
Feature

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

“Namenode 핵심 지표의 모니터링을 위한 전용 화면 제공”



제품 화면 상세



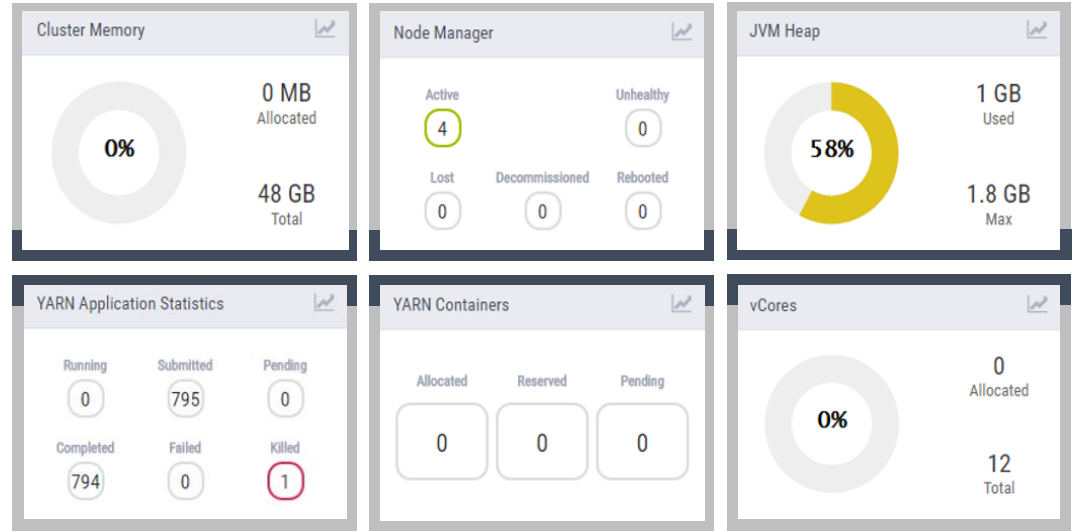
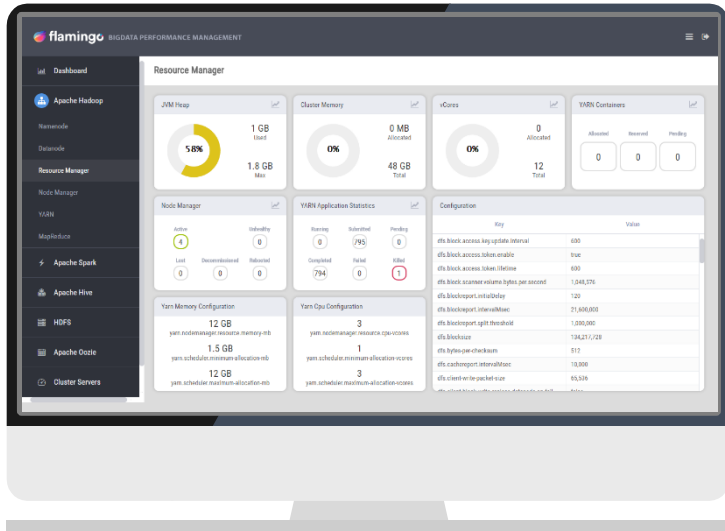
View Type NAMENODE VIEW

분산 파일 시스템을 관리하는 Namenode에 대해서 가장 중요한 성능 지표 (**JVM Heap**, **Active Namenode**, **HDFS Capacity**, **Block Status**) 모니터링에 대해서 쉽게 파악할 수 있습니다.

Feature

- Hadoop Namenode에서 Collector로 실시간 Pushing을 통해 수집하므로 데이터의 정확성 및 실시간성 보장
- Namenode HA시 Active Namenode에 상태 표시
- HDFS Metadata 수 증가시 JVM Heap 부족으로 인한 장애 모니터링
- Block 상태 모니터링을 통해 운영자가 HDFS에 파일 개짐에 대응
- HDFS의 용량 및 Block/File 개수의 실시간 모니터링

“Resource Manager의 자원 분배 및 상황에 대한 전용 화면 제공”



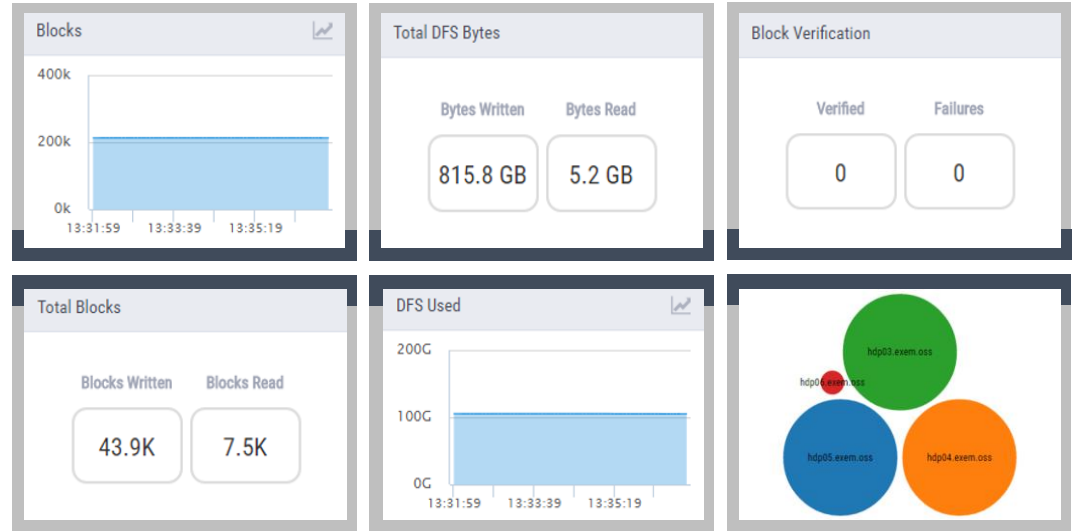
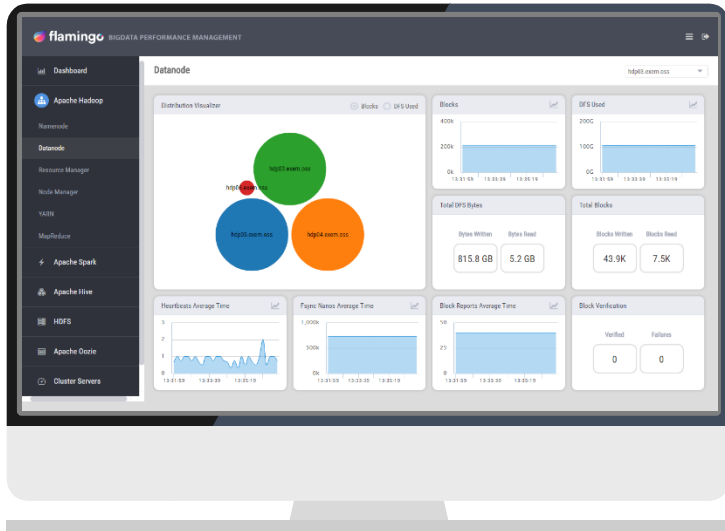
View Type RESOURCE MANAGER VIEW

RESOURCE MANAGER는 YARN Cluster의 전체 리소스를 관리하고 NODE MANAGER를 관리하는 주체입니다. Flamingo는 RESOURCE MANAGER에 핵심 모니터링 지표를 모니터링하여 관리자가 애플리케이션 튜닝을 위한 정보를 손쉽게 파악할 수 있습니다.

Feature

- 10초 주기 실시간 모니터링 정보 수집 (Polling이 아닌 Pushing 데이터 수집을 통한 정확성 확보 및 누락 데이터 방지)
- YARN Cluster 메모리 사용량
- vCore, Container 사용량
- Node Manager 정상 동작 여부
- YARN Application 상태
- YARN 설정 정보

“Datanode 별 블록 및 파일 분포 상태에 대한 전용 화면 제공”



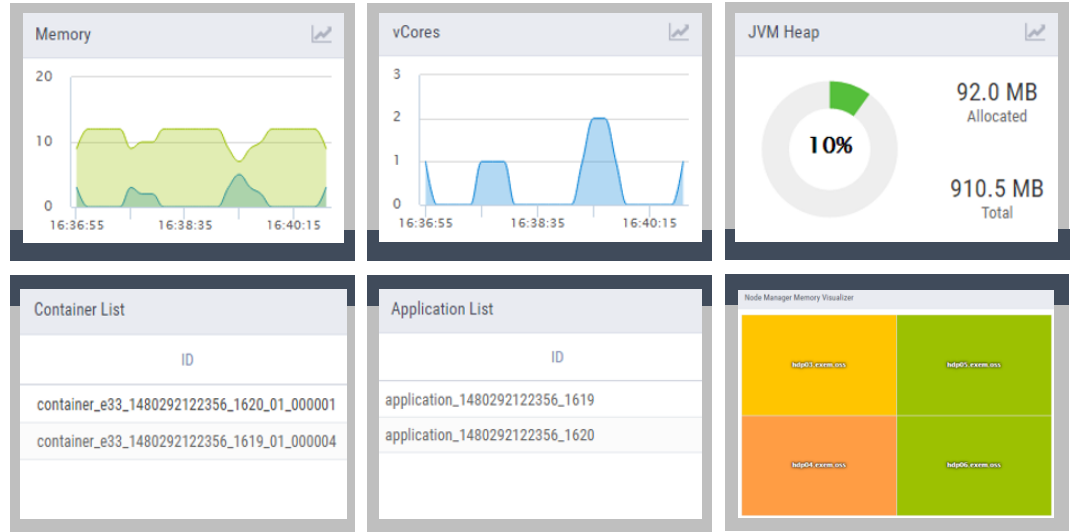
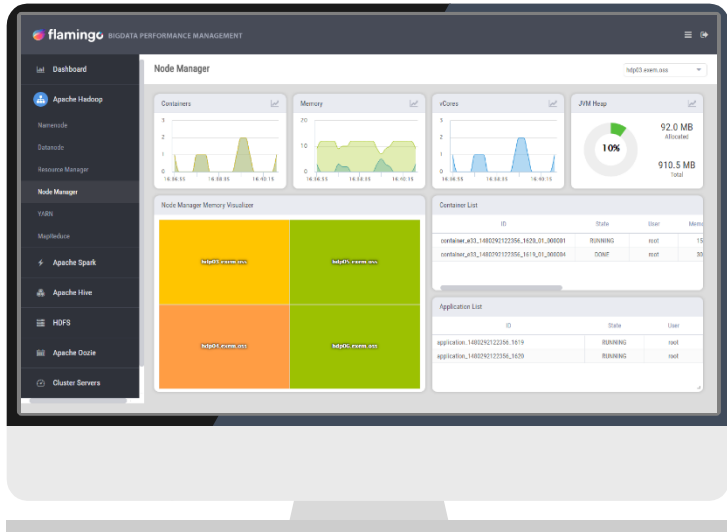
View Type DATANODE VIEW

다수의 노드로 구성되는 DATANODE는 파일의 블록을 저장하고 분석 작업을 수행하는 노드로서 Flamingo에서는 각 노드별 지표를 확인할 수 있도록 지원하여 특정 데이터 노드의 문제점을 파악할 수 있습니다.

Feature

- 10초 주기 실시간 모니터링 정보 수집 (Polling이 아닌 Pushing 데이터 수집을 통한 정확성 확보 및 누락 데이터 방지)
- 데이터 노드의 블록, 분산 파일 시스템 사용량에 대한 분포도
- 노드별 블록 개수
- DFS 크기, DFS 사용량 표시
- Namenode와 통신 상태

“각 노드별 분석 작업 및 vCore 등의 리소스 모니터링을 전용 화면 제공”



View Type NODE MANAGER VIEW

NODE MANAGER는 RESOURCE MANAGER의 요청에 따라서 컨테이너를 실행하고 vCore, RAM을 할당하여 작업을 모니터링합니다. FLAMINGO에서는 각 노드의 자원을 노드별로 사용량을 확인하고

Feature

- 10초 주기 실시간 모니터링 정보 수집 (Polling이 아닌 Pushing 데이터 수집을 통한 정확성 확보 및 누락 데이터 방지)
- 각 노드별 YARN Application 자원 분배 현황 확인
- 각 노드별 Application, Container 목록 확인
- vCore, RAM 사용량 확인
- JVM Heap Size 확인

YARN을 위한 Summary, Log 등의 핵심 정보에 대한 전용 화면 제공"

아플리케이션 유형별 통계

실행중 및 지연중인 작업의 개수

YARN Application의 실행 목록 및 상태

아플리케이션 실행 로그

View Type YARN VIEW

YARN은 Hadoop 2에서 자원을 관리하는 핵심 서비스로 FLAMINGO에서는 YARN을 관리하고 개발하고 분석하는데 있어서 가장 중요한 로그를 볼 수 있도록 하며, YARN 실행 이력을 표시하여 관리자들이 애플리케이션을 손쉽게 모니터링할 수 있습니다.

Feature

- 애플리케이션 유형별 통계 (예; MAPREDUCE, TEZ, SPARK 등)
- 실행중/지연중인 YARN Application 추이 그래프
- 실행중 및 완료된 YARN 애플리케이션 목록 및 각종 정보
- YARN 애플리케이션의 실행 로그
- YARN 애플리케이션의 실행에 대한 요약정보

MapReduce를 위한 Task, Conf, Counter 등의 핵심 지표에 대한 전용 화면 제공"

The screenshot displays the 'flamingo' interface for 'MapReduce' job monitoring. It includes a sidebar with navigation options like 'Dashboard', 'Apache Hadoop', 'YARN', 'MapReduce', 'Apache Spark', 'Apache Hive', 'HDFS', 'Apache Oozie', and 'Cluster Servers'. The main content area shows 'MapReduce Job Statistics Per Day' as a bar chart, 'MapReduce Job History (Total 1350)' as a table, and 'MR Job Summary' as a key-value table.

Annotations on the left side of the screenshot:

- MapReduce Job의 일별 통계 (MapReduce Job daily statistics) - points to the bar chart.

Annotations on the right side of the screenshot:

- MapReduce Job의 실행 목록 및 상태 (MapReduce Job execution list and status) - points to the job history table.
- MapReduce Job의 세부 실행 상태 (MapReduce Job detailed execution status) - points to the MR Job Summary table.

Job ID	Workflow Name	Status	User	Queue	Map	Reduce	Elapsed	Start	Finish
1	job_1480292122356_0779	SUCCEEDED	oozie	default	1	0	0:05	2016-12-02 13:30:05	2016-12-02 13:30:10
2	job_1480292122356_0778	SUCCEEDED	hdfs	default	2	1	0:12	2016-12-02 13:00:29	2016-12-02 13:00:41
3	job_1480292122356_0777	SUCCEEDED	hdfs	default	1	0	0:07	2016-12-02 13:00:18	2016-12-02 13:00:25
4	job_1480292122356_0776	SUCCEEDED	hdfs	default	1	0	0:08	2016-12-02 13:00:04	2016-12-02 13:00:12

Key	Value
submitTime	2016-12-02 13:30:00
startTime	2016-12-02 13:30:05
finishTime	2016-12-02 13:30:10

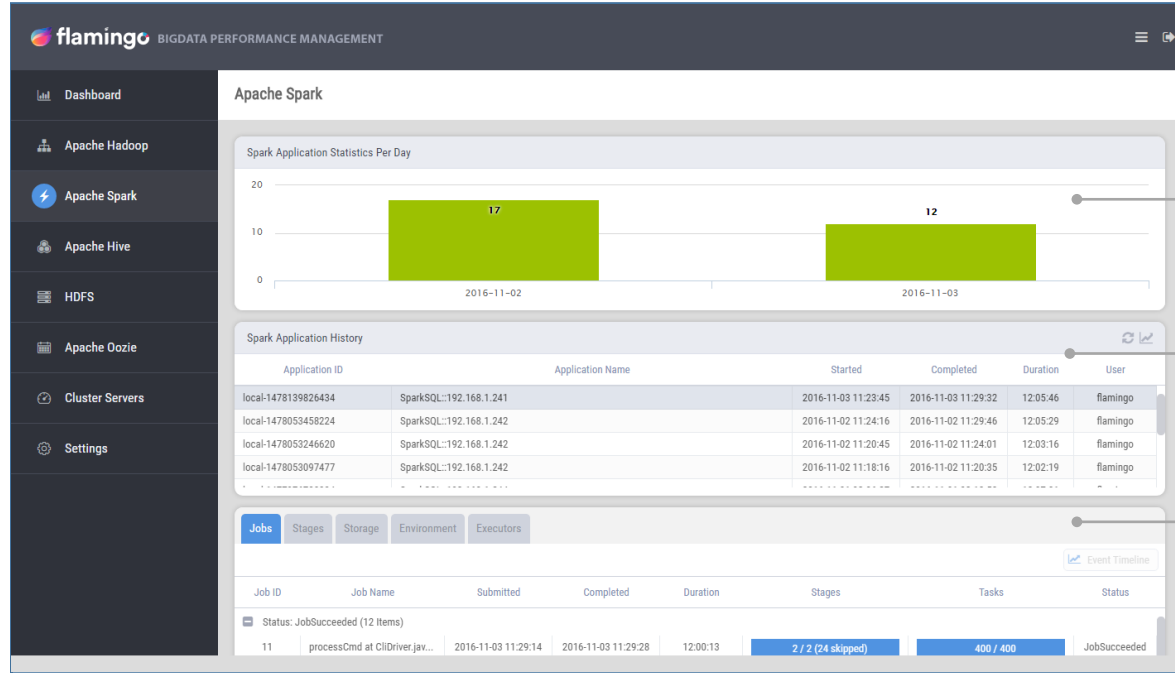
View Type MAPREDUCE VIEW

MAPREDUCE는 Map과 Reduce로 구성되어 있으며 분석 작업을 구현하는 프레임워크입니다. FLAMINGO는 Map, Reduce Task에 대한 로그 정보에서부터 Timeline에 이르는 등, 관리자가 MapReduce Job 모니터링을 위해서 필요한 핵심 지표를 모니터링 할 수 있도록 지원합니다.

Feature

- MapReduce Job의 일별 통계
- MapReduce Job의 실행 이력
- MapReduce Job의 Summary 정보 (실행 시간, 구간별 처리 시간 상태, MR Task 개수등)
- MapReduce Job Counter Tree
- MapReduce Job Configuration
- MapReduce Job Task 목록 및 Task별 실행 로그
- MapReduce Job Task의 Timeline

“Spark Job에 특화된 전용 모니터링 화면 제공



일별 Spark Job 실행 통계

Spark Job의 실행 이력

개별 Spark Job의 상세 모니터링 정보

View Type SPARK VIEW

SPARK은 인메모리 분석 프레임워크로서 Hadoop EcoSystem의 핵심 서비스가 되어가고 있습니다. FLAMINGO에서는 Spark Job의 실행 이력부터 모니터링 정보를 상세하게 정보를 표시하는 모니터링을 제공합니다.

Feature

- Spark Job의 일별 통계
- Spark Job의 실행 이력
- Spark Job의 Job ID 별 실행 이력 및 Event Timeline
- Spark Job의 Stage 별 실행 상태
- Spark Job의 Storage 상태
- Spark Job의 Executor의 동작 상태 및 Timeline

“HDFS의 파일 및 디렉토리를 관리하기 위한 관리자를 위한 모든 기능을 제공

The screenshot shows the 'HDFS Browser' interface in the Flamingo BigData Performance Management system. The interface includes a sidebar with navigation options like Dashboard, Apache Hadoop, Apache Spark, Apache Hive, HDFS, HDFS Browser Audit, HDFS Audit, Apache Oozie, Cluster Servers, and Settings. The main content area displays a 'Directory Usage Map' (a red and green heatmap) and a 'Directory Navigation' section with a table of HDFS files and directories. The table has columns for Name, Size, Consumed, Replication, Modified, Owner, Group, and Authority. Annotations with arrows point to specific features: '파일 및 디렉토리 관리 기능' (File and Directory Management Function) points to the sidebar; '디렉토리 사용량 맵' (Directory Usage Map) points to the heatmap; '디렉토리 이동 네비게이션 바' (Directory Navigation Bar) points to the navigation controls; and '파일 및 디렉토리 목록' (File and Directory List) points to the table.

Name	Size	Consumed	Replication	Modified	Owner	Group	Authority
app-logs	321.5 MB	964.4 MB		2016-11-28 21:04:32	yam	hadoop	rwXrwxrwx
apps	96.1 GB	290.3 GB		2016-09-13 15:25:56	hdfs	hdfs	rwXr-xr-x
ats	0 bytes	0 bytes		2016-09-13 15:22:14	yam	hadoop	rwXr-xr-x
flamingo	1.3 MB	3.8 MB		2016-11-03 12:53:05	admin	admin	rwXr-xr-x
hdfs	0 bytes	0 bytes		2016-09-28 09:07:05	hdfs	hdfs	rwXr-xr-x
hdp	701.3 MB	2.1 GB		2016-11-03 12:38:04	hdfs	hdfs	rwXr-xr-x
many	2.7 MB	8 MB		2016-09-26 17:39:43	hdfs	hdfs	rwXr-xr-x
mapred	0 bytes	0 bytes		2016-09-13 15:22:18	mapred	hdfs	rwXr-xr-x
mr-history	266.4 MB	799.3 MB		2016-09-13 15:22:25	mapred	hadoop	rwXrwxrwx
oozie	69.8 KB	209.3 KB		2016-11-30 14:48:26	hdfs	hdfs	rwXr-xr-x
output	0 bytes	0 bytes		2016-11-04 15:31:48	hdfs	hdfs	rwXr-xr-x
spark-history	185.1 MB	553.3 MB		2016-12-02 13:53:49	spark	hadoop	rwXrwxrwx
system	0 bytes	0 bytes		2016-10-12 15:38:05	hdfs	hdfs	rwXr-xr-x

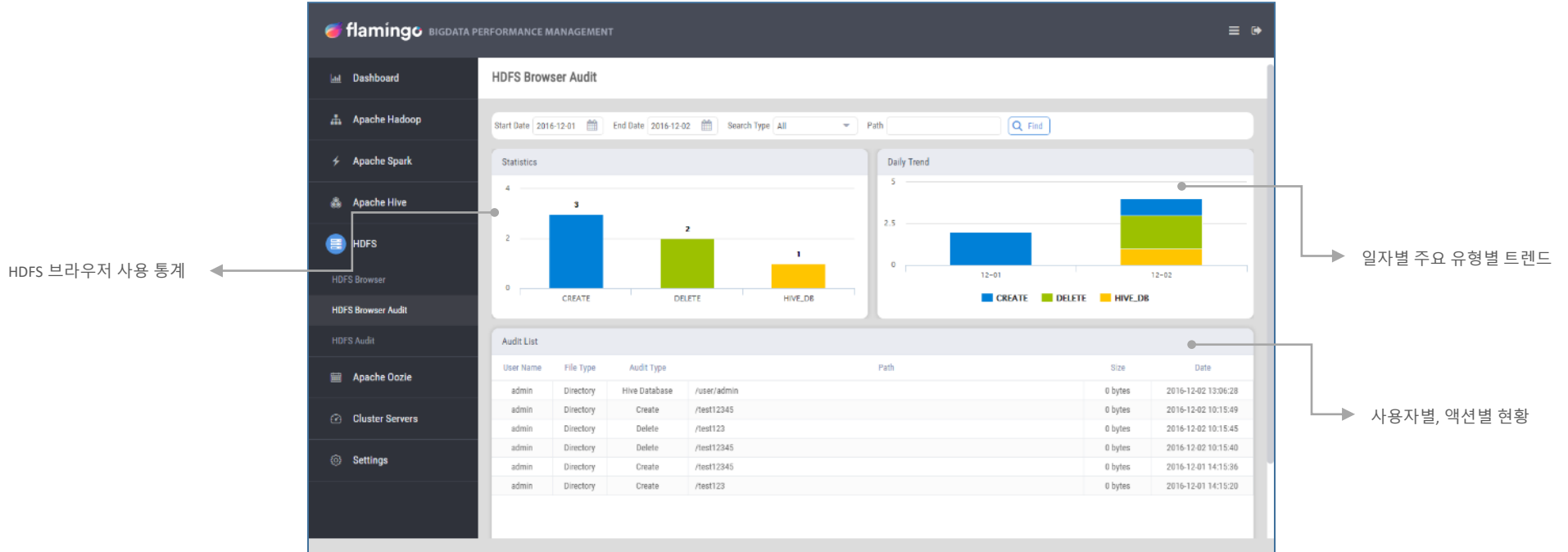
View Type HDFS BROWSER VIEW

HDFS 브라우저는 HDFS 상의 디렉토리 및 파일을 관리하는 핵심 기능으로 FLAMINGO의 핵심 기능입니다. 파일을 업로드하고 다운로드할 수 있으며, 수십 만개의 파일도 거뜬히 표시합니다.

Feature

- HDFS의 파일 및 디렉토리 관리
- 하나의 디렉토리에 수십만개의 파일이 있더라도 UI 문제없이 페이징
- 디렉토리 및 파일의 권한 관리
- 파일 내용 보기 및 파일 업로드/다운로드
- 문자열 기반 디렉토리 경로 입력시 해당 디렉토리로 이동
- 디렉토리별 사용량 표시 맵

“HDFS 브라우저의 사용자 행위에 대한 세부 모니터링 화면 제공”



View Type

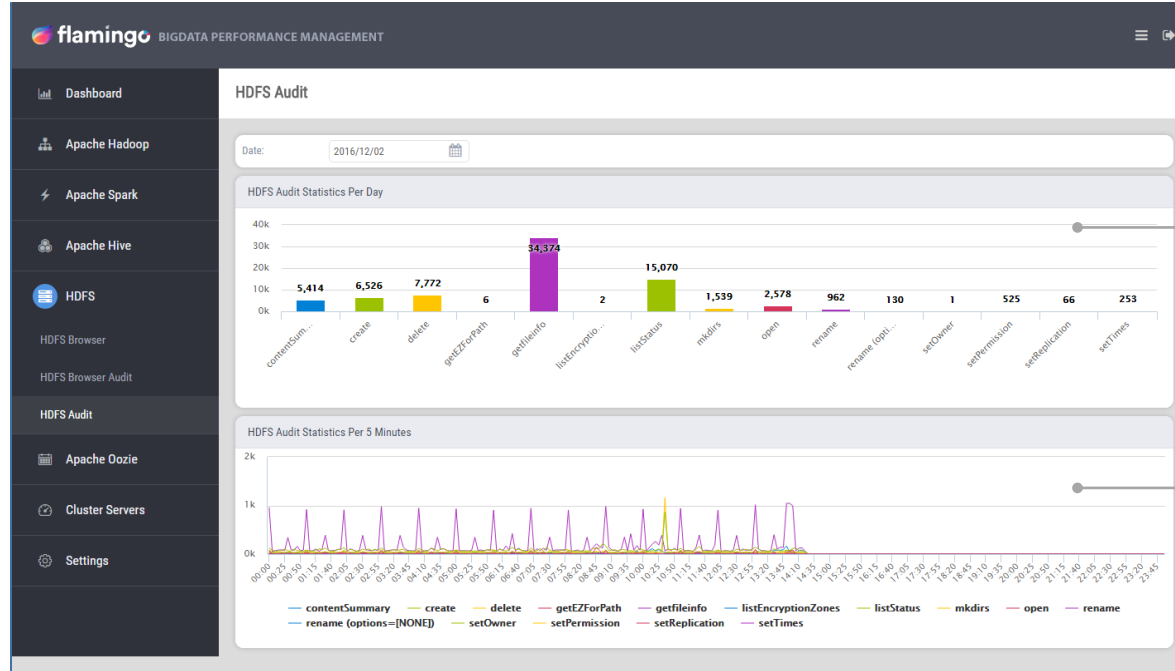
HDFS BROWSER AUDIT VIEW

NAMENODE View를 통해 Hadoop Cluster의 핵심 지표(**Resource Manager, Namenode, HDFS, Hive, YARN, Node Manager, Oozie** 등) 모니터링에 대해서 쉽게 파악할 수 있습니다.

Feature

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

HDFS 파일 시스템의 모든 이벤트를 모니터링”



HDFS 이벤트별 통계

HDFS 이벤트별 추이 그래프

View Type

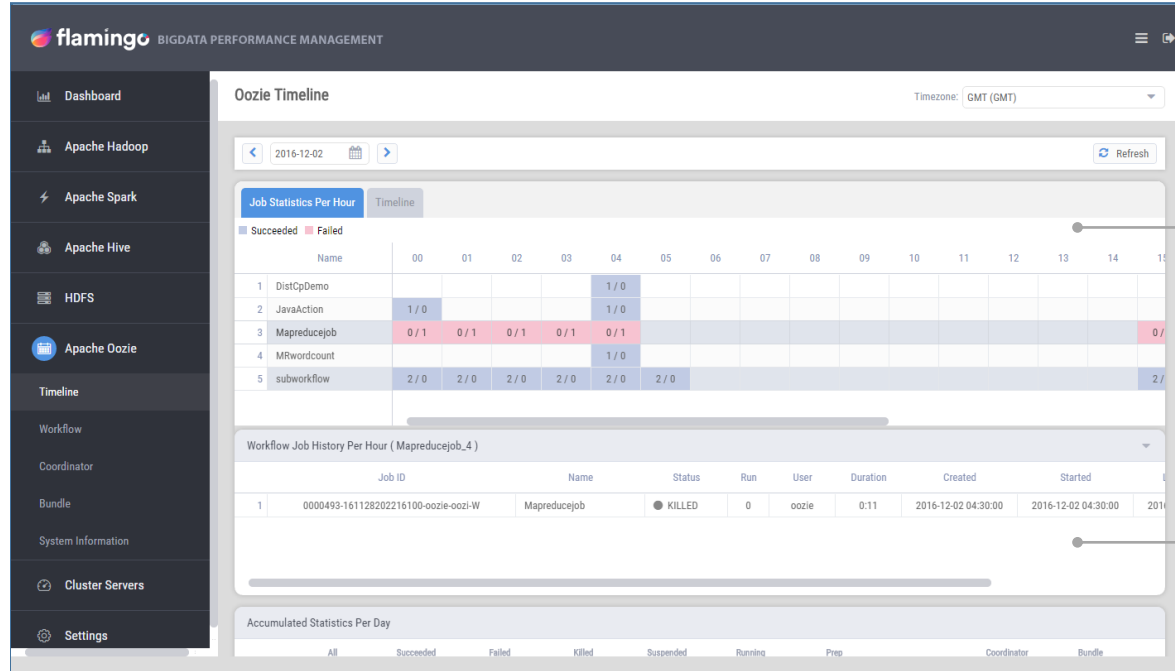
HDFS FILESYSTEM AUDIT VIEW

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

Feature

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

“Oozie 모니터링의 재해석”



Oozie Workflow Timeline

시간대별 Oozie Workflow 이력

View Type OOZIE TIMELINE VIEW

Oozie Workflow 모니터링은 운영자/관리자 관점에서 시간대별로 현황이 한눈에 보일 수 있어야 합니다. FLAMINGO는 운영자/관리자 관점에서 Oozie 모니터링을 재해석했습니다.

Feature

- Oozie Workflow Job에 대한 Timeline
- 해당 시간대에 실행된 Oozie Workflow 실행 이력 제공
- 해당 Oozie Workflow의 누적 통계
- Timezone 지원
- Oozie Workflow Job의 세부 실행 정보 표출

“세상에서 가장 빠른 Oozie 관리 및 모니터링 제공”

조회 조건 및 Workflow 관리

Job ID	Started	Status	Name	Duration	User	Last Modified
0000499-161128202216100-oozie-oozi-W	2016-12-02 05:30:00	KILLED	Mapreducejob	0:11	oozie	2016-12-02 05:30:11
0000498-161128202216100-oozie-oozi-W	2016-12-02 05:00:13	SUCCEEDED	subworkflow	0:29	hdfs	2016-12-02 05:00:42
0000497-161128202216100-oozie-oozi-W	2016-12-02 05:00:00	SUCCEEDED	subworkflow	0:42	hdfs	2016-12-02 05:00:42
0000496-161128202216100-oozie-oozi-W	2016-12-02 04:55:02	SUCCEEDED	DistCpDemo	0:26	root	2016-12-02 04:55:28
0000495-161128202216100-oozie-oozi-W	2016-12-02 04:54:42	SUCCEEDED	JavaAction	0:15	root	2016-12-02 04:54:57
0000494-161128202216100-oozie-oozi-W	2016-12-02 04:54:35	SUCCEEDED	MRwordcount	0:31	root	2016-12-02 04:55:06
0000493-161128202216100-oozie-oozi-W	2016-12-02 04:30:00	KILLED	Mapreducejob	0:11	oozie	2016-12-02 04:30:11
0000492-161128202216100-oozie-oozi-W	2016-12-02 04:00:13	SUCCEEDED	subworkflow	0:30	hdfs	2016-12-02 04:00:43
0000491-161128202216100-oozie-oozi-W	2016-12-02 04:00:00	SUCCEEDED	subworkflow	0:43	hdfs	2016-12-02 04:00:43
0000490-161128202216100-oozie-oozi-W	2016-12-02 03:30:00	KILLED	Mapreducejob	0:11	oozie	2016-12-02 03:30:11
0000489-161128202216100-oozie-oozi-W	2016-12-02 03:00:14	SUCCEEDED	subworkflow	0:30	hdfs	2016-12-02 03:00:44

워크플로우 실행 이력

View Type OOZIE WORKFLOW VIEW

Oozie Workflow는 다수의 실행 작업이 포함되어 있는 워크플로우입니다. FLAMINGO는 기존의 Oozie Console의 문제점(성능, 편의성, 정보 표현의 부족 등)을 획기적으로 개선한 모니터링을 제공합니다.

Feature

- 날짜별, 상태별, ID 등의 기준에 따라서 Workflow 조회
- Oozie Workflow의 Stop, Kill, Rerun 지원
- Oozie Workflow의 실행 이력 및 모니터링
- Oozie Workflow의 Action별 실행 이력 및 모니터링
- Oozie Workflow의 XML 및 각종 정보 모니터링

“Hive 모니터링을 위한 Total 솔루션”

중요 장애 지표 모니터링

Hive Query 실행 이력 및
Hive Topology View

Hive Table 용량 변화 감시

Hive Metastore 관리

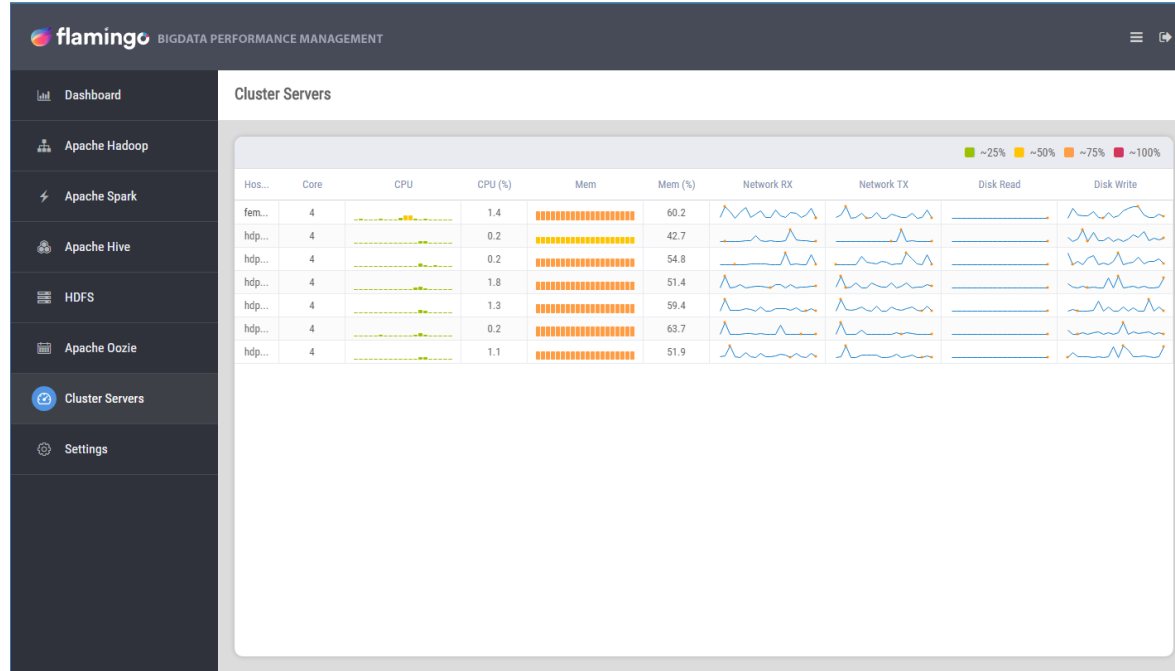
View Type HIVE VIEW

FLAMINGO는 HIVE SERVER 및 HIVE JOB을 모니터링 하기 위한 Total 솔루션입니다. Hive Query는 다수의 YARN, MapReduce로 동작하여 모니터링이 어렵지만 FLAMINGO는 Topology View를 통해서 그래프로 추적합니다. 또한 Hive Server의 주요 장애 지표를 모니터링합니다.

Feature

- 10초 주기 실시간 모니터링 정보 수집 (Polling이 아닌 Pushing 데이터 수집을 통한 정확성 확보 및 누락 데이터 방지)
- Hive Server 2의 JVM Heap Size 모니터링을 통한 장애 대응
- Hive Server 2의 Open Session 모니터링을 통한 접속 사용자수 확인
- Hive Metastore 관리기능을 이용한 테이블 및 데이터베이스 관리
- Hive Query Topology View (Hive Query Plan)
- Hive Table 용량 모니터링

하둡 클러스터 서버의 시스템 자원 모니터링 전용 화면 제공”



View Type

CLUSTER RESOURCE VIEW

FLAMINGO는 Hadoop Cluster를 구성하는 각 노드의 각종 자원 사용량을 모니터링하는 전용 화면을 제공합니다.

Feature

- 실시간 자원 모니터링
- Core, Memory, Network, Disk 사용량 모니터링
- Core별 CPU 사용량 모니터링
- 노드별/프로세스별 CPU, RAM 사용량 모니터링

THANK YOU