EBIGS (Exem Bigdata System) 1.0

빅데이터사업본부 기술컨설팅팀

© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

INDEX

- 제품 개요
- 시스템 아키텍쳐
- 시스템 구성
- 주요 구성 정보
- 시스템 구성 정보
- 시스템 H/W 구성
- 주요 기능
- 제품 스펙

© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

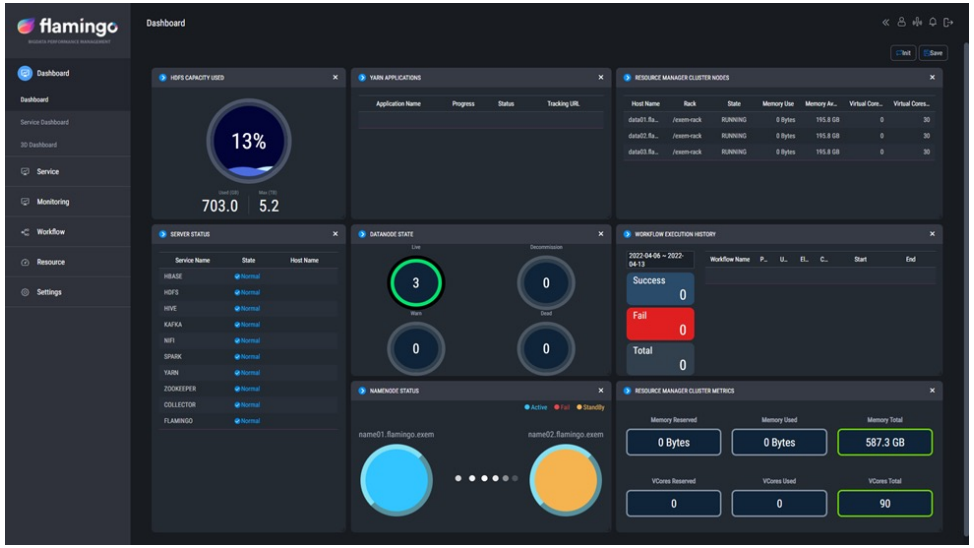
제품 개요

EBIGS(Exem Bigdata System)는 빅데이터를 수집, 정제, 저장, 분석하기 위한 소프트웨어 패키지 솔루션으로 자체 호환성 검증을 통한 Apache Hadoop Ecosystem과 이를 쉽게 설치, 변경, 모니터링, 운영하기 위한 오픈 소스 기반의 빅데이터 저장 관리 플랫폼입니다.

빅데이터 시스템에 대한 구성, 보안, 운영 등의 이슈들을 해결하고 Apache Hadoop Ecosystem 대한 Hadoop Cluster를 모니터링 및 운영할 수 있는 일관되고 안전한 플랫폼을 제공하여 시스템 관리의 부담을 최소화 합니다.

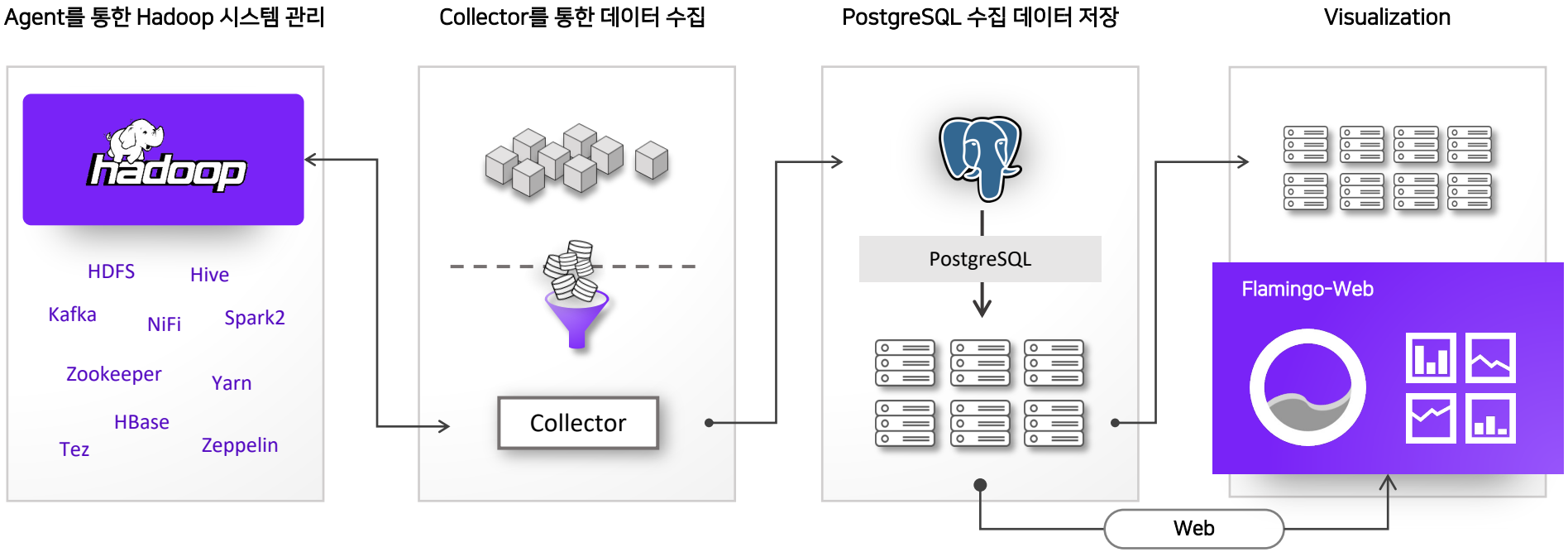
Feature

- 빅데이터 시스템에 대한 구성, 보안, 운영 등의 이슈를 해결합니다.
- Hadoop Ecosystem의 다양한 지표를 통합적인 관점에서 실시간 모니터링하고 관리하는 기능을 제공합니다.
- 단순 시스템, 성능 모니터링 차원을 넘어 손쉽고 간편한 운영 관리 모두를 제공합니다.
- Cluster 규모에 상관없이 손쉬운 유지보수 및 관리가 가능합니다.
- 정형, 반정형, 비정형 데이터의 배치, 실시간 수집 및 가공, 분석 처리를 지원합니다.



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

시스템 아키텍쳐



exem © Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

시스템 구성

exem © Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

주요 구성 정보

분산 파일 시스템의
병렬 처리 프레임워크에서 실행되어
대용량 데이터 처리 지원

Hadoop 전체 서비스에 대한
보안정책의 적용 및 관리

웹UI로 시스템간의 데이터 흐름 자동화
및 형식이 다른 프로토콜, 데이터의 분석
가능하며 데이터 배치 처리 지원

빠른 데이터 처리와 시스템 안정성 제공
및 데이터 실시간 스트리밍 지원

인-메모리 기반의 빠른 처리로
Hive의 레이턴시를 극복하여
데이터 처리 성능 향상

Hadoop Ecosystem의
성능 데이터를 최소한의 부하로
수집하여 모니터링 및 저장

© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

시스템 구성 정보 (1/2)

Name	상세 기능	채택 이유
Flamingo	Hadoop Ecosystem을 모니터링하고 서비스를 운영/관리하기 위한 시스템입니다.	여러 대의 서버와 서비스로 구성되어 운영이 어려운 Hadoop Ecosystem을 프로비저닝*, 관리, 모니터링하고 서비스 운영/관리를 통해 운영을 용이하게 하는 Flamingo를 개발/채택하였습니다.
Ranger	Hadoop Ecosystem 전반에 걸쳐 포괄적인 데이터 보안을 활성화, 모니터링 및 관리하는 시스템입니다.	Hadoop 내의 데이터 액세스에 대한 여러 사용 사례를 지원하는 동시에 보안 정책의 중앙 관리 및 사용자 액세스 모니터링을 제공하여 모든 보안 관련 작업 관리가 가능합니다.
HDFS	HDFS는 Java 언어로 작성된 분산, 확장 및 이식 등이 가능한 Hadoop의 주요 파일 시스템입니다.	방대한 양의 데이터를 처리하기 위해 예전 방식으로는 컴퓨팅 파워의 한계, 비용 문제가 발생합니다. 이를 극복하기 위해 스케일 아웃이 가능하며 분산 클러스터 환경에서 대용량 데이터 처리가 가능한 HDFS를 채택하였습니다.
Hive	Hadoop 기반의 데이터베이스 솔루션입니다. SQL과 매우 유사한 HiveQL 을 제공합니다.	파일시스템인 Hadoop은 데이터를 조회하기 위해서는 MapReduce코드를 작성해야 하는데 매우 번거롭고 어렵습니다. 하이브는 쿼리를 MapReduce 프로그램으로 변환해주는 쿼리 엔진으로 사용자에게 익숙한 쿼리를 이용하여 손쉽게 데이터 조회가 가능합니다.
Kafka	카프카는 웹사이트, 애플리케이션, 센서 등에서 취합한 데이터 스트림을 실시간으로 관리하는 시스템입니다.	데이터 스트림을 실시간으로 관리하고, 로드 밸런싱과 내고장성을 보장함으로 빠른 데이터 처리와 시스템의 안정성을 제공합니다. 대용량의 실시간 로그 데이터 처리에 특화되어 있으며 데이터의 유실없이 안전하게 전달합니다.
NiFi	Apache NiFi는 소프트웨어 시스템 간 데이터 흐름을 자동화하도록 설계된 아파치 소프트웨어 재단의 프로젝트입니다.	데이터 흐름을 모니터링 하기 위한 프레임워크로서 여러 네트워크를 통과하는 데이터 흐름을 웹UI에서 그래프로 표현하며, 프로토콜과 데이터 형식이 다르더라도 분석을 가능하게 해줍니다.
Spark2	Spark는 배치 처리, 스트리밍 분석/처리, 기계 학습, SQL 등을 지원하고 이들의 빠른 성능 지원을 위한 인-메모리 방식을 사용하는 분산 처리 시스템 입니다.	Hive의 높은 I/O 발생과 그로 인한 레이턴시를 극복하기 위해 데이터 가공/처리를 인메모리에서 빠르게 처리함으로 데이터 처리 성능을 향상시킵니다.

* 프로비저닝: 어떤 종류의 서비스든 사용자의 요구에 맞게 시스템 자체를 제공하는 것

exem © Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

시스템 구성 정보 (2/2)

Name	상세 기능	채택 이유
Yarn	Yarn은 수십, 수백개의 노드로 구성된 클러스터에서 데이터 처리 작업과 노드의 자원을 관리해주는 분산 자원 관리 솔루션입니다.	수십, 수백 개의 노드로 구성된 Hadoop Cluster에서 실행되는 수많은 작업들을 관리, 스케줄링하고 자원을 적절히 분산하여 관리해 줍니다.
Zookeeper	Zookeeper는 분산 컴퓨팅 환경에서 필요한 동기화, 그룹 서비스 등의 정보를 관리하는 시스템입니다.	수십, 수천 대의 서버에 설치돼 있는 분산 환경을 서버 간의 중요한 이벤트를 관리하면서 상호 작용을 조율함으로써 보다 효율적으로 관리할 수 있습니다.
HBase	HBase는 HDFS상에서 동작하는 NoSQL(비관계형) 데이터베이스 입니다.	하둡 기반의 컬럼 지향 NoSQL 데이터베이스로서 스키마 변경을 자유롭게 하고, 리전이라는 수십~수 백대의 분산 서버로의 복제 등의 기능을 지원해 성능과 안정성을 보장합니다.
Tez	Yarn기반의 데이터 처리를 위한 프레임워크 입니다.	Hive는 데이터 처리 작업 중간마다 발생되는 임시 데이터들을 디스크 상에 저장하여 매우 느립니다. 임시 데이터를 메모리에서 처리하는 Tez를 Hive의 쿼리엔진으로 사용하여 Hive의 높은 I/O 발생으로 인한 성능 저하를 향상시켜줍니다.
Zeppelin	웹 기반 notebook이며 일부 시각화 기능도 제공합니다.	빅데이터 분석가를 위한 웹 기반의 분석 도구이며, 분석결과를 즉시 시각화 합니다. iPython, Notebook과 유사한 노트북 기능을 제공하며, 분석가는 이를 통해 손쉽게 데이터를 추출, 정제, 분석, 공유할 수 있습니다.

exem © Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

시스템 H/W 구성

Manage Node #1	
OS	CentOS7.8
Flamingo Services	system-agent
	Collector
	Web
	Zeppelin
HDFS	Journal Node
YARN	-
Hadoop Ecosystem	Ranger
	ZooKeeper
	Hive
	Spark
	HBase
Etc	OpenJDK1.8
	PostgreSQL
	Apache NiFi

Name Node #1	
OS	CentOS7.8
Flamingo Services	system-agent
	rm-agent
	nm-agent
HDFS	Name Node1
	Failover Controller
	Journal Node
YARN	ResourceManager1
	History-server
Hadoop Ecosystem	ZooKeeper
	Hadoop client
	Hive Metastore
Etc	OpenJDK1.8

Name Node #2	
OS	CentOS7.8
Flamingo Services	system-agent
	rm-agent
	nm-agent
HDFS	NameNode2
	Failover Controller
	Journal Node
YARN	ResourceManager2
	History-server
Hadoop Ecosystem	ZooKeeper
	Hadoop client
Etc	OpenJDK1.8

Data Node #1~3	
OS	CentOS7.8
Flamingo Services	system-agent
HDFS	Data Node
YARN	Node Manager
Hadoop Ecosystem	Spark-worker
	Hbase-Region
	Hadoop client
	Kafka
Etc	OpenJDK1.8

Collector Node #1	
OS	CentOS7.8
Flamingo Services	system-agent
Hadoop Ecosystem	Hadoop client
	Kafka
	Apache NiFi
Etc	OpenJDK1.8

* 배치, 실시간 처리 지원

exem © Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

Hadoop Ecosystem 데이터 액세스 권한 관리를 위한 화면 제공

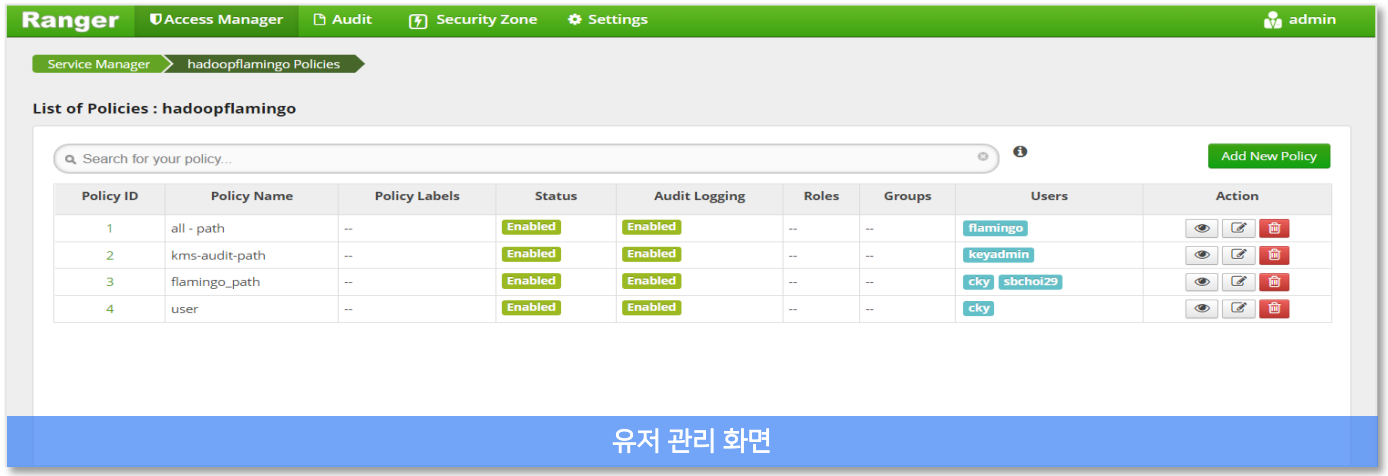
View Type

데이터 액세스 권한 관리

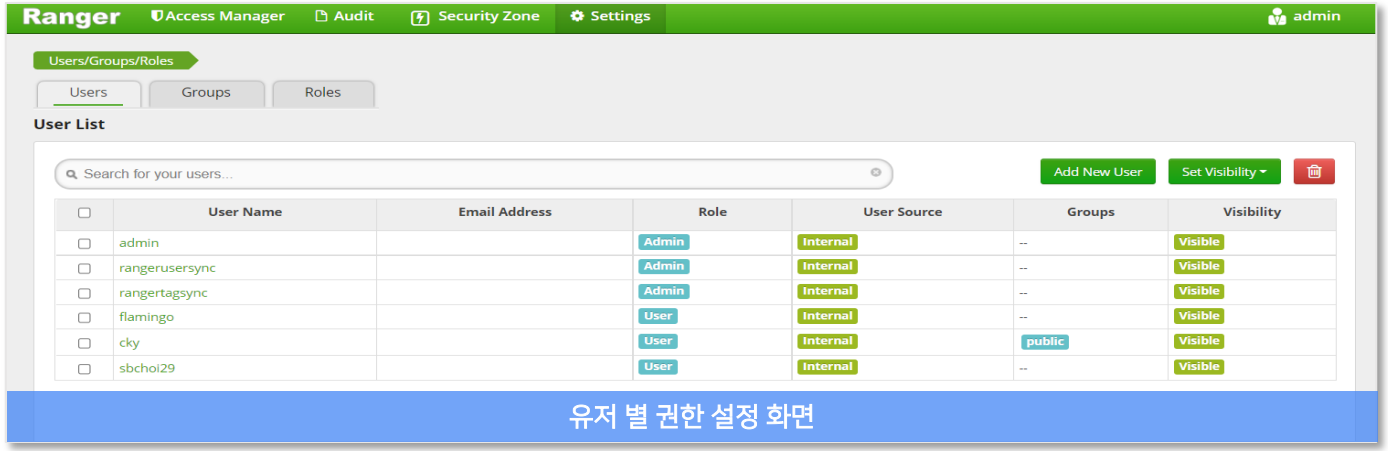
Hadoop Ecosystem 데이터 액세스를 위해, 유저 별 권한 관리 및 유저 관리가 가능합니다.

FEATURE

- 유저(사용자) 관리
- 사용자별 권한 관리
- 그룹관리
- Hadoop Ecosystem 데이터 액세스 권한 관리



유저 관리 화면



유저 별 권한 설정 화면

© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

Hadoop Cluster 종합 감시를 위한 전용 화면 제공

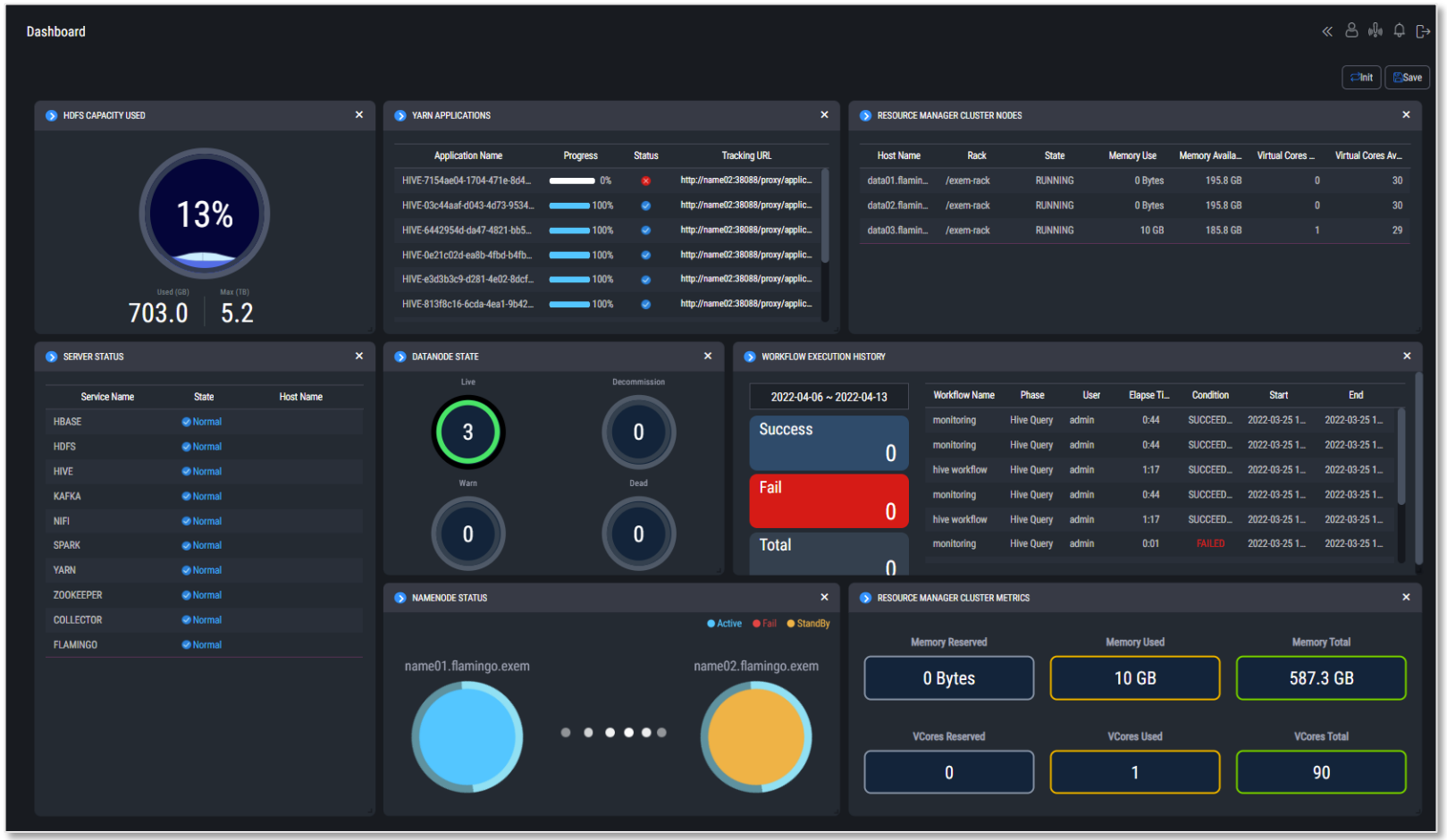
View Type

Dashboard

Dashboard View를 통해 Hadoop Cluster의 핵심 지표를 쉽고 직관적으로 파악할 수 있습니다.
(Resource Manager, NameNode, HDFS, Hive, YARN, Node Manager 등)

FEATURE

- 실시간 모니터링 정보 수집
- Hadoop Ecosystem 상태
- Workflow 실행 결과 및 상태
- 서비스가 설치되어 있는 Node 확인 가능
- 위치, 항목을 개인화 하여 조정



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

Hadoop Cluster Node 모니터링을 위한 전용 화면 제공

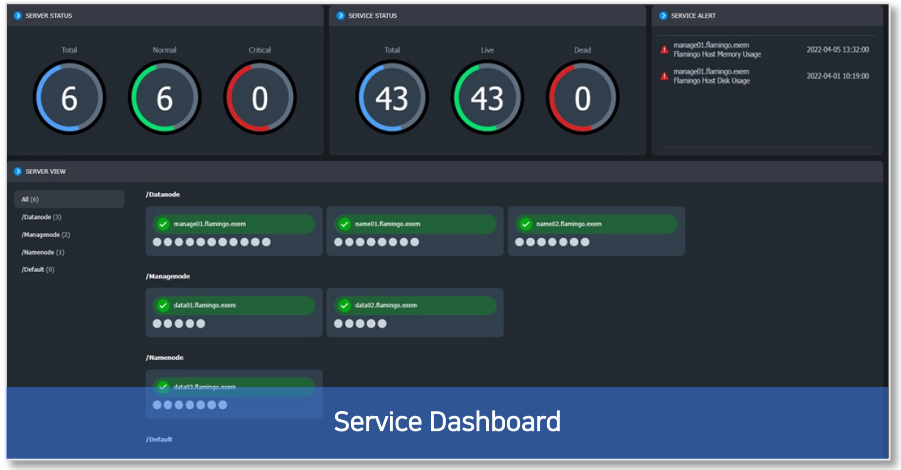
View Type

Service Dashboard

Hadoop Cluster Node를 시각화 하여 서버, 서비스, 노드 상태에 대한 모니터링을 가능하게 합니다.
또한, Hadoop Cluster Node를 시각적으로 입체화한 3D Dashboard도 지원합니다.

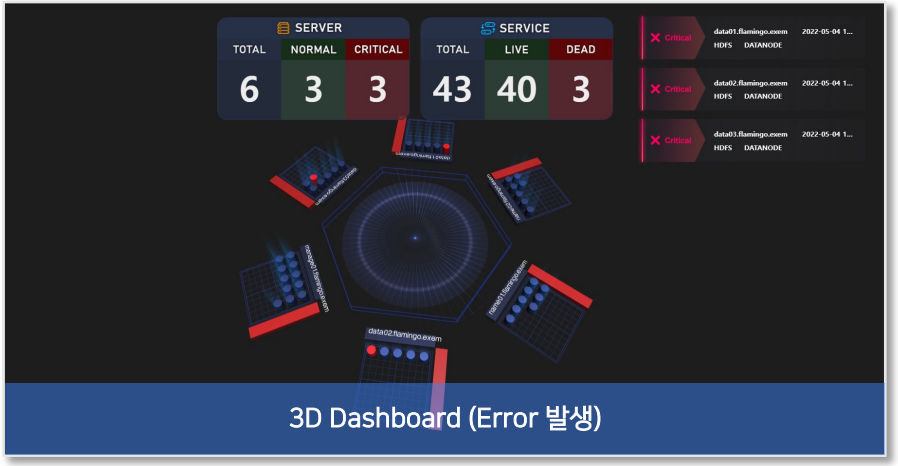
FEATURE

- 실시간 모니터링 정보 수집
- 서버, 서비스 상태
- 노드 별 서비스 상태
- 임계치 상태
- 상태 정보 Alert



Service Dashboard

3D Dashboard (서버별 상세)



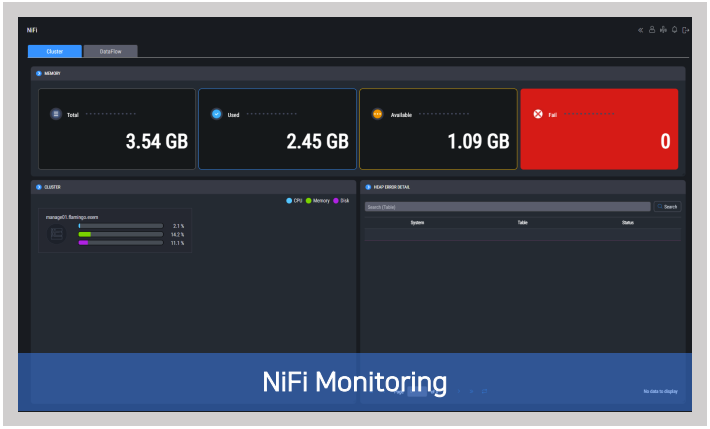
3D Dashboard (Error 발생)

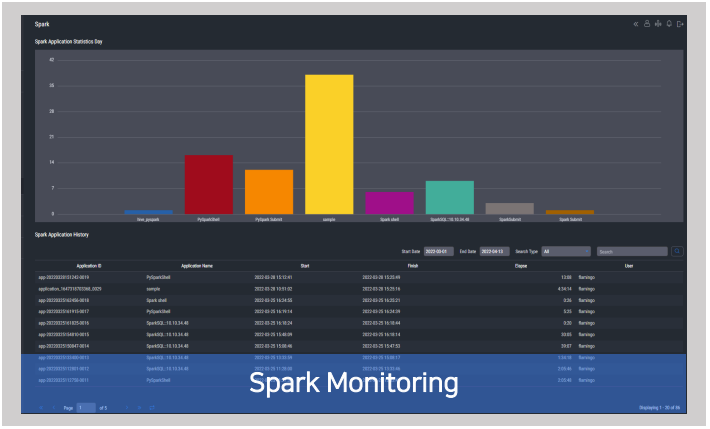
3D Dashboard (정상 동작)

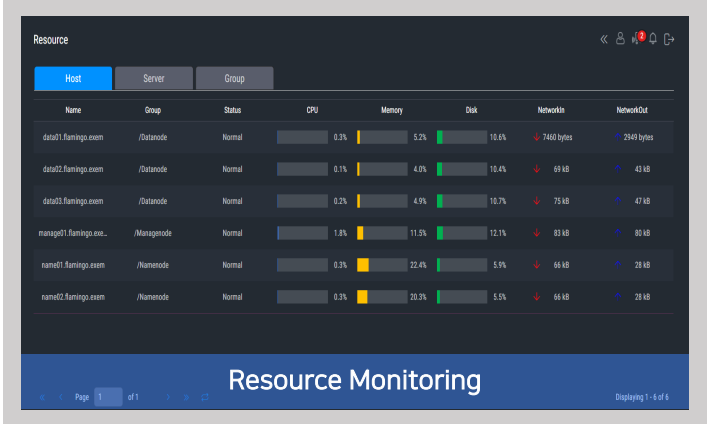
© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

Hadoop Cluster의 성능 관리에 최적화된 Monitoring 화면

HDFS / Yarn / NiFi / Spark / Resource







© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

HDFS 핵심 지표 모니터링을 위한 전용 화면 제공

View Type

HDFS Monitoring

분산 파일 시스템을 관리하는 NameNode의 가장 중요한 성능 지표 (JVM Heap, GC, Active NameNode, HDFS Capacity, Block Status)를 쉽게 파악할 수 있습니다.

Feature

- Collector를 활용하여 Hadoop의 데이터를 안정적으로 수집
- NameNode HA시 Active NameNode에 상태 표시
- Node 상태를 확인 할 수 있는 기능 제공
- Block 상태 모니터링을 통해 운영자가 HDFS 장애에 대응
- NameNode GC(garbage collector), JVM 상태 확인
- DataNode 여유 용량 확인

© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

HDFS 브라우저의 사용자 행위에 대한 **세부 모니터링 화면 제공**

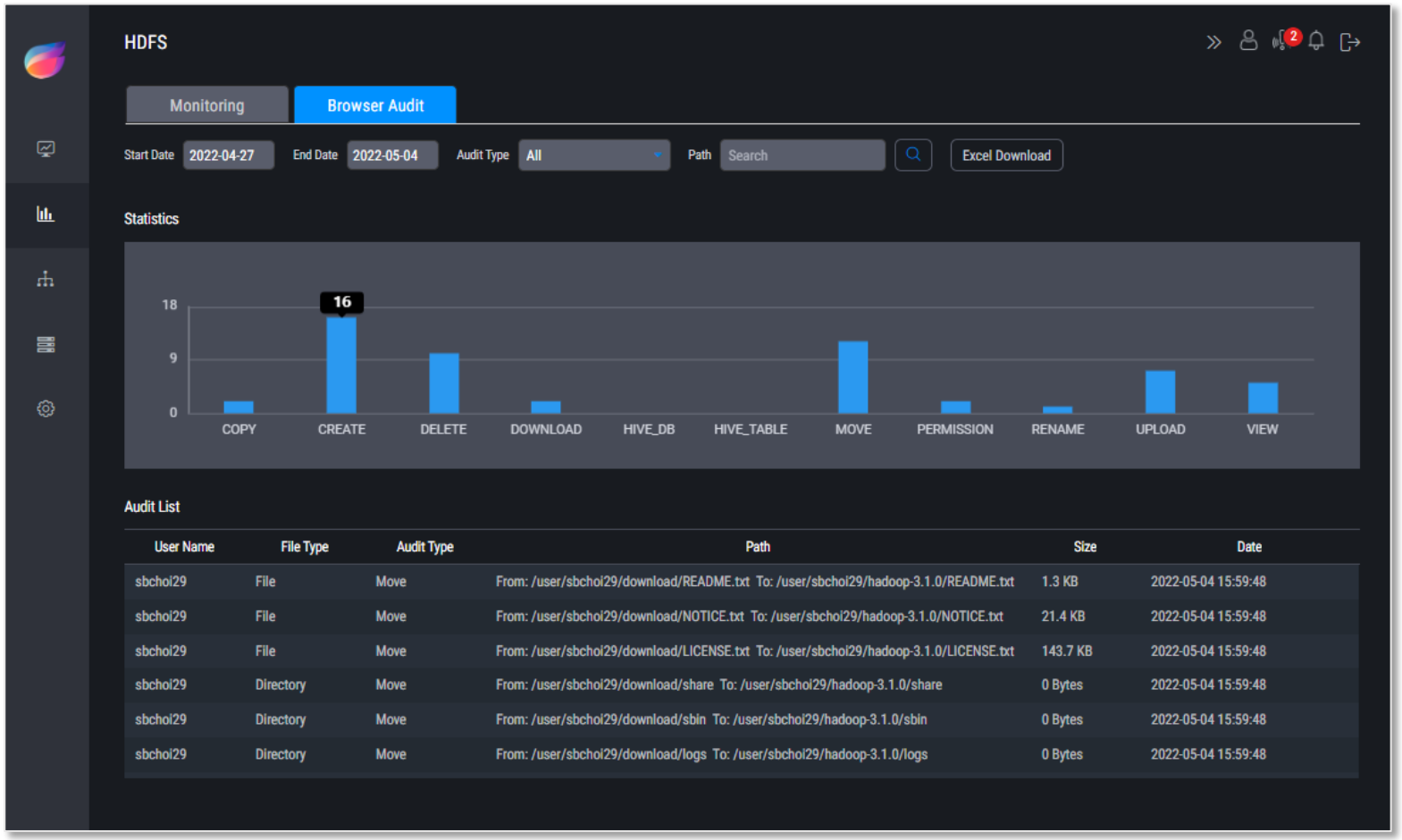
View Type

HDFS Browser Audit

HDFS Browser를 통하여 작업한 이력을 추적할 수 있습니다.

Feature

- HDFS 브라우저 사용 통계
- 주요 유형별 사용빈도 차트
- 사용자 별, 액션 별 현황
- 사용 목록 Excel 다운로드



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

Yarn을 위한 Summary, Log 등의 핵심 정보에 대한 전용 화면 제공

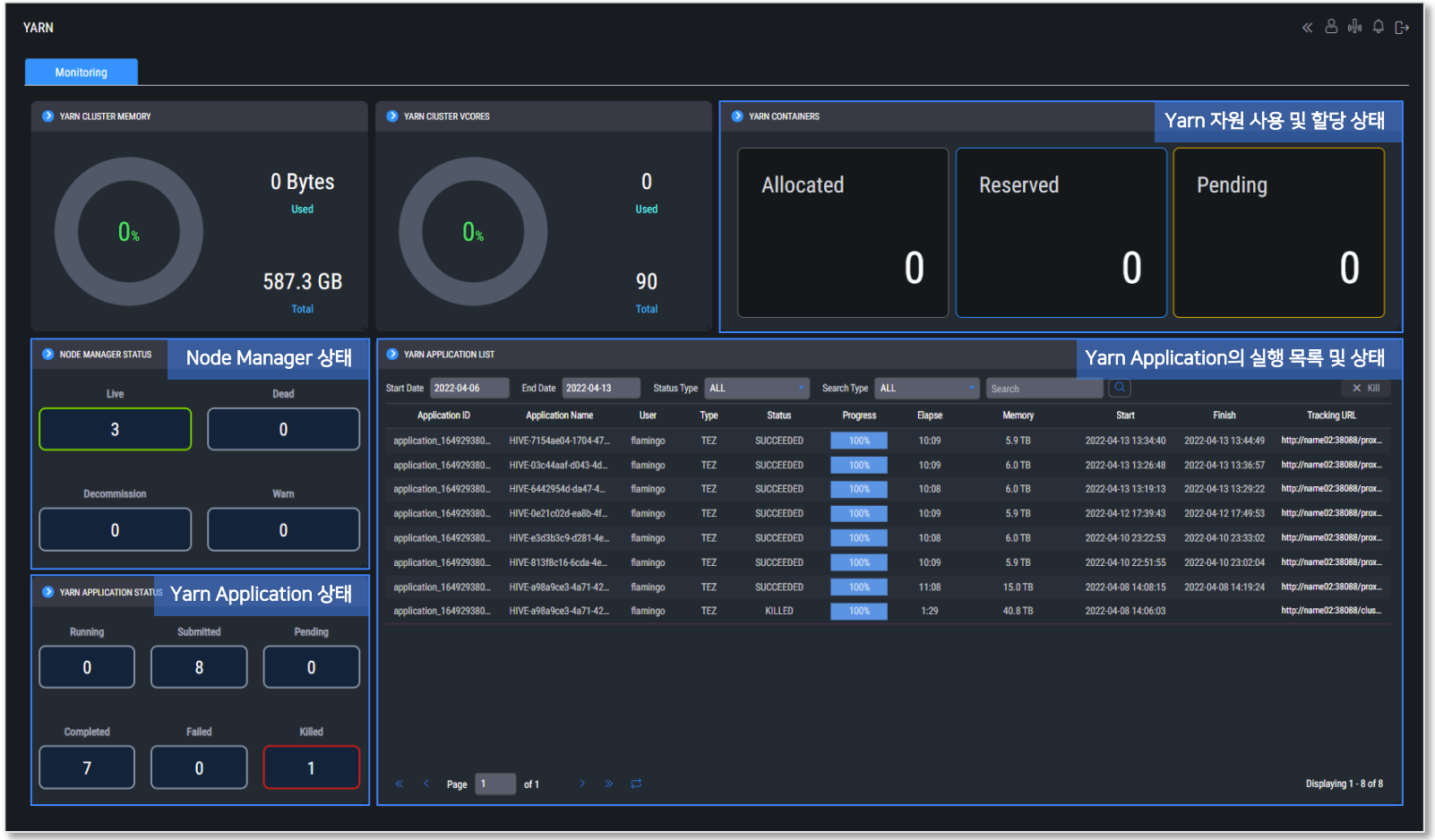
View Type

Yarn Monitoring

Yarn Cluster 사용량 확인 및 실행 이력을 표시하여 관리자들이 애플리케이션을 손쉽게 모니터링 할 수 있습니다.

Feature

- Yarn Cluster 사용량 확인
- 실행 중/지연 중인 Yarn Application Progress bar
- 실행 중 및 완료된 Yarn 애플리케이션 목록 및 각종 정보
- Yarn 애플리케이션의 실행 Tracking
- Node Manager, Yarn Application 상태
- Yarn Application 강제 종료



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

데이터 수집(Ni-Fi) 모니터링으로 안정적인 데이터 확보

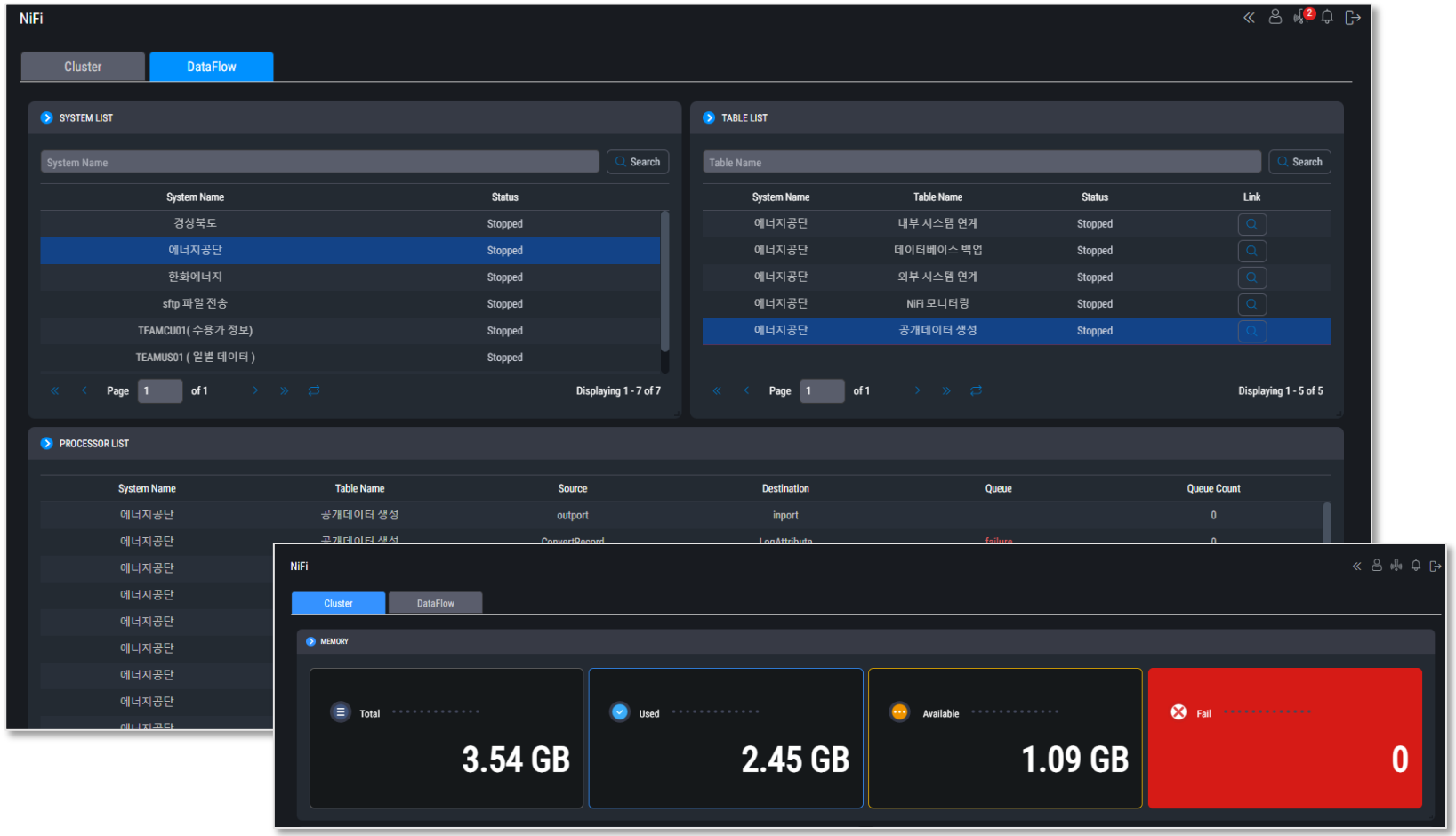
View Type

NiFi Monitoring

NiFi에서 구성한 Workflow의 장애 상황을 쉽게 확인할 수 있도록 모니터링 기능을 제공하여 안정적인 데이터 확보가 가능하도록 지원합니다.
NiFi Cluster를 구성하는 노드 자원 및 상태 정보, DataFlow 의 요약 및 상태 정보 기능을 제공합니다.

Feature

- 데이터 수집 서버 Heap 사용 지표 시각화
- Workflow 장애 모니터링 및 원인 파악
- NiFi Cluster Memory 사용량 제공
- 장애 Workflow와 NiFi 구성 화면 direct 이동 기능
- NiFi DataFlow의 시스템, 테이블 별 프로세서 리스트 확인 가능



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

Spark Application 정보 제공

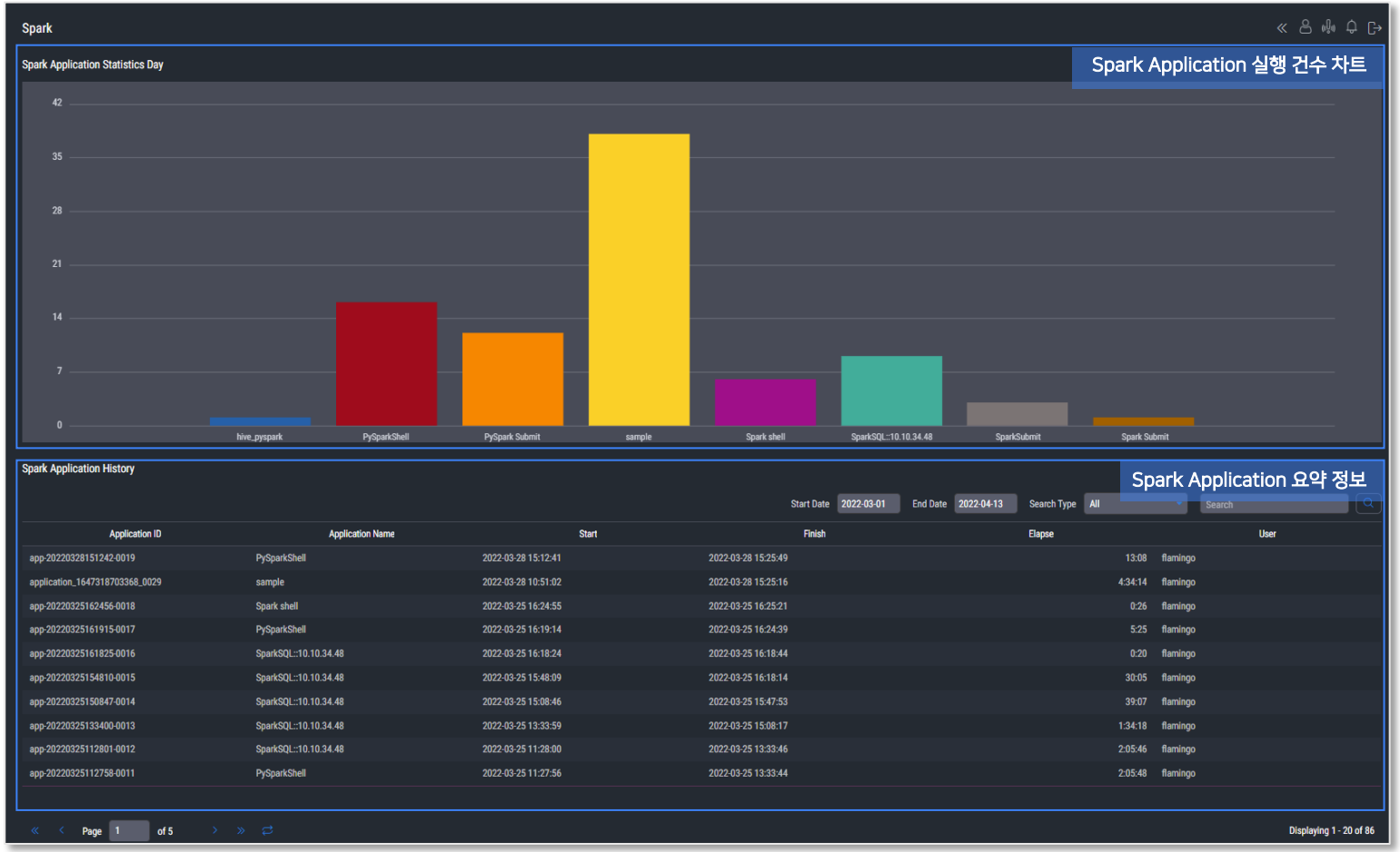
View Type

Spark Monitoring

Spark Application을 수행한 모든 애플리케이션 현황을 보여주며, 애플리케이션 통계 및 각 애플리케이션별 상세 설정 정보, Event Timeline 차트를 보여줍니다.

Feature

- Spark Application 실행 건수 조회 기능
- 특정 기간 Application 현황 조회 기능
- Application별 상세 설정 정보 조회



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

주요 기능 – Resource Monitoring

Hadoop Cluster 서버의 시스템 **자원 모니터링** 전용 화면 제공

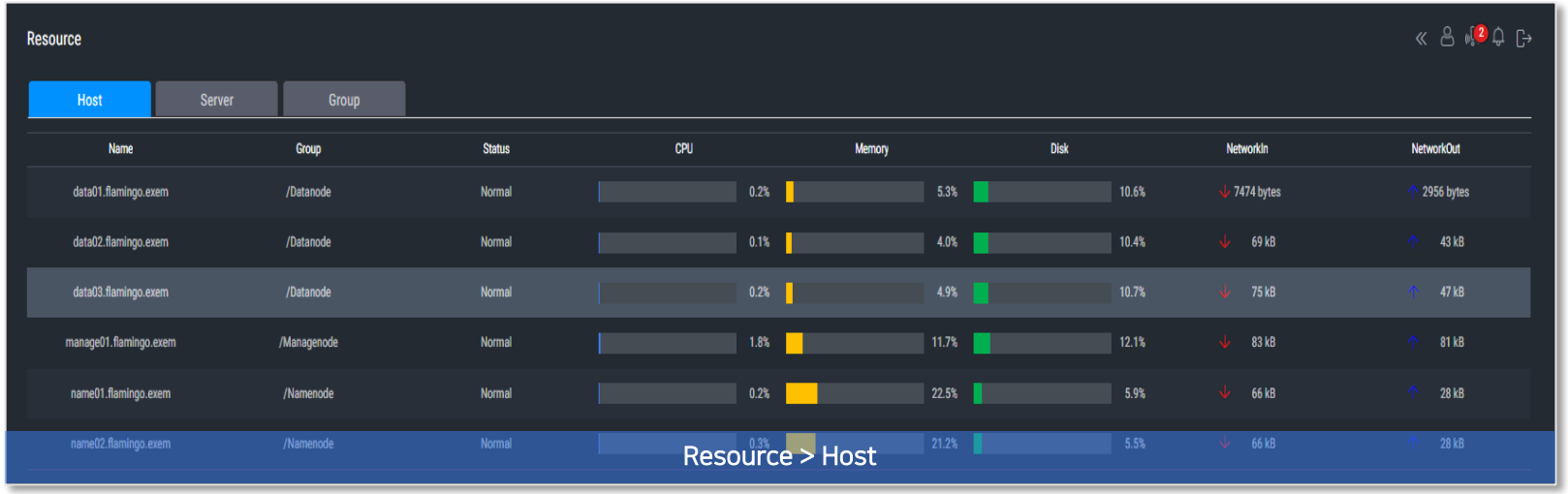
View Type

Resource Monitoring

Hadoop Cluster를 구성하는 각 노드의 각종 자원 사용량을 모니터링하는 전용 화면으로 **호스트, 서버, 그룹별**로 제공합니다.

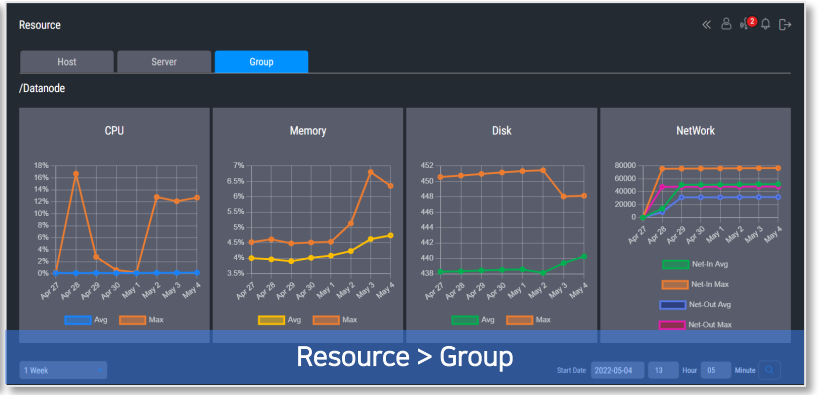
Feature

- Hadoop Cluster 자원 모니터링
- Core, Memory, Disk 사용량 모니터링
- 서버 Health Check로 손쉽게 상태 확인



Resource > Host

Resource > Server



Resource > Group

© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

주요 기능 – Workflow Designer

다양한 데이터 분석 플로우를 손쉽게 구성할 수 있는 **워크플로우 디자이너** 제공

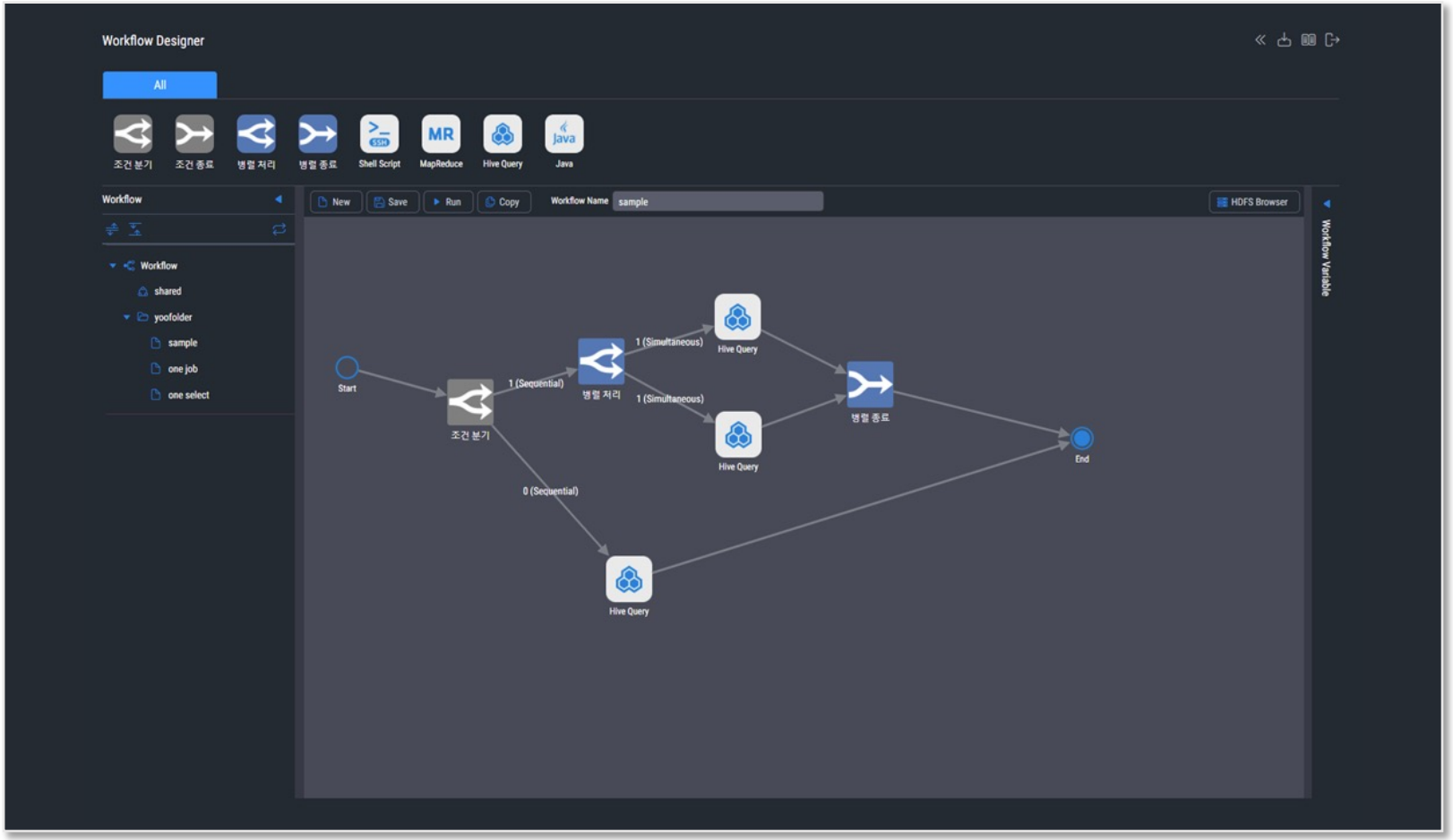
View Type

Workflow Designer

지속적인 데이터 처리가 가능한 Workflow를 제공하여 데이터의 품질을 향상시킬 수 있도록 도와드립니다.

Feature

- Drag & Drop 방식 지원
- 워크플로우를 UI 기반으로 빠르게 드로잉 및 실행
- 지속적인 데이터 처리를 위한 모듈(Java, shell, MR, hive Query 등)
  탑재를 통해 빠르게 MapReduce 방식으로 Flow를 구성하여 실행할 수 있으며,
  지속적인 처리를 위한 스케줄링 기능 제공



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

주요 기능 – Service

Hadoop Cluster의 서비스들을 관리하기 위한 관리기능 제공

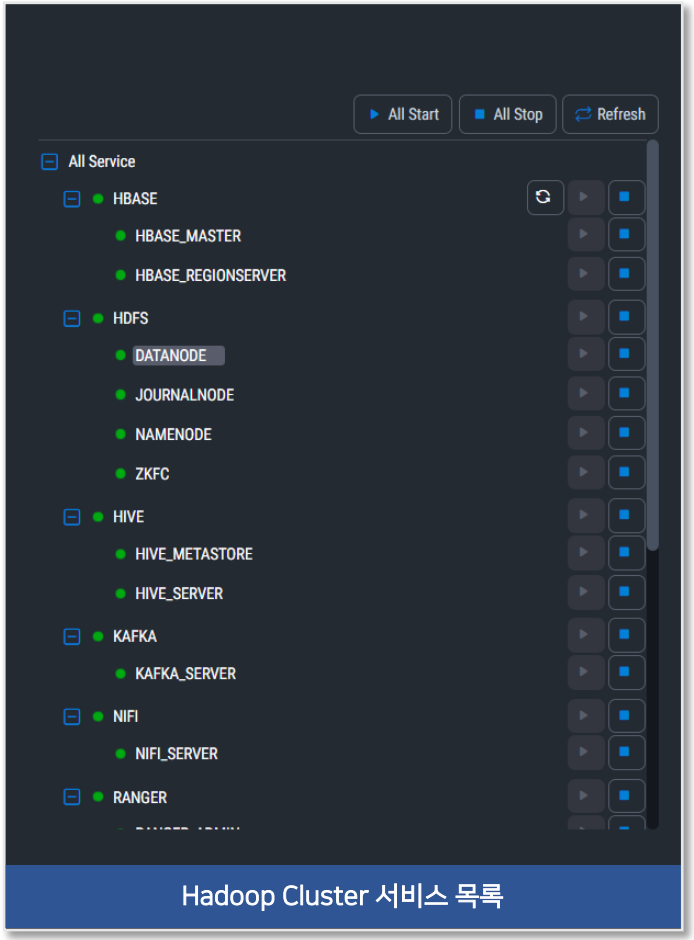
View Type

Service

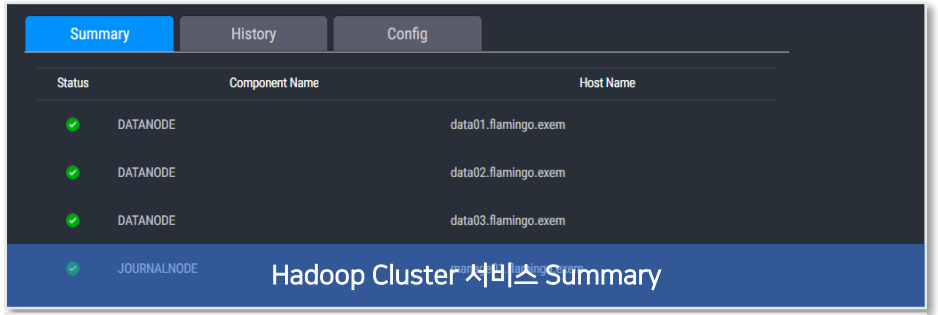
클러스터에 설치된 여러 서비스의 관리 페이지로 이동해 각 서비스에 존재하는 컴포넌트를 관리 (실행, 중지, Config값 수정 등) 할 수 있습니다.

Feature

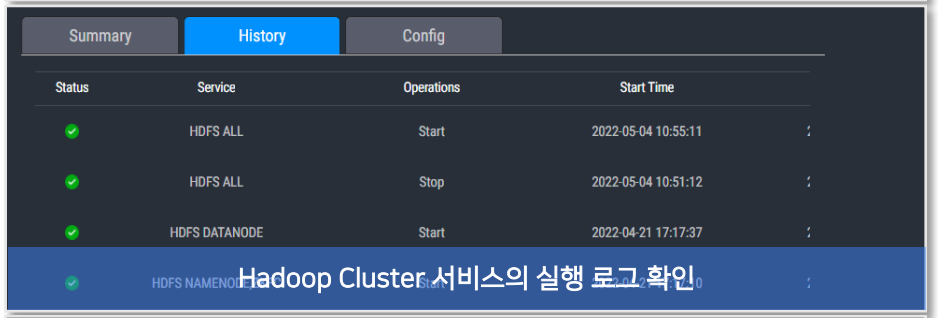
- Hadoop Ecosystem 서비스 실행 상태
- Hadoop Ecosystem 실행/중지
- Hadoop Ecosystem 실행/중지 이력 확인
- Hadoop Ecosystem 실행/중지 상세 로그 확인
- Hadoop Ecosystem Config 수정 및 배포



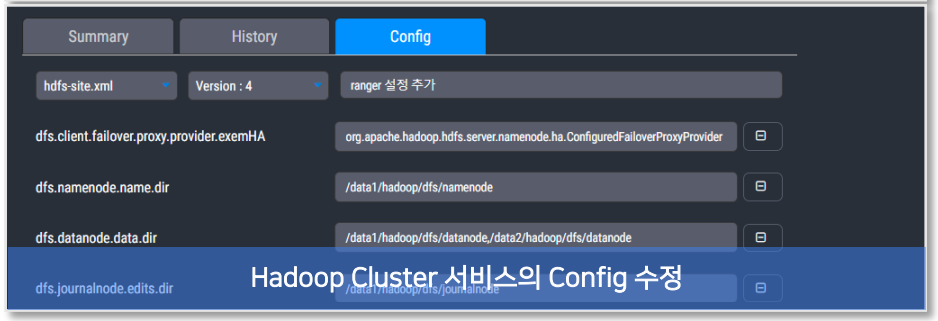
Hadoop Cluster 서비스 목록



Hadoop Cluster 서비스 Summary



Hadoop Cluster 서비스의 실행 로그 확인



Hadoop Cluster 서비스의 Config 수정

© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

HDFS의 파일 및 디렉토리를 관리하기 위한 모든 관리 기능 제공

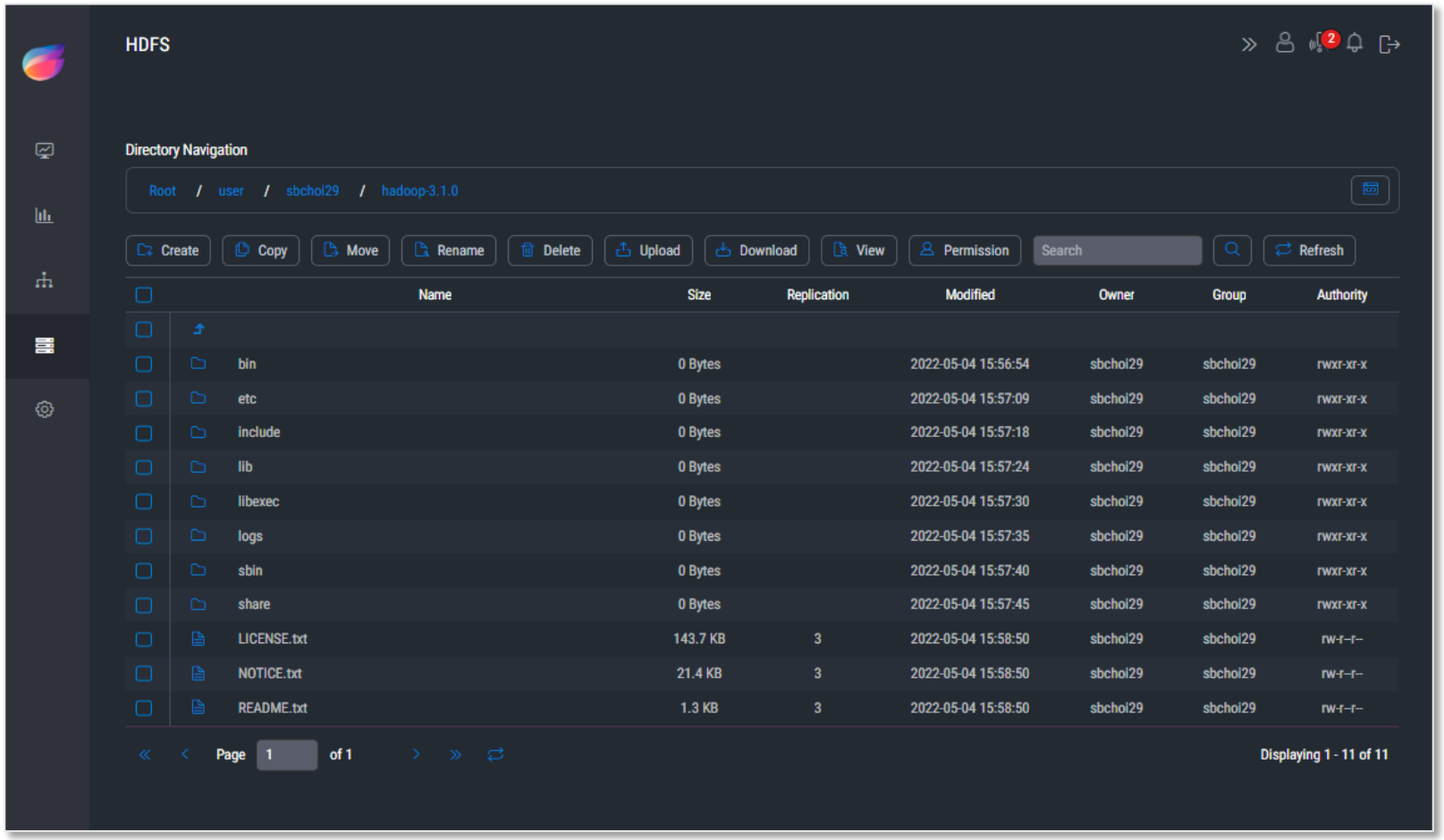
View Type

HDFS Browser

HDFS 브라우저는 HDFS상의 디렉토리 및 파일을 관리하는 기능으로 파일 업로드와 다운로드가 가능하며, 대량의 파일도 표시가 가능합니다.

Feature

- HDFS의 파일 및 디렉토리 관리
- 디렉토리 및 파일의 권한 관리
- 파일 내용 보기 및 파일 업로드/다운로드
- 문자열 기반 디렉토리 경로 입력 시 해당 디렉토리로 이동
- 디렉토리 별 사용량 표시
- HDFS내의 파일 수정



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

한 화면에 구성된 활용성 높은 Hive Editor

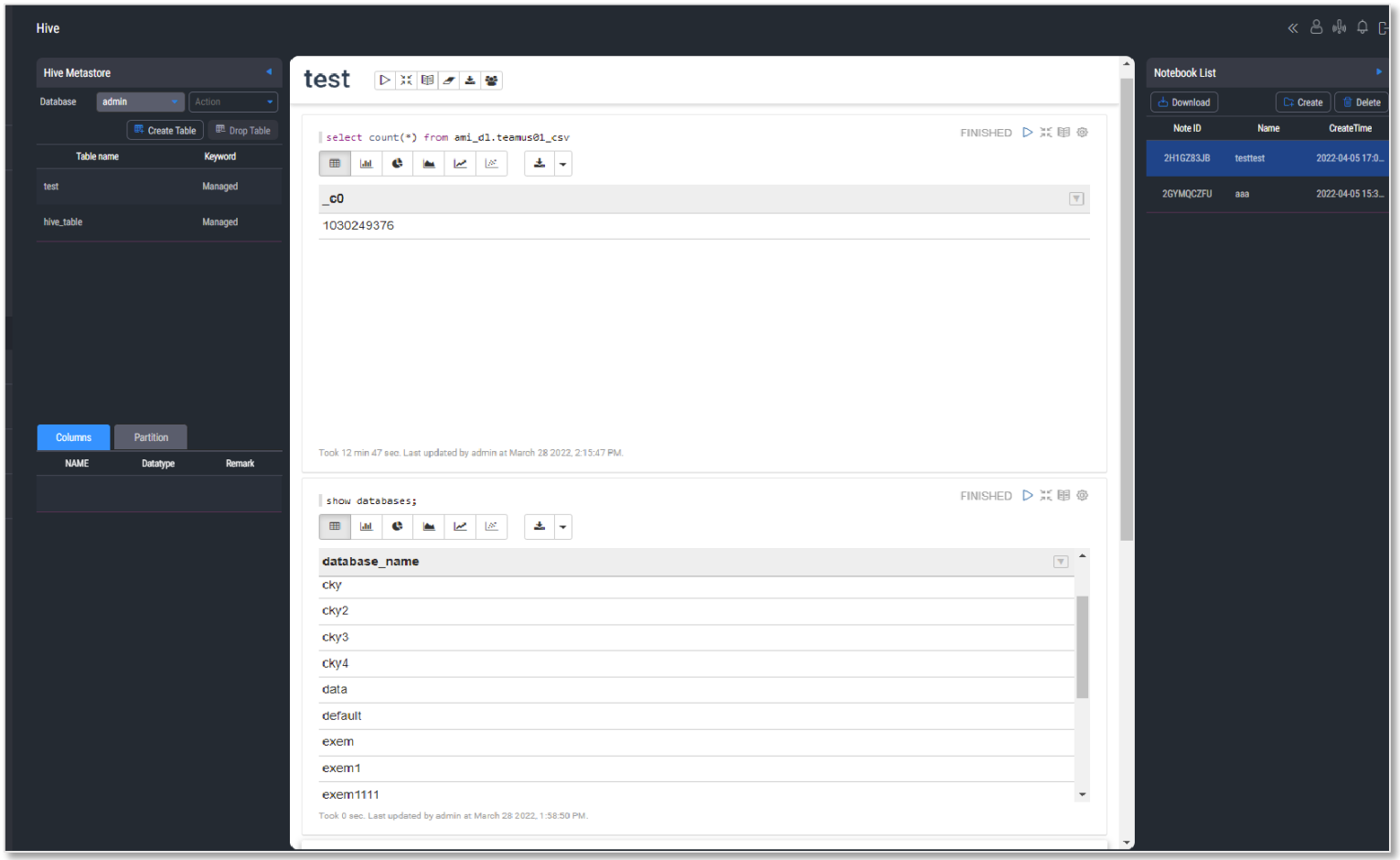
View Type

Hive Workpace

데이터 관리를 위하여 Hive Metastore를 제공합니다. Zeppelin을 활용하여 Hadoop의 데이터를 RDBMS처럼 쉽게 활용할 수 있습니다.

Feature

- Zeppelin을 활용한 손쉬운 Hive 활용
- Query History가 가능한 Zeppelin Note
- Hive Metadata 관리기능을 이용한 테이블 및 데이터베이스 관리
- 조회 데이터 시각화 제공



© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

Flamingo 설정 화면 제공

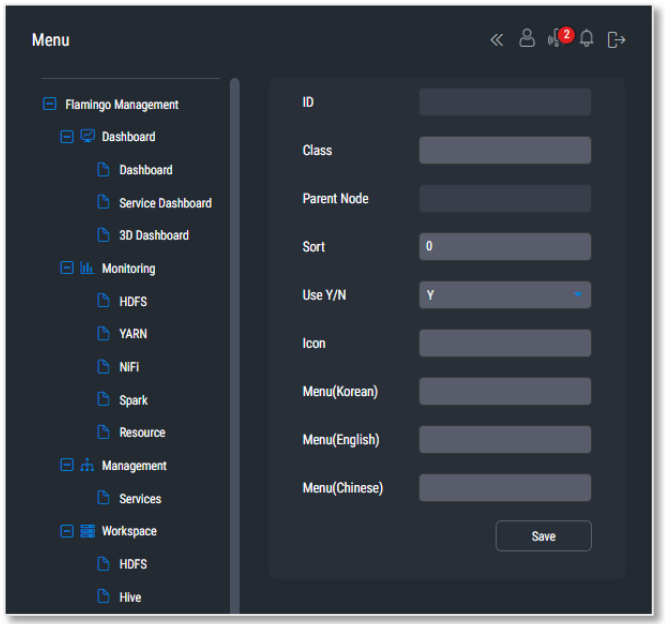
View Type

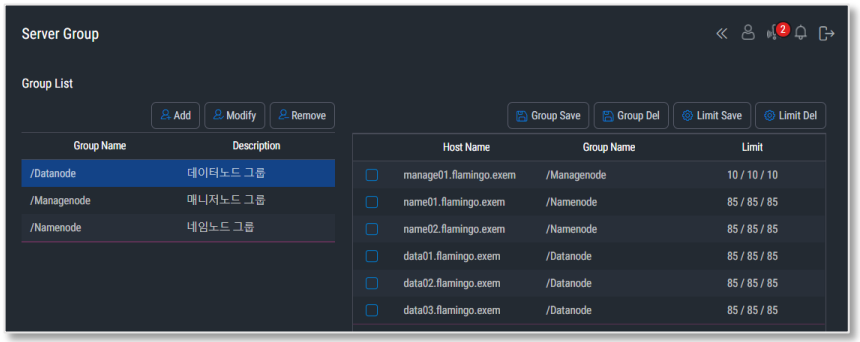
Settings

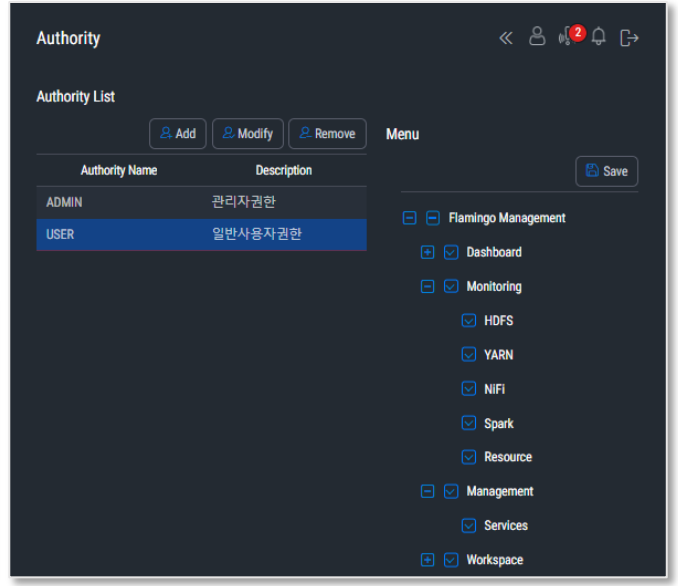
메뉴 관리, 사용자 관리, 권한 관리, 서버 그룹 관리 기능을 제공합니다.

Feature

- 메뉴 사용 여부, 정렬 설정
- 유저(사용자) 관리
- 사용자 권한 관리
- 그룹별 호스트 관리
- 서버별 임계치 설정







© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

주요 기능 – Quick Icon

빠른 정보 취득을 위한 퀵 아이콘 제공

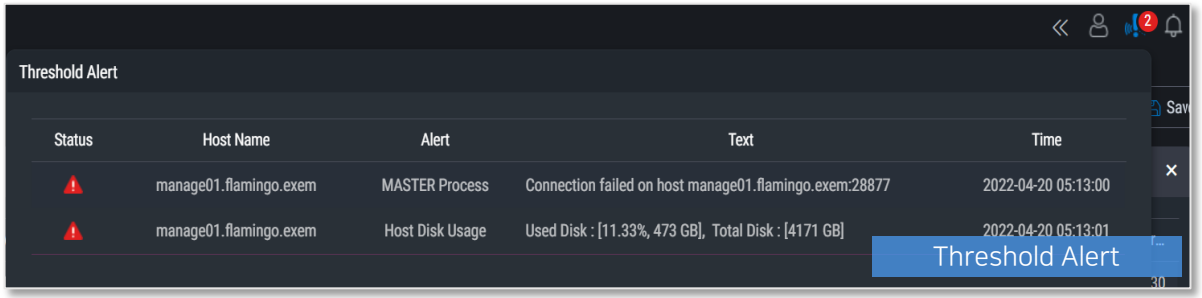
View Type

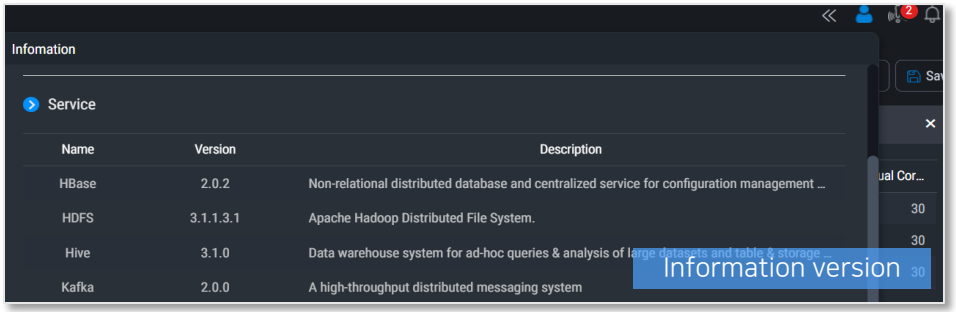
Quick Icon

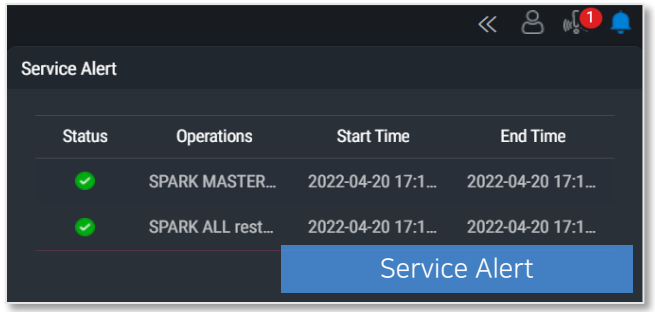
상단 퀵 아이콘을 통해 손쉽게 정보를 취득할 수 있습니다.

Feature

- 플라밍고 라이선스 정보 확인
- 각 서비스의 버전 정보 확인
- 임계치 상태 알림 확인
- 서비스 상태 알림 확인







© Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.

제품 스펙

- Web Browser

브라우저 : Edge, Chrome, Firefox, Safari

해상도 : 1920 x 1080 (권장) / 1440 x 900 (최소)

- Server

OS : Linux (CentOS, Ubuntu)

Hadoop : Apache Hadoop 2.0, 3.0 지원

DB : PostgreSQL 9.4 이상

CPU : 8Core (최소)

RAM : 64GB 이상 (Hadoop Node에 따라 변경)

DISK : Hadoop Node에 따라 변경

JAVA : 1.8 버전 이상

- Flamingo Agent **(Hadoop Node 개수만큼 필요)**

OS : Linux (CentOS, Ubuntu)

Hadoop : Apache Hadoop 2.0, 3.0 지원

CPU : 4Core (권장)

RAM : 8GB 이상

DISK : 100GB

JAVA : 1.8 버전

- Flamingo Collector / Flamingo Web

OS : Linux (CentOS, Ubuntu)

Hadoop : Apache Hadoop 2.0, 3.0 지원

DB : PostgreSQL 9.4 이상

CPU : 4Core (권장)

RAM : 8GB 이상

DISK : 100GB

JAVA : 1.8 버전

exem © Copyright 2001~2021 EXEM CO., LTD. All Rights Reserved.