

수집 파일럿 실행

» 수집 아키텍처

➤ 수집 요구 사항

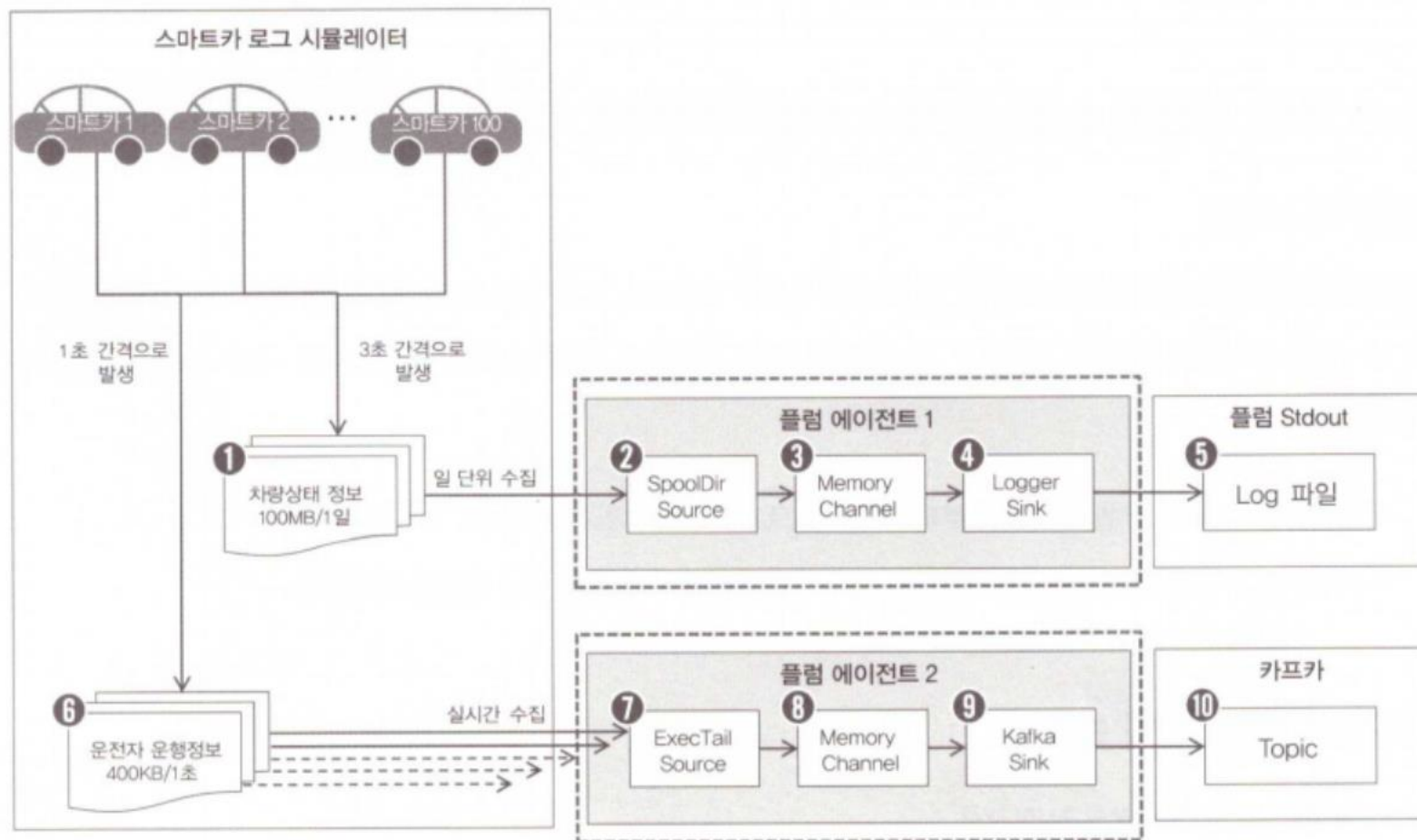
- 요구 사항1: 차량의 다양한 장치로부터 발생하는 로그 파일을 수집해서 **기능별 상태 점검**.
- 요구 사항2: 운전자의 운행 정보가 담긴 로그를 실시간으로 수집해서 **주행 패턴을 분석**.

» 수집 아키텍처

➤ 요구사항 구체화 및 분석

수집 요구사항 구체화	분석 및 해결 방안
1. 스마트카로부터 로그 파일들이 주기적으로 발생한다.	플럼을 이용해 대용량 배치 파일 및 실시간 로그 파일을 수집
2. 스마트카의 배치 로그 파일 이벤트를 감지해야 한다.	플럼의 Source 컴포넌트 중 SpoolDir를 이용해 주기적인 로그 파일 발생 이벤트를 감지
3. 스마트카의 실시간 로그 발생 이벤트를 감지해야 한다.	플럼의 Source 컴포넌트 중 Exec-Tail을 이용해 특정 로그 파일에서 로그 생성 이벤트를 감지
4. 스마트카가 만들어내는 로그 데이터에 가비지 데이터가 있을 수 있다.	플럼의 Interceptor를 이용해 정상 패턴의 데이터만 필터링
5. 수집 도중 장애가 발생해도 데이터를 안전하게 보관 및 재처리할 수 있어야 한다.	플럼의 메모리 Channel 및 카프카 Broker 사용으로 로컬 디스크의 파일시스템에 수집 데이터를 임시 저장
6. 스마트카의 실시간 로그 파일은 비동기 처리로 빠른 수집 처리를 해야 한다.	플럼에서 수집한 데이터를 카프카 Sink 컴포넌트를 이용해 카프카 Topic에 비동기 전송

» 수집 아키텍처



» 수집 아키텍처

- 로그 시뮬레이터 : 스마트카의 상태 정보와 운전자의 운행 정보 로그를 가상으로 만드는 자바 로그 발생기.
 - 스마트카 상태 정보 : 100명의 스마트가 장치들의 상태 정보를 3초 간격으로 발생시키며, 1일 100MB의 로그 파일이 생성.
 - 스마트가 운전자 운행 정보 : 100명의 스마트카 운전자들의 운행 정보를 실시간으로 발생시키며, 발생된 하나의 운행 정보 로그는 4KB 미만. 동시에 최대 400KB 용량으로 실시간 데이터 발생.
- 플럼 에이전트 1 : 스마트가 상태 정보를 기록한 로그 파일을 일별로 수집하기 위한 배치성 플럼 에이전트.
 - SpoolDir Source : 약속된 로그 발생 디렉토리를 모니터링하다가 정의된 로그 파일 발생 시 해당 파일의 내용을 읽어서 수집하는 기능을 제공.
 - Memory Channel : SpoolDir Source로 부터 수집된 데이터를 메모리 Channel에 중간 적재. 버퍼링 기능을 제공하며, Sink와 연결되어 트랜잭션 처리를 지원함.
 - Logger Sink : Channel로 부터 읽어들이는 데이터를 플럼의 표준 로그 파일로 출력.
- 플럼 에이전트 2 : 스마트가 운전자의 운행 정보를 실시간으로 수집하기 위한 실시간성 플럼 에이전트.
 - Exec-Tail Source : 로그가 쌓이고 있는 파일에 Tail 파이프라인을 이용해 실시간으로 데이터를 수집하는 기능
 - Memory Channel : Exec-Tail Source로 부터 수집된 데이터를 메모리 Channel 버퍼링 처리를 하면서 임시 적재.
 - Kafka Sink : Channel로부터 읽어들이는 데이터를 카프카 broker의 특정 토픽에 비동기 방식으로 전송하는 provider 역할을 수행.
- etc
 - Flume Stdout : 플럼의 Logger-Sink를 통해 표준 출력 로그가 출력.
 - Kafka Topic : 플럼의 Kafka-Sink는 수집된 실시간 로그를 임시 적재.

CM을 이용한 Flume / Kafka 설치

Flume 설치

The screenshot shows the Cloudera Manager web interface. The browser address bar indicates the URL is `server01.hadoop.com:7180/cmf/home`. The main navigation bar includes links for '홈' (Home), '상태' (Status), '모든 상태 문제' (All Status Problems), '구성' (Configuration), and '모든 최근 명령' (All Recent Commands). A 'Switch to Table View' button and a '추가' (Add) button are also present.

A yellow banner at the top contains a warning message: '내장된 PostgreSQL 데이터베이스를 사용하는 바-프로덕션 모드에서 Cloudera Manager를 실행 중입니다. 프로덕션으로 이동하기 전에 지원되는 외부 데이터베이스를 사용하여 전환하십시오. 자세한 세부 정보' (Running Cloudera Manager in a bar-production mode using the built-in PostgreSQL database. Before moving to production, use supported external databases to migrate. See detailed information). Below this, a message states: 'Service Monitor 요청에 실패했습니다. 이로 인해 페이지 응답이 느려질 수 있습니다. Service Monitor 상태를 확인하십시오.' (Failed to request Service Monitor. This may slow down page response. Check Service Monitor status).

The left sidebar shows 'Cluster 1' selected. A dropdown menu is open for '서비스 추가' (Add Service), with 'Add Hosts' highlighted. The menu options include: 시작 (Start), 중지 (Stop), 재시작 (Restart), 롤링 재시작 (Rolling Restart), 클라이언트 구성 배포 (Deploy Client Configuration), Kerberos 클라이언트 구성 배포 (Deploy Kerberos Client Configuration), 클러스터 업그레이드 (Upgrade Cluster), 클러스터 새로고침 (Refresh Cluster), 동적 리소스 풀 새로고침 (Refresh Dynamic Resource Pool), Inspect Hosts in Cluster, Kerberos 설정 (Configure Kerberos), Delete Kerberos Credentials, Configure the Cluster for Auto-TLS, HDFS 데이터 저장 암호화 설정 (Configure HDFS Data Storage Encryption), 클라이언트 구성 URL 보기 (View Client Configuration URL), 클러스터 이름 바꾸기 (Rename Cluster), and 유지관리 모드 진입 (Enter Maintenance Mode).

The main content area is titled '차트' (Charts) and shows a warning message: '쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다.' (Cannot execute query. Host Monitor is not running). Below the warning are three charts: '클러스터 CPU', '클러스터 디스크 IO', and '클러스터 네트워크 IO', all showing '쿼리 오류' (Query Error).

The bottom status bar shows the URL `server01.hadoop.com:7180/cmf/clusters/1/add-service/index`.

Flume 설치

Cluster 1에 서비스 추가 - Cloud x +

← → ↻ 주의 요함 | server01.hadoop.com:7180/cmf/clusters/1/add-service/index

Cloudera Manager

Cluster 1에 서비스 추가

추가할 서비스 유형을 선택합니다.

서비스 유형	설명
<input type="radio"/> ADLS Connector	The ADLS Connector service provides key management for accessing ADLS Gen1 accounts and ADLS Gen2 containers from CDH services.
<input type="radio"/> Accumulo	The Apache Accumulo sorted, distributed key/value store is a robust, scalable, high performance data storage and retrieval system. This service only works with releases meant to run on top of CDH6.
<input checked="" type="radio"/> Flume	Flume은 대부분의 소스에서 데이터를 수집 및 집계하여 HDFS 같은 영구 스토리지로 저장합니다.
<input type="radio"/> HBase	Apache HBase는 대규모 데이터 세트에 일의의 실시간 읽기/쓰기 액세스를 제공합니다(HDFS와 ZooKeeper 필요).
<input type="radio"/> HDFS	Apache HDFS(Hadoop Distributed File System)는 Hadoop 애플리케이션이 사용하는 기본 스토리지 시스템입니다. HDFS는 데이터 블록에 대한 여러 개의 복제본을 생성하고 이들 클러스터 전반에 걸쳐 컴퓨팅 호스트에 배포하여 안정적이고 매우 빠른 계산을 지원합니다.
<input type="radio"/> Hive	Hive는 SQL과 유사한 언어인 HiveQL을 제공하는 데이터 웨어하우스 시스템입니다.
<input type="radio"/> Hue	Hue는 CDH(Cloudera Distribution Including Apache Hadoop)에서 작동하는 GUI(그래픽 사용자 인터페이스)입니다(HDFS, MapReduce, Hive 필요).
<input type="radio"/> Impala	Impala에서는 HDFS 및 HBase에 저장된 데이터에 대해 실시간 SQL 쿼리 인터페이스를 제공합니다. Impala에는 Hive 서비스가 필요하며 Hue와 Hive Metastore를 공유합니다.
<input type="radio"/> Isilon	EMC Isilon is a distributed filesystem.
<input type="radio"/> Java KeyStore KMS	The Hadoop Key Management Service with file-based Java KeyStore. Maintains a single copy of keys, using simple password-based protection. Requires CDH 6.0+. Not recommended for production use.
<input type="radio"/> Kafka	Apache Kafka is publish-subscribe messaging rethought as a distributed commit log.
<input type="radio"/> Key-Value Store Indexer	Key-Value Store Indexer는 HBase에 포함된 테이블 안의 데이터의 변경 사항을 수신 대기하고 Solr을 사용하여 인덱싱합니다.
<input type="radio"/> Kudu	Kudu is a true column store for the Hadoop ecosystem.
<input type="radio"/> Oozie	Oozie는 클러스터의 데이터 처리 작업을 관리하는 워크플로우 조정 서비스입니다.
<input type="radio"/> S3 Connector	The S3 Connector Service securely provides a single set of AWS credentials to Impala and Hue. This enables Hue administrators to browse the S3 filesystem and define Impala tables backed by S3 data authorized to that AWS identity, and also enables Impala users to query S3-backed tables without directly providing AWS credentials, subject to having the proper permissions defined via Sentry. The S3 Connector only supports the S3A protocol.
<input type="radio"/> Sentry	Sentry 서비스는 인증 정책 메타데이터를 저장하고 동시 클라이언트와 이 메타데이터에 대해 안전하게 보호된 액세스를 제공합니다.
<input type="radio"/> Solr	Solr은 HDFS에 저장된 데이터를 인덱싱 및 검색하는 벡터 서비스입니다.
<input type="radio"/> Spark	Apache Spark is an open source cluster computing system. This service runs Spark as an application on YARN.
<input type="radio"/> Sqoop 1 Client	Configuration and connector management for Sqoop 1.
<input type="radio"/> YARN (MR2 Included)	YARN이라고도 하는 MRv2(Apache Hadoop MapReduce 2.0)는 MapReduce 애플리케이션을 지원하는 데이터 계산 프레임워크입니다(HDFS 필요).
<input type="radio"/> ZooKeeper	Apache ZooKeeper는 구성 데이터를 유지하고 동기화하는 중앙 집중식 서비스입니다.

뒤로 계속

Flume 설치

Cluster 1에 Flume 서비스 추가 - x

← → ↺ 주의 요함 | server01.hadoop.com:7180/cmf/clusters/1/add-service/index?serviceType=FLUME#step=roleA...

Cloudera Manager

Cluster 1에 Flume 서비스 추가

Select Dependencies

역할 할당 사용자 지정

변경 내용 검토

명령 세부 정보

요약

역할 할당 사용자 지정

여기에서 새 서비스에 대한 역할 할당을 사용자 지정할 수 있지만 단일 호스트에 너무 많은 수의 역할을 할당하는 등 올바르게 않게 할당할 경우, 성능이 저하될 수 있습니다.

역할 할당을 호스트별로 볼 수도 있습니다. [호스트별로 보기](#)

Agent

호스트 선택

할당된 호스트가 너무 적습니다. 최소 1개입니다.

뒤로 계속

server01.hadoop.com:7180/cmf/clusters/1/add-service/index?serviceType=FLUME#

Flume 설치

Cluster 1에 Flume 서비스 추가 - x

← → ↺ 주의 요함 | server01.hadoop.com:7180/cmf/clusters/1/add-service/index?serviceType=FLUME#step=roleA...

Cloudera Manager

지원 admin

1개의 호스트가 선택됨

새 역할 또는 기존 역할에 대한 호스트를 선택하십시오. 호스트 목록은 유효하지 않은 호스트를 제거하도록 필터링되었습니다. 이러한 호스트에는 상태가 불량하거나, 다른 클러스터의 구성원이거나, 호환되지 않는 CDH 버전이 설치되어 있는 호스트가 포함됩니다.

호스트 이름(host01, host[01-10]), IP 주소 또는 랙을 입력하십시오. 검색

팁: 첫 번째 확인란을 클릭하고 Shift 키를 누른 채로 마지막 확인란을 클릭하여 범위를 선택합니다.

<input type="checkbox"/>	호스트 이름 ↑	IP 주소 ↓	랙 ↓	코어 ↓	물리적 메모리 ↓	기존 역할	추가된 역할
<input type="checkbox"/>	server01.hadoop.com	192.168.56.101	/default	1	3.4 GiB	B NN SNN C AP C ES C HM C SM JHS RM	
<input checked="" type="checkbox"/>	server02.hadoop.com	192.168.56.102	/default	1	2.9 GiB	DN NM S	A

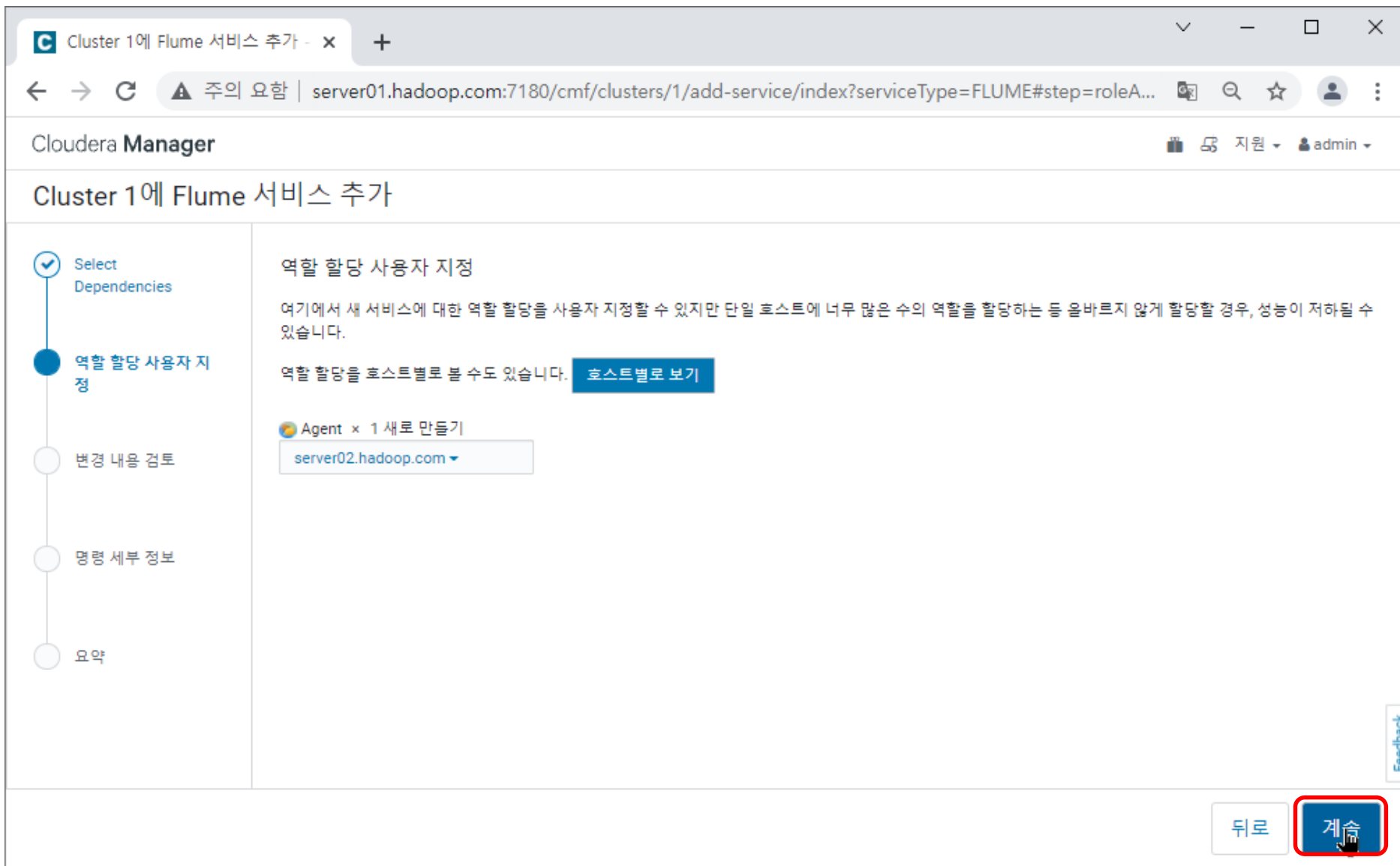
Displaying 1 - 2 of 2

취소 확인

뒤로

계속

Flume 설치



Flume 설치

Cluster 1에 Flume 서비스 추가 - x

Cluster 1에 Flume 서비스 추가 - x

server01.hadoop.com:7180/cmf/clusters/1/add-service/index?serviceType=FLUME#step=finish...

Cloudera Manager

Cluster 1에 Flume 서비스 추가

Select Dependencies

역할 할당 사용자 지정

변경 내용 검토

요약

요약

새 서비스가 클러스터에 설치 및 구성되었습니다.

참고: 여전히 새 서비스를 시작해야 할 수 있습니다. 시작하기 전에 오래된 구성이 포함된 모든 종속 서비스를 재시작하는 것이 좋습니다. 이러한 작업은 아래에서 완료를 클릭하면 주 페이지에서 수행할 수 있습니다.

뒤로

완료

Flume 설치

The screenshot shows the Cloudera Manager web interface in a browser window. The address bar displays 'server01.hadoop.com:7180/cmf/home'. The page header includes navigation tabs like '홈', '상태', and '모든 상태 문제'. A warning banner at the top indicates a PostgreSQL database issue. The main content area shows 'Cluster 1' with a list of services: CDH 6.3.2 (Parcel), Flume (highlighted with a red box and a mouse cursor), HDFS, YARN (MR2 I...), and ZooKeeper. To the right, a '차트' (Charts) section displays three monitoring charts: '클러스터 CPU', '클러스터 디스크 IO', and '클러스터 네트워크 IO', all showing '쿼리 오류' (Query Error). A red-bordered box highlights a message: '쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다.' (Cannot execute query. Host Monitor is not running).

상태 - 홈 - Cloudera Manager

server01.hadoop.com:7180/cmf/home

Cloudera Manager 클러스터 호스트 진단 감사 차트 관리

홈 상태 모든 상태 문제 구성 10 모든 최근 명령 Switch to Table View 추가

내장된 PostgreSQL 데이터베이스를 사용하는 비-프로덕션 모드에서 Cloudera Manager를 실행 중입니다. 프로덕션으로 이동하기 전에 지원되는 외부 데이터베이스를 사용하여 전환하십시오. 자세한 세부 정보

Service Monitor 요청에 실패했습니다. 이로 인해 페이지 응답이 느려질 수 있습니다. Service Monitor 상태를 확인하십시오.

Cluster 1

CDH 6.3.2 (Parcel)

Flume 2

Flume

HDFS 3

YARN (MR2 I...

ZooKeeper 1

Cloudera Management Service

Cloudera Man... 4

차트

30분 1시간 2시간 6시간 12시간 1일 7d 30d

쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다.

클러스터 CPU

쿼리 오류

클러스터 디스크 IO

쿼리 오류

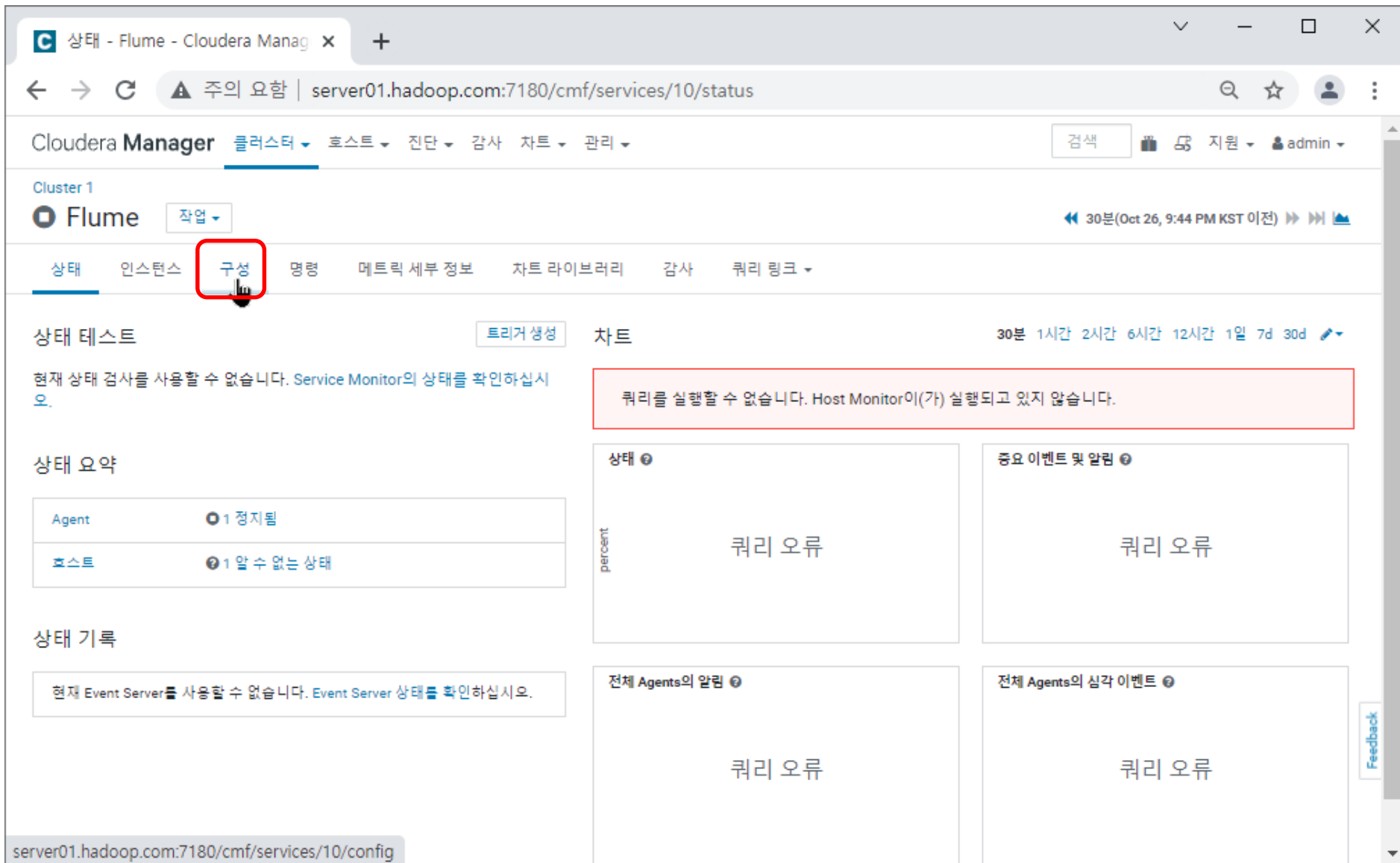
클러스터 네트워크 IO

쿼리 오류

Feedback

server01.hadoop.com:7180/cmf/services/10/status

Flume 설치



Flume 설치

구성 - Flume - Cloudera Manag x +

server01.hadoop.com:7180/cm/services/10/config#filterfreeText=java%20heap

Cloudera Manager 클러스터 호스트 진단 감사 차트 관리

Cluster 1

Flume 작업

Oct 26, 9:46 PM KST

상태 인스턴스 구성 명령 메트릭 세부 정보 차트 라이브러리 감사 쿼리 링크

java heap 역할 그룹

모든 설명 표시

필터

범위

Flume (서비스 자원)	0
Agent	3

범주

Flume-NG Solr 싱크	0
고급	2
기본	0
로그	0
리소스 관리	1
모니터링	0
보안	0
성능	0
스택 수집	0

메모리가 부족하면 힙 덤프 ☒ Agent Default Group

힙 덤프 디렉토리 oom_heap_dump_dir Agent Default Group

/tmp

Agent의 Java 힙 크기(단위: 바이트) Agent Default Group

100 MiB

25 페이지 기준

1 Edited Value 변경 이유

변경 내용 저장

Flume 설치

구성 - Flume - Cloudera Manag x +

← → ↺ 주의 요함 | server01.hadoop.com:7180/cm/services/10/config#filterfreeText=java%20heap

Cloudera Manager 클러스터 호스트 진단 감사 차트 관리

Cluster 1

Flume 작업 Oct 26, 9:48 PM KST

상태 인스턴스 구성 명령 메트릭 세부 정보 차트 라이브러리 감사 쿼리 링크

java heap 역할 그룹

모든 설명 표시

필터

범위

Flume (서비스 자원)	0
Agent	3

범주

Flume-NG Solr 링크	0
고급	2
기본	0
로그	0
리소스 관리	1
모니터링	0
보안	0
성능	0
스택 수집	0

메모리가 부족하면 힙 덤프 ☒ Agent Default Group

힙 덤프 디렉토리 oom_heap_dump_dir Agent Default Group

/tmp

Agent의 Java 힙 크기(단위: 바이트) Agent Default Group

100 MiB

25 페이지 기준

Feedback

변경 내용 저장

server01.hadoop.com:7180/cm/

Flume 설치

The screenshot shows the Cloudera Manager web interface. The browser tab is '상태 - 홈 - Cloudera Manager'. The address bar shows 'server01.hadoop.com:7180/cmf/home'. The page title is 'Cloudera Manager'. The navigation bar includes '홈', '상태', '모든 상태 문제', '구성 10', and '모든 최근 명령'. A warning message is displayed: '내장된 PostgreSQL 데이터베이스를 사용하는 비-프로덕션 모드에서 Cloudera Manager를 실행 중입니다. 프로덕션으로 이동하기 전에 지원되는 외부 데이터베이스를 사용하여 전환하십시오. 자세한 세부 정보'. Below this, a message states: 'Service Monitor 요청에 실패했습니다. 이로 인해 페이지 응답이 느려질 수 있습니다. Service Monitor 상태를 확인하십시오.'

The main content area shows 'Cluster 1' with a dropdown menu. The 'Flume' service is listed with a dropdown arrow highlighted by a red box. The dropdown menu is open, showing 'Flume개 작업' (Flume Actions) with options: '시작' (Start), '중지' (Stop), '재시작' (Restart), '롤링 재시작' (Rolling Restart), '인스턴스' (Instances), '구성' (Configuration), and '역할 인스턴스 추가' (Add Role Instance). The '시작' option is highlighted by a mouse cursor.

The right side of the page shows '차트' (Charts) with a time range of '30분'. There are three chart boxes: '클러스터 CPU', '클러스터 디스크 IO', and '클러스터 네트워크 IO'. Each box displays '쿼리 오류' (Query Error) and a message: '쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다.'

The bottom status bar shows the URL: 'server01.hadoop.com:7180/cmf/services/10/do?command=Start'.

Flume 설치

The screenshot shows the Cloudera Manager web interface in a browser window. The address bar displays 'server01.hadoop.com:7180/cmf/home'. The main navigation bar includes 'Cloudera Manager', '클러스터', '호스트', '진단', '감사', '차트', and '관리'. The left sidebar shows the '홈' (Home) tab selected, with a list of services for 'Cluster 1': CDH 6.3.2 (Parcel), 2개의 호스트 (2 hosts), Flume, HDFS, YARN (MR2 I...), and ZooKeeper. The main content area displays a confirmation dialog box titled '시작' (Start) with the question 'Flume 서비스에 시작 명령을 실행하시겠습니까?' (Do you want to run the start command for the Flume service?). The dialog has two buttons: '취소' (Cancel) and '시작' (Start). A mouse cursor is clicking the '시작' button. The background shows a 'Cluster 1' overview with sections for '클러스터 CPU', '클러스터 디스크 IO', and '클러스터 네트워크 IO', each displaying '쿼리 오류' (Query Error).

Cloudera Manager 클러스터 호스트 진단 감사 차트 관리

홈 상태 모든 상태 문제 구성

내장된 PostgreSQL 데이터베이스를 사용하는 비-제한 세부 정보

Service Monitor 요청에 실패했습니다. 이로 인해

Cluster 1

CDH 6.3.2 (Parcel)

2개의 호스트

Flume

HDFS

YARN (MR2 I...

ZooKeeper

Cloudera Management Service

Cloudera Man...

시작

Flume 서비스에 시작 명령을 실행하시겠습니까?

취소 시작

클러스터 CPU

쿼리 오류

클러스터 디스크 IO

쿼리 오류

클러스터 네트워크 IO

쿼리 오류

Flume 설치

The screenshot shows the Cloudera Manager web interface. A modal window titled "시작" (Start) is displayed in the center. The modal contains the following information:

- 상태** (Status): 완료됨 (Completed) with a green checkmark icon.
- 컨텍스트** (Context): Flume with a link icon.
- 시간** (Time): Oct 26, 9:56:05 PM.
- 지속 시간** (Duration): 25.11s.
- 메시지** (Message): Successfully started service.
- 진행률** (Progress): 1/1 단계가 완료되었습니다. (1/1 step completed).
- 옵션** (Options):
 - ☒ Show All Steps
 - ☐ Show Only Failed Steps
 - ☐ Show Only Running Steps
- 기록** (Log): A single entry with a green checkmark, text "서비스의 1 역할 시작" (Start 1 role of service), time "Oct 26, 9:56:05 PM", and duration "25.08s".
- 버튼** (Buttons): A blue button labeled "닫기" (Close) at the bottom right.

The background interface shows the Cloudera Manager navigation bar with links like "홈" (Home), "클러스터" (Cluster), "호스트" (Hosts), "진단" (Diagnosis), "감사" (Audit), "차트" (Charts), and "관리" (Management). The main content area is partially visible, showing "Cloudera Management Service" and "클러스터 네트워크 IO" (Cluster Network IO).

Kafka 설치

Cloudera Manager 클러스터 ▾ 호스트 ▾ 진단 ▾ 감사 ▾ 차트 ▾ 관리 ▾

홈 상태 모든 상태 문제 구성 10 모든 최근 명령 Switch to Table View 추가 ▾

내장된 PostgreSQL 데이터베이스를 사용하는 비-프로덕션 모드에서 Cloudera Manager를 실행 중입니다. 프로덕션으로 이동하기 전에 지원되는 외부 데이터베이스를 사용하여 전환하십시오. 자세한 세부 정보

Service Monitor 요청에 실패했습니다. 이로 인해 페이지 응답이 느려질 수 있습니다. Service Monitor 상태를 확인하십시오.

Cluster 1

서비스 추가

- Add Hosts
- 시작
- 중지
- 재시작
- 롤링 재시작
- 클라이언트 구성 배포
- Kerberos 클라이언트 구성 배포
- 클러스터 업그레이드
- 클러스터 새로고침
- 동적 리소스 풀 새로고침
- Inspect Hosts in Cluster

차트 30분 1시간 2시간 6시간 12시간 1일 7d 30d

쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다.

클러스터 CPU 쿼리 오류

클러스터 디스크 IO 쿼리 오류

클러스터 네트워크 IO 쿼리 오류

Feedback

server01.hadoop.com:7180/cmf/clusters/1/add-service/in...

Kafka 설치

Cluster 1에 서비스 추가 - Cloud x +

← → ↻ 주의 요함 | server01.hadoop.com:7180/cmf/clusters/1/add-service/index

Cloudera Manager

Cluster 1에 서비스 추가

추가할 서비스 유형을 선택합니다.

서비스 유형	설명
<input type="radio"/> ADLS Connector	The ADLS Connector service provides key management for accessing ADLS Gen1 accounts and ADLS Gen2 containers from CDH services.
<input type="radio"/> Accumulo	The Apache Accumulo sorted, distributed key/value store is a robust, scalable, high performance data storage and retrieval system. This service only works with releases meant to run on top of CDH6.
<input type="radio"/> Flume	Flume은 대부분의 소스에서 데이터를 수집 및 집계하여 HDFS 같은 영구 스토리지로 저장합니다.
<input type="radio"/> HBase	Apache HBase는 대규모 데이터 세트에 임의의 실시간 읽기/쓰기 액세스를 제공합니다(HDFS와 ZooKeeper 필요).
<input type="radio"/> HDFS	Apache HDFS(Hadoop Distributed File System)는 Hadoop 애플리케이션이 사용하는 기본 스토리지 시스템입니다. HDFS는 데이터 블록에 대한 여러 개의 복제본을 생성하고 이를 클러스터 전반에 걸쳐 컴퓨팅 호스트에 배포하여 안정적이고 매우 빠른 계산을 지원합니다.
<input type="radio"/> Hive	Hive는 SQL과 유사한 언어인 HiveQL을 제공하는 데이터 웨어하우스 시스템입니다.
<input type="radio"/> Hue	Hue는 CDH(Cloudera Distribution Including Apache Hadoop)에서 작동하는 GUI(그래픽 사용자 인터페이스)입니다(HDFS, MapReduce, Hive 필요).
<input type="radio"/> Impala	Impala에서는 HDFS 및 HBase에 저장된 데이터에 대해 실시간 SQL 쿼리 인터페이스를 제공합니다. Impala에는 Hive 서비스가 필요하며 Hue와 Hive Metastore를 공유합니다.
<input type="radio"/> Isilon	EMC Isilon is a distributed filesystem.
<input type="radio"/> Java KeyStore KMS	The Hadoop Key Management Service with file-based Java KeyStore. Maintains a single copy of keys, using simple password-based protection. Requires CDH 6.0+. Not recommended for production use.
<input checked="" type="radio"/> Kafka	Apache Kafka is publish-subscribe messaging rethought as a distributed commit log.
<input type="radio"/> Key-Value Store Indexer	Key-Value Store Indexer는 HBase에 포함된 테이블 안의 데이터의 변경 사항을 수신 대기하고 Solr를 사용하여 인덱싱합니다.
<input type="radio"/> Kudu	Kudu is a true column store for the Hadoop ecosystem.
<input type="radio"/> Oozie	Oozie는 클러스터의 데이터 처리 작업을 관리하는 워크플로우 조정 서비스입니다.
<input type="radio"/> S3 Connector	The S3 Connector Service securely provides a single set of AWS credentials to Impala and Hue. This enables Hue administrators to browse the S3 filesystem and define Impala tables backed by S3 data authorized to that AWS identity, and also enables Impala users to query S3-backed tables without directly providing AWS credentials, subject to having the proper permissions defined via Sentry. The S3 Connector only supports the S3A protocol.
<input type="radio"/> Sentry	Sentry 서비스는 인증 정책 메타데이터를 저장하고 동시 클라이언트와 이 메타데이터에 대해 안전하게 보호된 액세스를 제공합니다.
<input type="radio"/> Solr	Solr은 HDFS에 저장된 데이터를 인덱싱 및 검색하는 배포 서비스입니다.
<input type="radio"/> Spark	Apache Spark is an open source cluster computing system. This service runs Spark as an application on YARN.
<input type="radio"/> Sqoop 1 Client	Configuration and connector management for Sqoop 1.
<input type="radio"/> YARN (MR2 Included)	YARN이라고도 하는 MRv2(Apache Hadoop MapReduce 2.0)는 MapReduce 애플리케이션을 지원하는 데이터 계산 프레임워크입니다(HDFS 필요).
<input type="radio"/> ZooKeeper	Apache ZooKeeper는 구성 데이터를 유지하고 동기화하는 중앙 집중식 서비스입니다.

뒤로 계속

Kafka 설치

Cluster 1에 Kafka 서비스 추가 - x

Cluster 1에 Kafka 서비스 추가

Cloudera Manager

admin

✓ Select Dependencies

역할 할당 사용자 지정

변경 내용 검토

명령 세부 정보

요약

역할 할당 사용자 지정

여기에서 새 서비스에 대한 역할 할당을 사용자 지정할 수 있지만 단일 호스트에 너무 많은 수의 역할을 할당하는 등 올바르게 할당할 경우, 성능이 저하될 수 있습니다.

역할 할당을 호스트별로 볼 수도 있습니다. [호스트별로 보기](#)

Kafka Broker

호스트 선택

Kafka MirrorMaker

호스트 선택

Gateway

호스트 선택

할당된 호스트가 너무 복잡하여, 최소 1개입니다.

server01.hadoop.com:7180/cmf/clusters/1/add-service/index?serviceType=KAFKA#

뒤로

계속

Kafka 설치

Cluster 1에 Kafka 서비스 추가 - x

← → ↺ 주의 요함 | server01.hadoop.com:7180/cmf/clusters/1/add-service/index?serviceType=KAFKA#step=role...

Cloudera Manager

지원 admin

1개의 호스트가 선택됨

새 역할 또는 기존 역할에 대한 호스트를 선택하십시오. 호스트 목록은 유효하지 않은 호스트를 제거하도록 필터링되었습니다. 이러한 호스트에는 상태가 불량하거나, 다른 클러스터의 구성원이거나, 호환되지 않는 CDH 버전이 설치되어 있는 호스트가 포함됩니다.

호스트 이름(host01, host[01-10]), IP 주소 또는 랙을 입력하십시오. 검색

팁: 첫 번째 확인란을 클릭하고 Shift 키를 누른 채로 마지막 확인란을 클릭하여 범위를 선택합니다.

<input type="checkbox"/>	호스트 이름 ↑	IP 주소 ↓	랙 ↓	코어 ↓	물리적 메모리 ↓	기존 역할	추가된 역할
<input type="checkbox"/>	server01.hadoop.com	192.168.56.101	/default	1	3.4 GiB	B NN SNN AP ES HM SM JHS RM	
<input checked="" type="checkbox"/>	server02.hadoop.com	192.168.56.102	/default	1	2.9 GiB	A DN NM S	KB

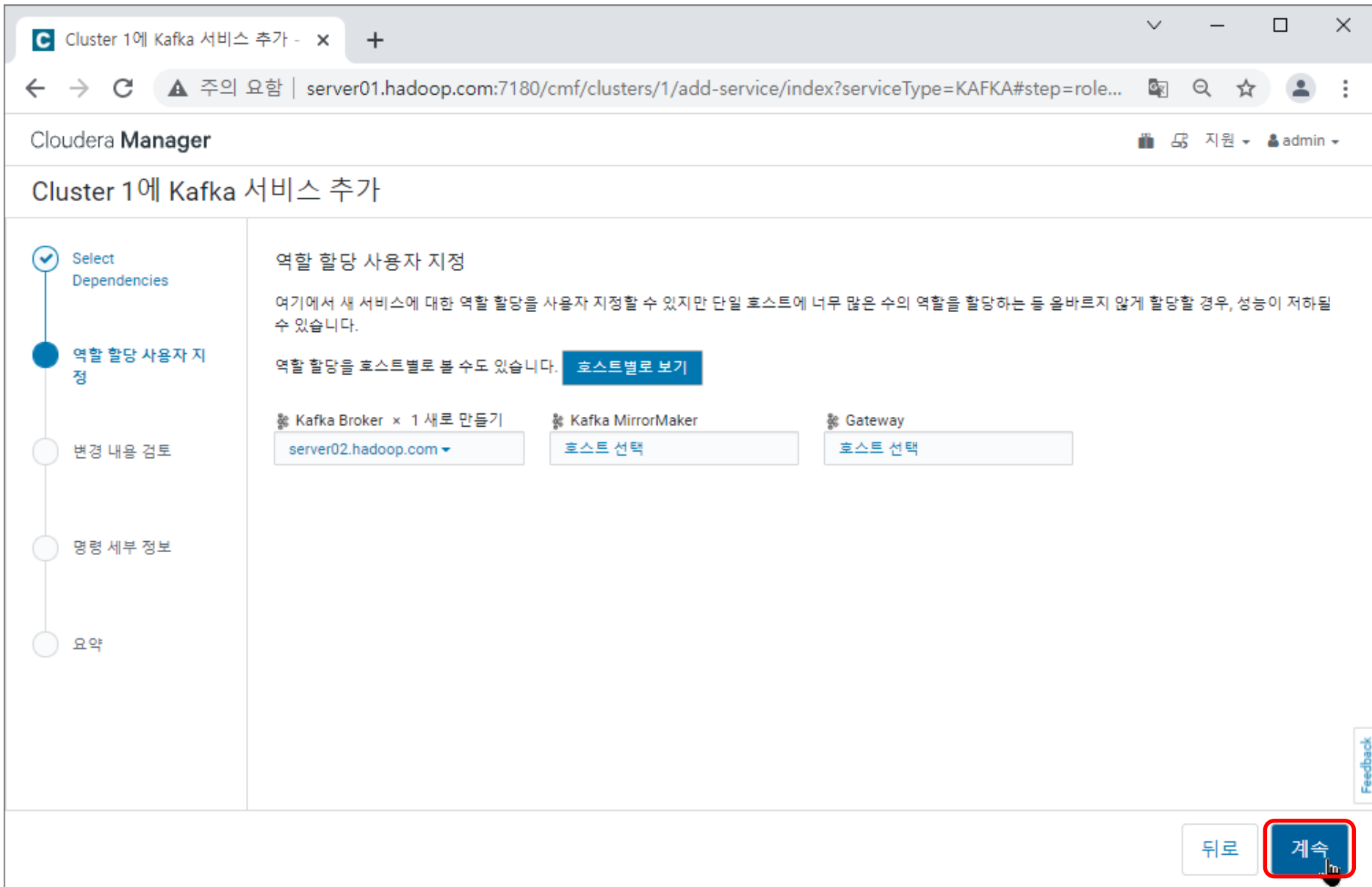
Displaying 1 - 2 of 2

취소 **확인**

뒤로

계속

Kafka 설치



Kafka 설치

Cluster 1에 Kafka 서비스 추가

Cluster 1에 Kafka 서비스 추가

Select Dependencies

역할 할당 사용자 지정

변경 내용 검토

명령 세부 정보

요약

변경 내용 검토

ZooKeeper Root

zookeeper.chroot

Kafka(서비스 전체)

?

Enable Kerberos Authentication

kerberos.auth.enable

☐ Kafka(서비스 전체)

?

Topic Auto Creation

auto.create.topics.enable

☒ Kafka(서비스 전체)

?

Default Replication Factor

default.replication.factor

Kafka(서비스 전체)

1

?

Offset Commit Topic Number of Partitions

offsets.topic.num.partitions

Kafka(서비스 전체)

50

?

Offset Commit Topic Replication Factor

offsets.topic.replication.factor

Kafka(서비스 전체)

3

?

TCP Port

port

Kafka Broker Default Group

9092

?

TLS/SSL Port

ssl_port

Kafka Broker Default Group

9093

?

뒤로

계속

Feedback

Kafka 설치

Cluster 1에 Kafka 서비스 추가 - x +

← → ↺ 주의 요함 | server01.hadoop.com:7180/cmf/clusters/1/add-service/index?serviceType=KAFKA#step=com...

Cloudera Manager 지원 admin

Cluster 1에 Kafka 서비스 추가

✓ Select Dependencies

✓ 역할 할당 사용자 지정

✓ 변경 내용 검토

● 명령 세부 정보

○ 요약

첫 번째 실행 명령

상태 완료됨 Oct 26, 10:21:51 PM 46.82s

Finished First Run of the following services successfully: Kafka.

✓ 1/1 단계가 완료되었습니다.

☒ Show All Steps ☐ Show Only Failed Steps ☐ Show Only Running Steps

> Run a set of services for the first time	Oct 26, 10:21:51 PM	46.81s
---	---------------------	--------

뒤로

계속

Kafka 설치

Cluster 1에 Kafka 서비스 추가 - x

Cluster 1에 Kafka 서비스 추가 - x

server01.hadoop.com:7180/cmf/clusters/1/add-service/index?serviceType=KAFKA#step=finis...

주의 요함

Cloudera Manager

지원 admin

Cluster 1에 Kafka 서비스 추가

Select Dependencies

역할 할당 사용자 지정

변경 내용 검토

명령 세부 정보

요약

요약

새 서비스가 클러스터에 설치 및 구성되었습니다.

참고: 여전히 새 서비스를 시작해야 할 수 있습니다. 시작하기 전에 오래된 구성이 포함된 모든 종속 서비스를 재시작하는 것이 좋습니다. 이러한 작업은 아래에서 완료를 클릭하면 주 페이지에서 수행할 수 있습니다.

뒤로

완료

Kafka 설치

The screenshot shows the Cloudera Manager web interface. The browser address bar indicates the URL is `server01.hadoop.com:7180/cmf/home`. The page title is "Cloudera Manager". The navigation bar includes links for "홈" (Home), "상태" (Status), "모든 상태 문제" (All Status Problems), "구성" (Configuration), and "모든 최근 명령" (All Recent Commands). A search bar and a user profile icon labeled "admin" are also present.

A yellow banner at the top contains a warning message: "내장된 PostgreSQL 데이터베이스를 사용하는 비-프로덕션 모드에서 Cloudera Manager를 실행 중입니다. 프로덕션으로 이동하기 전에 지원되는 외부 데이터베이스를 사용하여 전환하십시오. 자세한 세부 정보" (Cloudera Manager is running in non-production mode using the built-in PostgreSQL database. Before moving to production, use supported external databases for migration. See details for more information).

Below the banner, a message states: "Service Monitor 요청에 실패했습니다. 이로 인해 페이지 응답이 느려질 수 있습니다. Service Monitor 상태를 확인하십시오." (The Service Monitor request failed. This may slow down page response. Check the Service Monitor status).

The main content area is titled "Cluster 1" and shows a list of services. The "Kafka" service is highlighted with a red box and a mouse cursor. The services listed are:

- CDH 6.3.2 (Parcel)
- 2개의 호스트 (2 Hosts)
- Flume
- Kafka (highlighted)
- YARN (MR2 I...)
- ZooKeeper

Each service has a status icon (green for healthy, orange for degraded, red for failed) and a wrench icon for configuration. The Kafka service shows a status of 3 (degraded) and 1 (failed).

On the right side, there are three monitoring charts:

- 클러스터 CPU (Cluster CPU): Query 오류 (Query Error)
- 클러스터 디스크 IO (Cluster Disk IO): Query 오류 (Query Error)
- 클러스터 네트워크 IO (Cluster Network IO): Query 오류 (Query Error)

A red box highlights a message at the top of the charts: "쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다." (Cannot execute query. Host Monitor is not running).

At the bottom left, the URL `server01.hadoop.com:7180/cmf/services/12/status` is visible.

Kafka 설치

The screenshot shows the Cloudera Manager interface for a Kafka cluster. The browser address bar indicates the URL is `server01.hadoop.com:7180/cm/services/12/status`. The page title is "Kafka" under "Cluster 1". The "구성" (Configuration) tab is highlighted with a red box. The "상태 테스트" (Status Test) section shows a message: "현재 상태 검사를 사용할 수 없습니다. Service Monitor의 상태를 확인하십시오." (Cannot use current status test. Please check the status of the Service Monitor). The "상태 요약" (Status Summary) table shows two rows: "Kafka Broker" and "호스트" (Host), both with a status of "1 알 수 없는 상태" (1 Unknown status). The "상태 기록" (Status History) section shows a message: "현재 Event Server를 사용할 수 없습니다. Event Server 상태를 확인하십시오." (Cannot use current Event Server. Please check the status of the Event Server). The "쿼리 오류" (Query Error) section shows two messages: "쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다." (Cannot execute query. Host Monitor is not running) and "모든 Kafka Brokers의 총 Messages Received" (Total Messages Received for all Kafka Brokers). The bottom of the page shows the URL `server01.hadoop.com:7180/cm/services/12/config`.

Cloudera Manager 클러스터 호스트 진단 감사 차트 관리

Cluster 1
Kafka 작업

30분(Oct 26, 10:29 PM KST 이전)

상태 인스턴스 구성 명령 차트 라이브러리 감사 쿼리 링크

상태 테스트 트리거 생성 차트 30분 1시간 2시간 6시간 12시간 1일 7d 30d

현재 상태 검사를 사용할 수 없습니다. Service Monitor의 상태를 확인하십시오.

상태 요약

Kafka Broker	1 알 수 없는 상태
호스트	1 알 수 없는 상태

상태 기록

현재 Event Server를 사용할 수 없습니다. Event Server 상태를 확인하십시오.

쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다.

Active Controllers

쿼리 오류

모든 Kafka Brokers의 총 Messages Received

쿼리 오류

server01.hadoop.com:7180/cm/services/12/config

Kafka 설치

구성 - Kafka - Cloudera Manager x +

← → ↺ 주의 요함 | server01.hadoop.com:7180/cm/services/12/config#filterfreeText=Data%20Retention%20Time 🔍 ☆ 👤 ⋮

Cloudera Manager 클러스터 ▾ 호스트 ▾ 진단 ▾ 감사 차트 ▾ 관리 ▾

Cluster 1

🔍 Kafka 작업 ▾ Oct 26, 10:31 PM KST

상태 인스턴스 구성 명령 차트 라이브러리 감사 쿼리 링크 ▾

Data Retention Time 역할 그룹

모든 설명 표시

필터

▼ 범위

Kafka (서비스 자원)	0
Gateway	0
Kafka Broker	1
Kafka MirrorMaker	0

▼ 범주

고급	0
기본	1
로그	0
리소스 관리	0
모니터링	0
보안	0
성능	0
스택 수집	0
프로파일링	0

Data Retention Time
log.retention.ms

Kafka Broker Default Group

10 분 ▾

25 ▾ 페이지 기준

1 Edited Value 변경 이유

변경 내용 저장

Flume 수집 기능 구현

The screenshot displays the Cloudera Manager web interface. The browser address bar shows the URL `server01.hadoop.com:7180/cmf/home`. The main navigation bar includes links for Home, Status, All Status Problems, Configuration (11 items), and All Recent Commands. A search bar and user profile (admin) are also present.

A yellow banner at the top contains a message about the PostgreSQL database and a link for more details. Below this, a red box highlights a warning: "Service Monitor 요청에 실패했습니다. 이로 인해 페이지 응답이 느려질 수 있습니다. Service Monitor 상태를 확인하십시오."

The left sidebar shows the "Cluster 1" configuration. Under "CDH 6.3.2 (Parcel)", the "Flume" service is highlighted with a red box and a mouse cursor. Other services listed include HDFS, Kafka, YARN (MR2 I...), and ZooKeeper.

The main content area is titled "차트" (Charts) and shows monitoring status for "Cluster 1". A red box at the top of this section contains the message: "쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다." Below this, three monitoring charts are displayed, each showing "쿼리 오류" (Query Error):

- 클러스터 CPU (Cluster CPU)
- 클러스터 디스크 IO (Cluster Disk IO)
- 클러스터 네트워크 IO (Cluster Network IO)

At the bottom left, the "Cloudera Management Service" section shows the "Cloudera Man..." service with 4 items.

The bottom status bar shows the URL `server01.hadoop.com:7180/cmf/services/10/status`.

Flume 수집 기능 구현

Cloudera Manager 클러스터 호스트 진단 감사 차트 관리

Cluster 1

Flume 작업

30분(Oct 26, 10:36 PM KST 이전)

상태 인스턴스 **구성** 명령 메트릭 세부 정보 차트 라이브러리 감사 쿼리 링크

상태 테스트 트리거 생성 차트 30분 1시간 2시간 6시간 12시간 1일 7d 30d

현재 상태 검사를 사용할 수 없습니다. Service Monitor의 상태를 확인하십시오.

상태 요약

Agent	1 알 수 없는 상태
호스트	1 알 수 없는 상태

상태 기록

현재 Event Server를 사용할 수 없습니다. Event Server 상태를 확인하십시오.

쿼리를 실행할 수 없습니다. Host Monitor이(가) 실행되고 있지 않습니다.

상태

percent

쿼리 오류

중요 이벤트 및 알림

쿼리 오류

server01.hadoop.com:7180/cmf/services/10/config

Flume 수집 기능 구현

구성 - Flume - Cloudera Manag x +

← → ↻ 주의 요함 | server01.hadoop.com:7180/cmf/services/10/config

기본값 아님
재정의 7
0

시스템 그룹

flume

Flume(서비스 전체)

flume

Agent 이름

Agent Default Group ↻
SmartCar_Agent

구성 파일

Agent Default Group ↻
SmartCar_Agent.sources = SmartCarInfo_SpoolSource DriverCarInfo_TailSource
SmartCar_Agent.channels = SmartCarInfo_Channel DriverCarInfo_Channel
SmartCar_Agent.sinks = SmartCarInfo_LoggerSink DriverCarInfo_KafkaSink
SmartCar_Agent.sources.SmartCarInfo_SpoolSource.type = spooldir

Flume 홈 디렉토리

Agent Default Group
/var/lib/flume-ng

플러그인 디렉토리

Agent Default Group

2 Edited Values 변경 이유

변경 내용 저장

Feedback

SmartCar Agent에 Interceptor 추가

- Interceptor는 Source와 Channel의 중간에서 데이터를 가공하는 역할.
- 플럼의 Source에서 유입되는 데이터 중 일부 데이터를 수정하거나 필요한 데이터만 필터링하는 등 중간에 데이터를 추가/가공/정제하는데 사용.
- Event : 플럼의 데이터 전송 단위 -> Header + Body로 구성.
- Interceptor는 Event의 Header에 특정 값을 추가하거나 Event의 Body에 데이터를 가공하는 기능으로 활용.
- 파일럿 프로젝트에서는 SmartCarInfo 로그 파일을 수집하는데 총 4개의 Interceptor를 추가할 것.
- 우선 Filter Interceptor 하나만 추가, 4장에서 나머지 3개 사용.

Kafka 기능 구현

➤ 카프카 Topic 생성

: kafka-topics --create --zookeeper server02.Hadoop.com:2181 --replication-factor 1
--partitions 1 --topic SmartCar-Topic

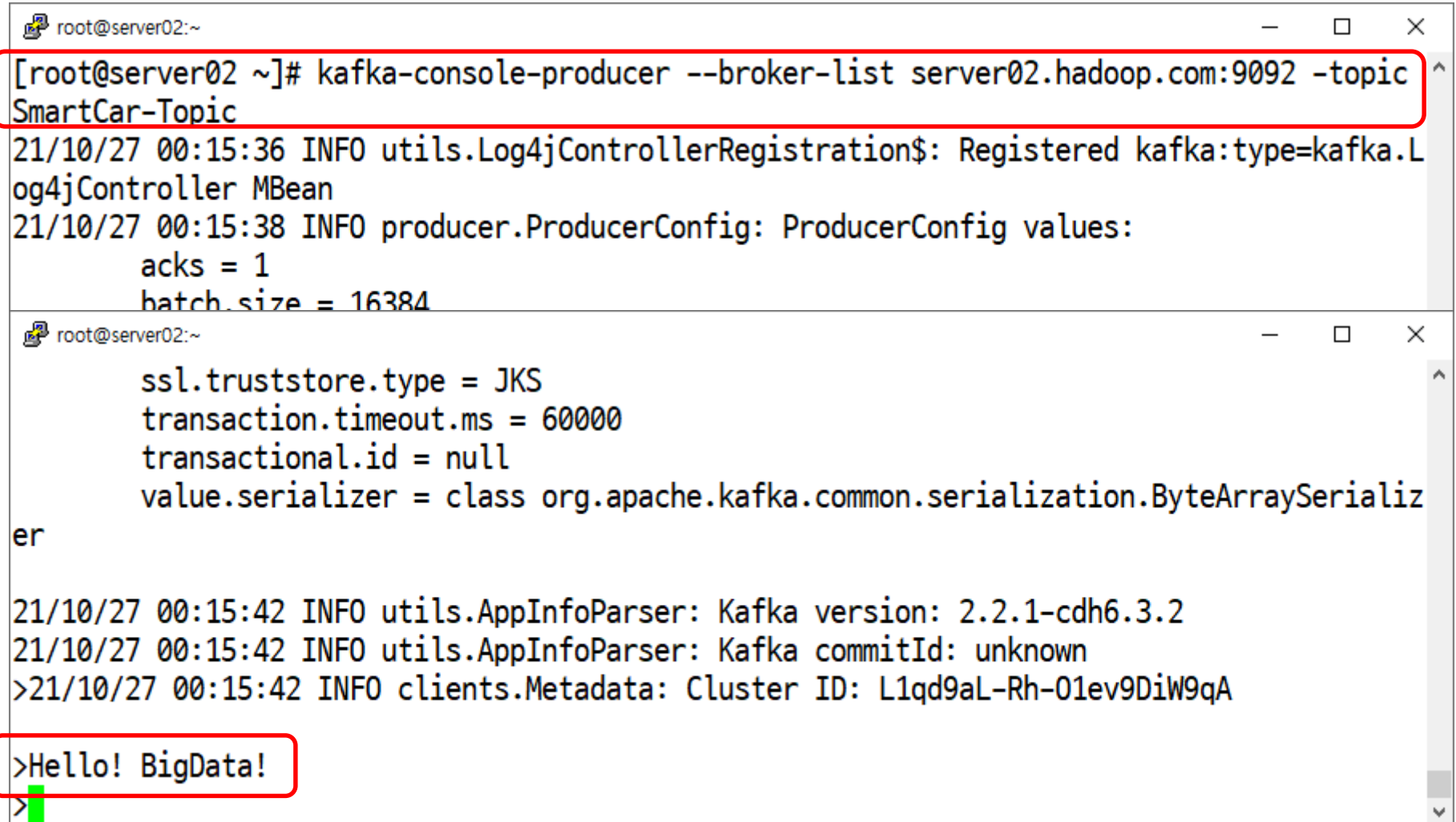


```
root@server02:~  
[root@server02 ~]# kafka-topics --create --zookeeper server02.hadoop.com:2181 --replication-factor 1  
--partitions 1 --topic SmartCar-Topic  
21/10/26 23:15:33 INFO utils.Log4jControllerRegistration$: Registered kafka:type=kafka.Log4jController MBean  
21/10/26 23:15:34 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Initializing a new session to server02.hadoop.com:2181.  
21/10/26 23:15:34 INFO zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.5-cdh6.3.2--1, built on 11/08/2019 13:12 GMT  
21/10/26 23:15:34 INFO zookeeper.ZooKeeper: Client environment:host.name=server02.hadoop.com  
21/10/26 23:15:34 INFO zookeeper.ZooKeeper: Client environment:java.version=1.8.0_181  
21/10/26 23:15:34 INFO zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corporation  
21/10/26 23:15:34 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Connected.  
21/10/26 23:15:35 INFO zk.AdminZkClient: Creating topic SmartCar-Topic with configuration {} and initial partition assignment Map(0 -> ArrayBuffer(28))  
Created topic SmartCar-Topic.  
21/10/26 23:15:35 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closing.  
21/10/26 23:15:35 INFO zookeeper.ClientCnxn: EventThread shut down  
21/10/26 23:15:35 INFO zookeeper.ZooKeeper: Session: 0x17cbc618aff0003 closed  
21/10/26 23:15:35 INFO zookeeper.ZooKeeperClient: [ZooKeeperClient] Closed.  
[root@server02 ~]#
```

Kafka 기능 구현

➤ 카프카 Producer 사용

: kafka-console-producer --broker-list server02.hadoop.com:9092 --topic SmartCar-Topic

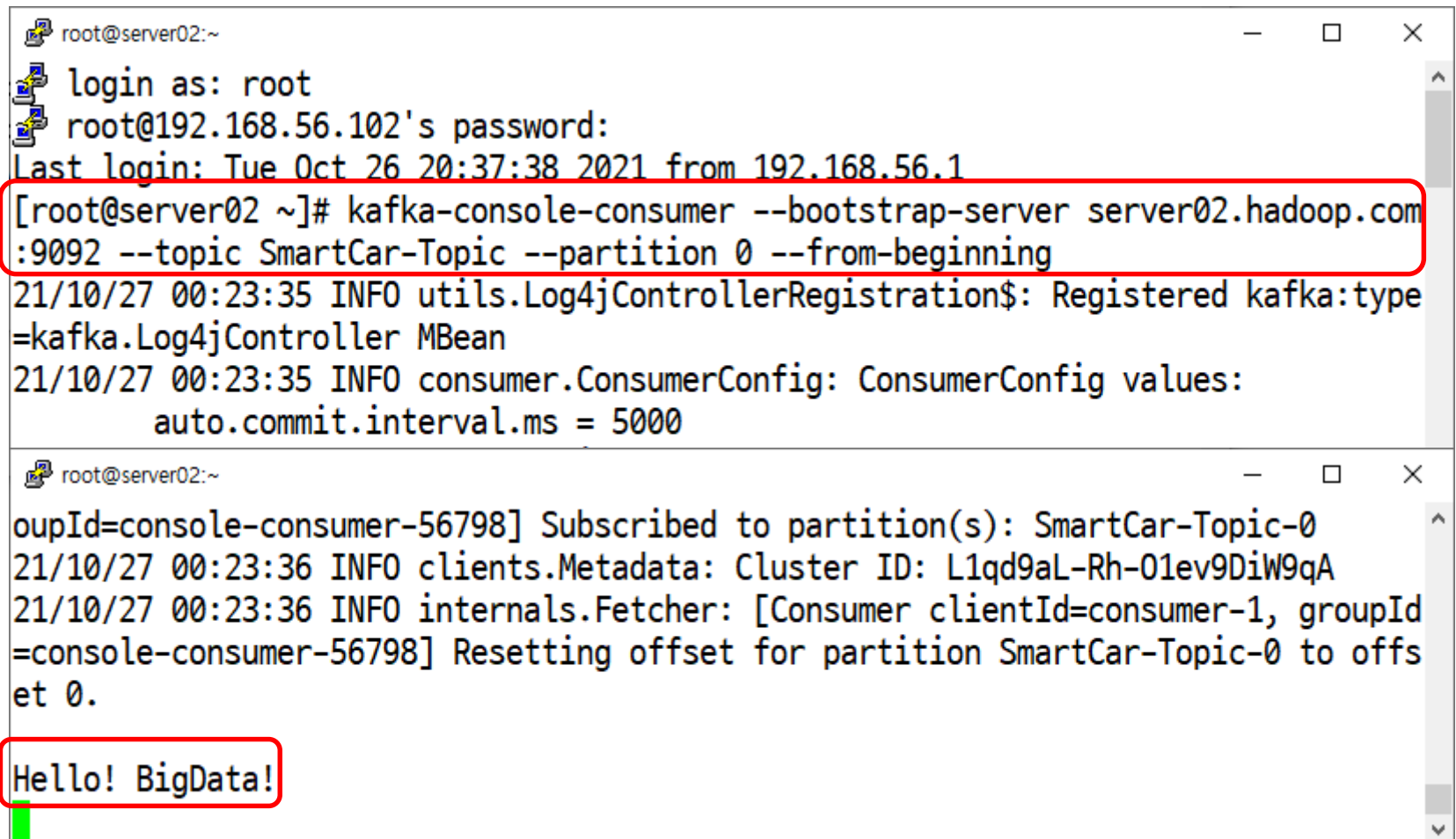


```
root@server02:~  
[root@server02 ~]# kafka-console-producer --broker-list server02.hadoop.com:9092 -topic SmartCar-Topic  
21/10/27 00:15:36 INFO utils.Log4jControllerRegistration$: Registered kafka:type=kafka.Log4jController MBean  
21/10/27 00:15:38 INFO producer.ProducerConfig: ProducerConfig values:  
    acks = 1  
    batch.size = 16384  
    ssl.truststore.type = JKS  
    transaction.timeout.ms = 60000  
    transactional.id = null  
    value.serializer = class org.apache.kafka.common.serialization.ByteArraySerializer  
21/10/27 00:15:42 INFO utils.AppInfoParser: Kafka version: 2.2.1-cdh6.3.2  
21/10/27 00:15:42 INFO utils.AppInfoParser: Kafka commitId: unknown  
>21/10/27 00:15:42 INFO clients.Metadata: Cluster ID: L1qd9aL-Rh-01ev9DiW9qA  
>Hello! BigData!  
>
```

Kafka 기능 구현

➤ 카프카 Consumer 사용

: kafka-console-consumer --bootstrap-server server02.Hadoop.com:9092 --topic SmartCar-Topic
--partition 0 --from-beginning



The image shows two terminal windows. The top window shows the command `kafka-console-consumer --bootstrap-server server02.hadoop.com:9092 --topic SmartCar-Topic --partition 0 --from-beginning` being executed, with the command line highlighted by a red box. The bottom window shows the output of the command, including log messages and the message `Hello! BigData!`, which is also highlighted by a red box.

```
root@server02:~  
login as: root  
root@192.168.56.102's password:  
Last login: Tue Oct 26 20:37:38 2021 from 192.168.56.1  
[root@server02 ~]# kafka-console-consumer --bootstrap-server server02.hadoop.com:9092 --topic SmartCar-Topic --partition 0 --from-beginning  
21/10/27 00:23:35 INFO utils.Log4jControllerRegistration$: Registered kafka:type=kafka.Log4jController MBean  
21/10/27 00:23:35 INFO consumer.ConsumerConfig: ConsumerConfig values:  
    auto.commit.interval.ms = 5000  
  
[root@server02 ~]#  
[2021-10-27 00:23:36] [Consumer clientId=consumer-1, groupId=console-consumer-56798] Subscribed to partition(s): SmartCar-Topic-0  
21/10/27 00:23:36 INFO clients.Metadata: Cluster ID: L1qd9aL-Rh-01ev9DiW9qA  
21/10/27 00:23:36 INFO internals.Fetcher: [Consumer clientId=consumer-1, groupId=console-consumer-56798] Resetting offset for partition SmartCar-Topic-0 to offset 0.  
  
Hello! BigData!
```

Kafka 기능 구현

➤ 특정 토픽 삭제 명령

: `kafka-topics --delete --zookeeper server02.hadoop.com:2181 --topic SmartCar-Topic`

Flume 수집 기능 점검

➤ 스마트카 로그 시뮬레이터 실행

- `java -cpbigdata.smartcar.loggen-1.0.jar com.wikibook.bigdata.smartcar.loggen.CarLogMain202xxxxx 3 &`
- `mv /home/pilot-pjt/working/SmartCar/SmartCarStatusInfo_202xxxxx.txt /home/pilot-pjt/working/car-batch-log/`

➤ 스마트카의 상태 정보 로그 파일이 플럼의 표준 출력 로그로 전송됐는지 리눅스 tail 명령어를 통해 확인.

- `tail -f /var/log/flume-ng/flume-cmf-flume-AGENT-server02.Hadoop.com.log`

➤ 플럼 에이전트 1 : 스마트가 상태 정보를 기록한 로그 파일을 일별로 수집하기 위한 배치성 플럼 에이전트.

- Logger Sink : Channel로 부터 읽어 들인 데이터를 플럼의 표준 로그 파일로 출력.

➤ 플럼 에이전트 2 : 스마트가 운전자의 운행 정보를 실시간으로 수집하기 위한 실시간성 플럼 에이전트.

- Exec-Tail Source : 로그가 쌓이고 있는 파일에 Tail 파이프라인을 이용해 실시간으로 데이터를 수집하는 기능

➤ etc

- Flume Stdout : 플럼의 Logger-Sink를 통해 표준 출력 로그가 출력.
- `ps -ef | grep smartcar.log`
- `kill -9 pid`