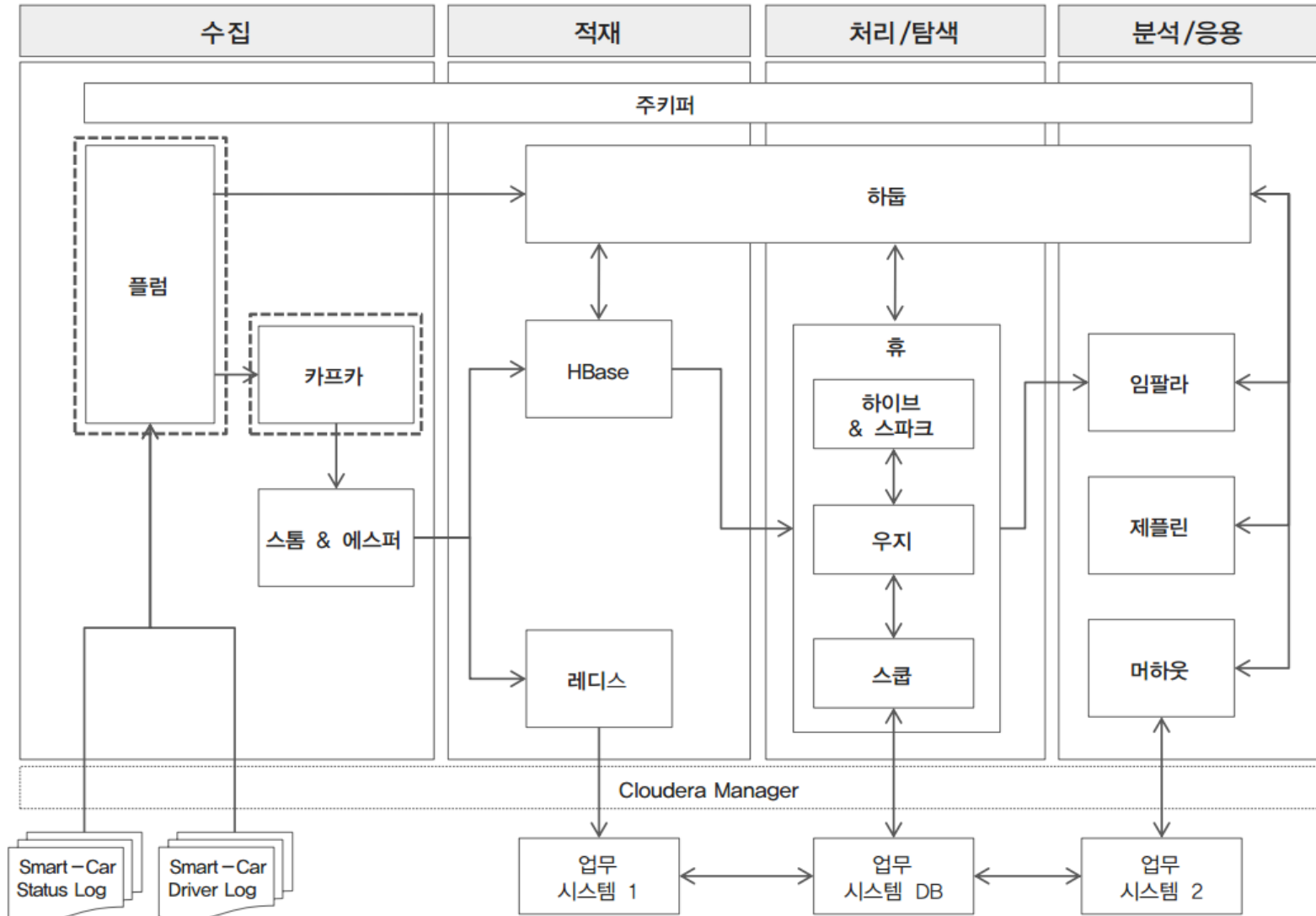


3.4 수집 파일럿 실행 2단계 - 환경 구성

파일럿 프로젝트 아키텍처



3.4 수집 파일럿 실행 2단계 - 환경 구성

플럼 설치

01. CM의 홈에서 [서비스 추가] 화면으로 이동한다. 추가할 서비스 유형 중 [Flume]을 선택하고 우측 하단의 [계속] 버튼을 클릭한다.



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 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.

그림 3.19 CM을 이용해 플럼 설치 - 플럼 컴포넌트 선택

3.4 수집 파일럿 실행 2단계 - 환경 구성

플럼 설치

02. 플럼을 설치할 서버 호스트를 server02.hadoop.com으로 선택하고 [확인] → [계속] 버튼을 클릭한다.

역할 할당 사용자 지정

여기에서 새 서비스에 대한 역할 할당을 사용자 지정할 수 있습니다. 역할 할당을 호스트별로 볼 수도 있습니다. **호스트별로**

1

Agent

호스트 선택

할당된 호스트가 너무 적습니다. 최소 1개입니다.

1개의 호스트가 선택됨

새 역할 또는 기존 역할에 대한 호스트를 선택하십시오. 호스트 목록은 유효하지 않은 호스트를 제거하도록

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 ↕
<input type="checkbox"/>	server01.hadoop.com	192.168.56.101	/default	1
<input checked="" type="checkbox"/>	server02.hadoop.com	192.168.56.102	/default	1
<input type="checkbox"/>	server03.hadoop.com	192.168.56.103	/default	1

2

3

뒤로 **계속**

그림 3.20 CM을 이용해 플럼 설치 - 설치 위치 지정

3.4 수집 파일럿 실행 2단계 - 환경 구성

플럼 설치

03. 아래와 같이 설치 완료 메시지가 보이면 플럼을 Server02에 성공적으로 설치한 것이다. [완료] 버튼을 클릭해 플럼 설치를 종료한다.

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

그림 3.21 CM을 이용해 플럼 설치 - 완료

3.4 수집 파일럿 실행 2단계 - 환경 구성

플럼 설치

04. CM의 홈 화면에 플럼 메뉴가 추가된 것을 확인할 수 있다. 플럼의 Default 그룹의 Heap Memory가 너무 작게 설정돼 있다. 파일럿 프로젝트용으로 좀 더 크게 설정한다.

CM 홈 → [Flume] → [구성]에서 검색란에 “java heap”이라고 입력하면 힙 메모리 설정창이 나타난다. 여기서 힙 크기를 “100”MiB 정도로만 늘려준다.

■ 변경 전: 50

■ 변경 후: 100

05. CM 홈 → [Flume] 우측의 콤보박스를 선택하고 [시작] 메뉴를 선택해 플럼 에이전트를 구동한다.

3.4 수집 파일럿 실행 2단계 - 환경 구성

카프카 설치

01. CM의 홈에서 [서비스 추가] 화면으로 이동한다. 추가할 서비스 유형 중 [Kafka]를 선택하고 우측 하단의 [계속] 버튼을 클릭한다.

The screenshot shows the Cloudera Manager interface for adding services to a cluster. The interface is divided into two main panels. The left panel shows the cluster configuration for 'Cluster 1', including CDH 6.3.2 (Parcel) and various services like HDFS, YARN (MR2), ZooKeeper, and Cloudera Manager. The right panel shows a list of available services to be added. The 'Kafka' service is highlighted with a red box and a circled '3'. The '계속' (Continue) button is highlighted with a red box and a circled '4'.

Service Icon	Service Name	Description
Impala	Impala	Impala에서는 HDFS 및 HBase에 저장된 데이터에 대해 실시간 SQL 쿼리 인터페이스를 제공합니다. Impala에는 Hive 서비스가 필요하며 Hue와 Hive Metastore를 공유합니다.
Isilon	Isilon	EMC Isilon is a distributed filesystem.
Java KeyStore KMS	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.
Kafka	Kafka	Apache Kafka is publish-subscribe messaging rethought as a distributed commit log.
Key-Value Store Indexer	Key-Value Store Indexer	Key-Value Store Indexer는 HBase에 포함된 테이블 안의 데이터의 변경 사항을 수신 대기하고 Solr을 사용하여 인덱싱합니다.
Kudu	Kudu	Kudu is a true column store for the Hadoop ecosystem.
Oozie	Oozie	Oozie는 클러스터의 데이터 처리 작업을 관리하는 워크플로우 조정 서비스입니다.

그림 3.22 CM을 이용해 카프카 설치 - 카프카 컴포넌트 선택

3.4 수집 파일럿 실행 2단계 - 환경 구성

카프카 설치

02. 카프카를 설치할 서버 호스트를 server02.hadoop.com으로 선택하고 [확인] → [계속] 버튼을 클릭한다. 참고로 카프카 MirrorMaker, Gateway는 선택하지 않는다.

역할 할당 사용자 지정

여기에서 새 서비스에 대한 역할
저하될 수 있습니다. ②

역할 할당을 호스트별로 볼 수도
①

Kafka Broker

호스트 선택

할당된 호스트가 너무 적습니다.
최소 1개입니다.

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 맥을 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	4.9 GiB	NN SNN JHS RM	
<input checked="" type="checkbox"/>	server02.hadoop.com	192.168.56.102	/default	1	4.9 GiB	A DN NM S	KB
<input type="checkbox"/>	server03.hadoop.com	192.168.56.103	/default	1	2.9 GiB	DN C AP C ES C HM C RM C SM NM	

Kafka MirrorMaker
호스트 선택

Gateway
호스트 선택

③
뒤로 계속

그림 3.23 CM을 이용해 카프카 설치 - 설치 위치 지정

3.4 수집 파일럿 실행 2단계 - 환경 구성

카프카 설치

03. 카프카 변경 내용 검토가 나타나면 기본값을 유지하고 [계속] 버튼을 클릭한다. 카프카 구성요소들이 Server02에 설치되고 카프카 서버가 시작된다.

첫 번째 실행 명령

상태 완료됨 Mar 10, 10:30:25 PM 47.42s

Finished First Run of the following services successfully: Kafka.

▼ 1/1단계가 완료되었습니다.

☒ Show All Steps

☐ Show Only Failed Steps

☐ Show Only Running Steps

▼ Run a set of services for the first time Start 명령이 Kafka 서비스에서 실행됨	Mar 10, 10:30:25 PM	47.42s
▼ 3 단계 순차 실행 Start 명령이 Kafka 서비스에서 실행됨	Mar 10, 10:30:25 PM	47.33s
> Ensuring that the expected software releases are installed on hosts.	Mar 10, 10:30:25 PM	5.04s
> 클라이언트 구성 배포** Cluster 1	Mar 10, 10:30:30 PM	16.14s
> Kafka 시작 Kafka	Mar 10, 10:30:46 PM	26.05s

그림 3.24 카프카 서버 설치 및 시작

3.4 수집 파일럿 실행 2단계 - 환경 구성

카프카 설치

04. 아래와 같이 설치 완료 메시지가 보이면 카프카가 Server02에 정상적으로 설치된 것이다.

요약

✓ 새 서비스가 클러스터에 설치 및 구성되었습니다.

i 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

그림 3.25 CM을 이용해 카프카 설치 - 완료

3.4 수집 파일럿 실행 2단계 - 환경 구성

카프카 설치

05. 파일럿 환경을 고려해 카프카에 저장될 메시지의 보관 기간을 짧게 조정한다. CM 홈에서 [Kafka] → [구성]을 선택하고, 검색어로 "Data Retention Time"를 입력한 후 7일에서 15분으로 수정한다.

▪ 변경 전: 7일

▪ 변경 후: 10분

3.4 수집 파일럿 실행 2단계 - 환경 구성

실습