

빅데이터 파일럿 프로젝트

빅데이터 파일럿 프로젝트

1. 파일럿 프로젝트 도메인의 이해

파일럿 프로젝트의 기본 도메인을 이해하고, 이와 관련된 빅데이터 요구사항을 도출 및 분석한다.



2. 빅데이터 파일럿 아키텍처 이해

스마트카의 빅데이터 분석을 위한 소프트웨어/하드웨어 아키텍처를 이해한다.



3. 빅데이터 파일럿 프로젝트용 PC 환경 구성

파일럿 프로젝트 환경을 구성하기 위해 독자들의 PC에서 사용할 기본 소프트웨어를 설치하고 구성한다. 이 과정에서 자바, 이클립스, 오라클 버추얼 박스 등을 설치한다.



4. 빅데이터 파일럿 프로젝트용 PC 서버 구성

3개의 가상 머신을 생성하고, 분산 클러스터 환경을 구성하기 위한 3대의 리눅스 서버(CentOS)를 설치 및 구성한다.



5. CM(Cloudera Manager) 설치

빅데이터 소프트웨어들을 설치/관리하는 Cloudera Manager를 설치한다. CM을 이용해 빅데이터 파일럿 프로젝트의 기본 소프트웨어인 하둡, 주키퍼를 설치한다.



6. 스마트카 로그 시뮬레이터 설치

스마트카의 상태 정보와 운행 정보를 시뮬레이션해 로그 데이터를 생성하는 자바 프로그램을 설치한다.



7. 파일럿 환경 관리

파일럿 환경을 안전하게 시작하고 종료하는 방법을 설명한다.

파일럿 프로젝트 도메인 이해

➤ 요구 사항 파악

- 요구사항1 : 차량의 다양한 장치로부터 발생하는 로그 파일을 수집해서 기능별 상태를 점검
- 요구사항2 : 운전자의 운행정보가 담긴 로그를 실시간으로 수집해서 주행 패턴을 분석

요구 사항 1

표 2.1 요구사항 1: 스마트카 상태 분석 요건

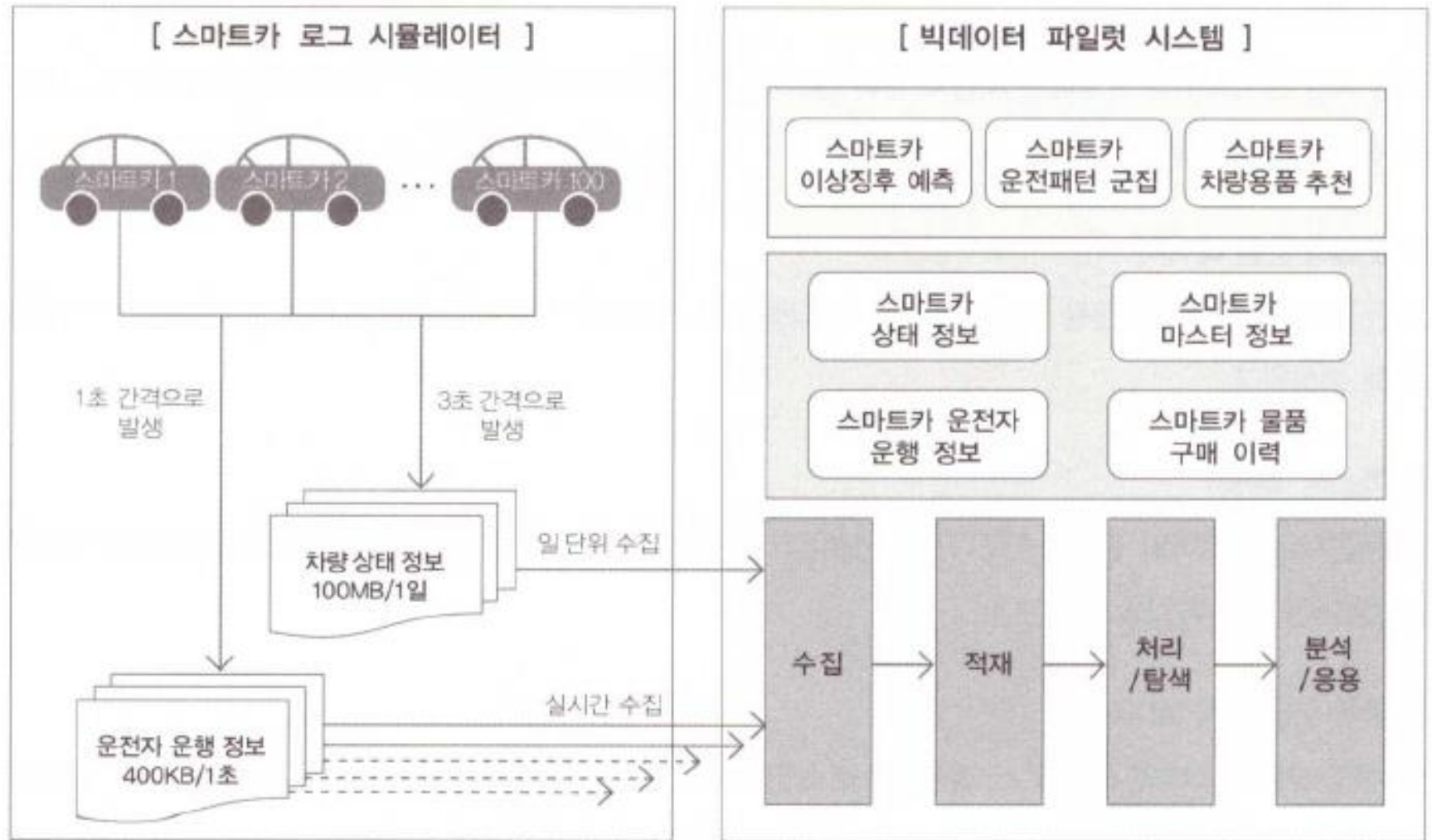
요구사항 1	차량의 다양한 장치로부터 발생하는 로그 파일을 수집해서 기능별 상태를 점검한다.		
데이터가 발생하는 위치	100대의 시범 운행 차량		
발생 데이터의 종류	대용량 로그 파일		
데이터 발생 주기	3초		
데이터 수집 주기	24시간		
데이터 수집 규모	1MB/1대(1일 수집 규모: 약 100MB/100대)		
데이터 타입	텍스트(UTF-8), 반정형		
데이터 분석 주기	일/주/월/년		
데이터 처리 유형	배치		
데이터 구분자	콤마(,)		
데이터 스키마	발생일시	로그 발생 일시	20151212081530(년월일시초)
	차량번호	차량 고유 번호	A12345
	타이어 FL	차량 앞왼쪽 타이어 상태	정상: 80~100, 비정상: 80 이하
	타이어 FR	차량 앞오른쪽 타이어 상태	정상: 80~100, 비정상: 80 이하
	타이어 BL	차량 뒤왼쪽 타이어 상태	정상: 80~100, 비정상: 80 이하
	타이어 BR	차량 뒤오른쪽 타이어 상태	정상: 80~100, 비정상: 80 이하
	라이트 FL	차량 전면왼쪽 라이트 상태	1: 정상, 2: 비정상
	라이트 FR	차량 전면오른쪽 라이트 상태	1: 정상, 2: 비정상
	라이트 BL	차량 후면왼쪽 라이트 상태	1: 정상, 2: 비정상
	라이트 BR	차량 후면오른쪽 라이트 상태	1: 정상, 2: 비정상
	엔진	차량 엔진의 상태	A: 정상, B: 점검 필요, C: 고장
	브레이크	차량 브레이크 상태	A: 정상, B: 점검 필요, C: 고장
	배터리	차량 배터리 충전 상태	1~100
	작업 요청일	수집 작업 요청일	20151212(년월일)

요구 사항 2

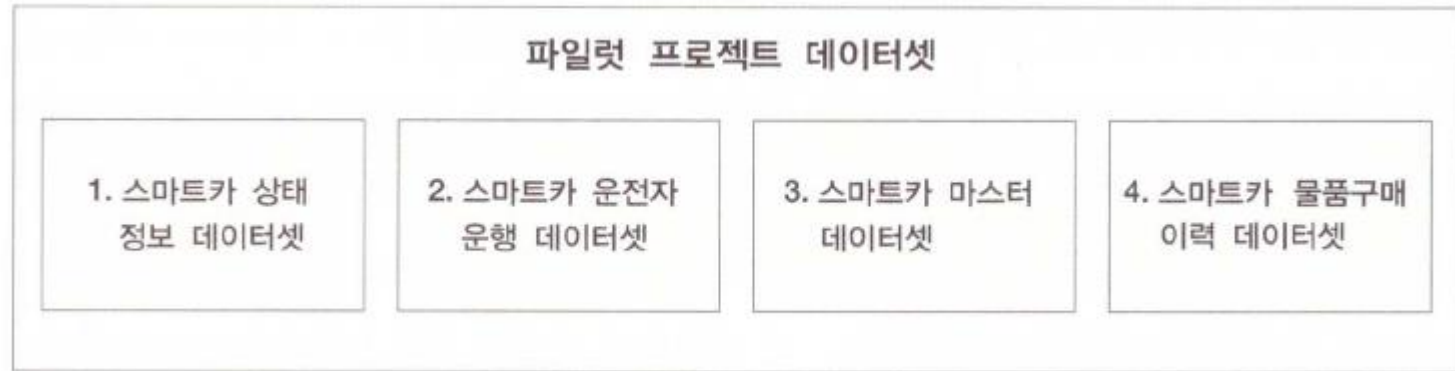
표 2.2 요구사항 2: 스마트카 운행 분석 요건

요구사항 2	운전자의 운행 정보가 담긴 로그를 실시간으로 수집해서 주행 패턴을 분석한다.		
데이터가 발생하는 위치	100대의 시범 운행 차량		
발생 데이터의 종류	실시간 로그 파일		
데이터 발생 주기	주행 관련 이벤트 발생 시		
데이터 수집 주기	1초		
데이터 수집 규모	4KB/1대(초당 수집 규모: 약 400KB/100대)		
데이터 타입	텍스트(UTF-8), 반정형		
데이터 분석 주기	실시간		
데이터 처리 유형	실시간		
데이터 구분자	콤마(,)		
데이터 스키마	발생 일시	로그 발생 일시	20151212081530(년월일시초)
	차량 번호	차량 고유 번호	A12345
데이터 스키마	가속 페달	가속페달 이벤트	0~5단계
	브레이크 페달	브레이크 이벤트	0~3단계
	운전대 회전각	운전대 회전 이벤트	F: 직진 L1: 좌회전각 1~10 L2: 좌회전각 11~20 L3: 좌회전각 21~30 R1: 우회전각 1~10 R2: 우회전각 11~20 R3: 우회전각 21~30
	방향지시등	방향지시등 이벤트	L: 왼쪽, R: 오른쪽, N: 없음
	주행 속도	차량 주행 속도	0~250km/h
	주행 지역	운전 중인 구역 번호	A구역: 1~10번, B구역: 1~10번 C구역: 1~10번, D구역: 1~10번 E구역: 1~10번, F구역: 1~10번
	작업 요청일	수집 작업 요청일	20151212(년월일)

요구 사항 이해



파일럿 프로젝트 데이터셋 유형



➤ 스마트카 상태 정보 데이터

- 스마트카의 각종 센서로부터 발생하는 차량의 상태 정보 데이터.

➤ 스마트카 운전자 운행 데이터

- 스마트카 운전자의 운전 패턴/운행 정보가 담긴 데이터.

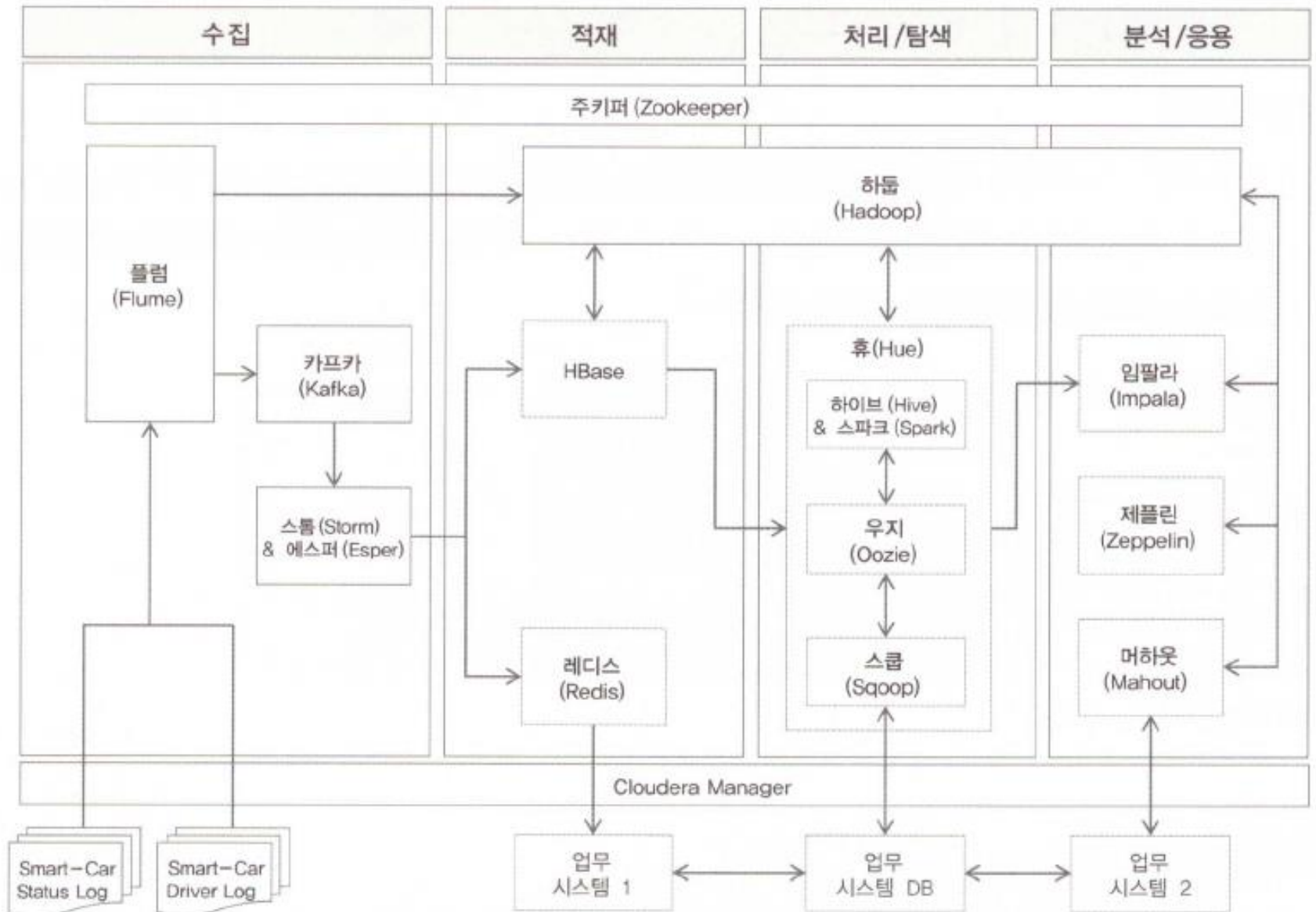
➤ 스마트카 마스터 데이터

- 스마트카 운전자의 프로파일 정보가 담긴 데이터.

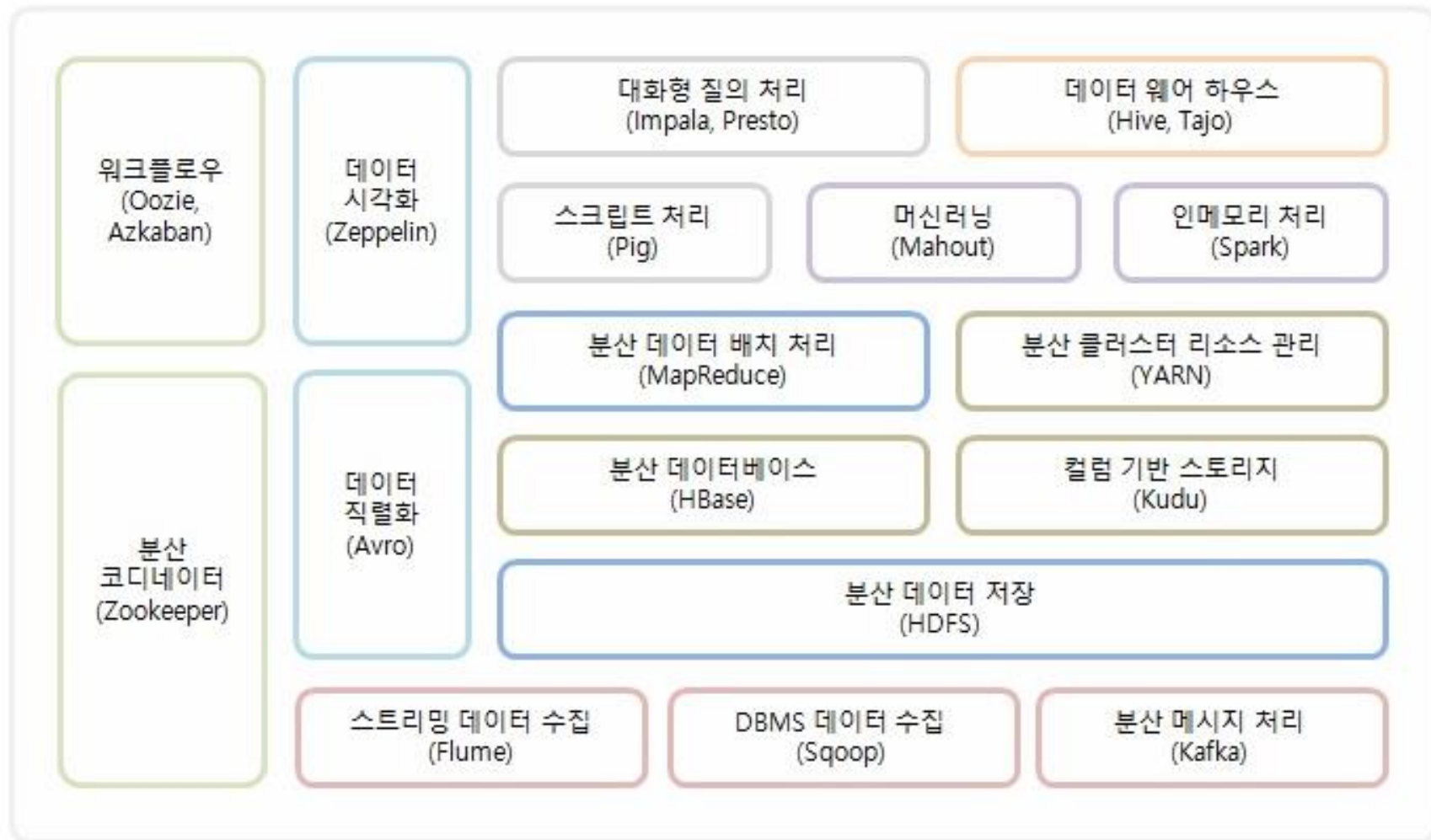
➤ 스마트카 물품 구매 이력 데이터

- 차량 내의 스마트 스크린을 통해 쇼핑몰에서 구입한 차량 물품 구매 목록 데이터

파일럿 소프트웨어 아키텍처



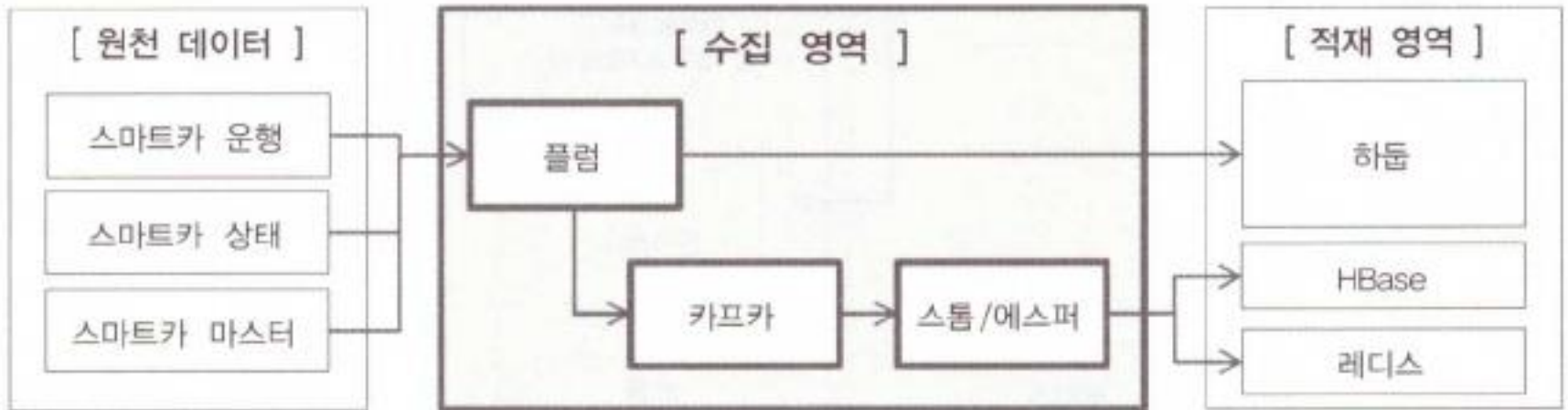
아키텍처 Layer



아키텍처 Layer



수집 Layer



➤ 플럼

- 차량의 로그 수집.

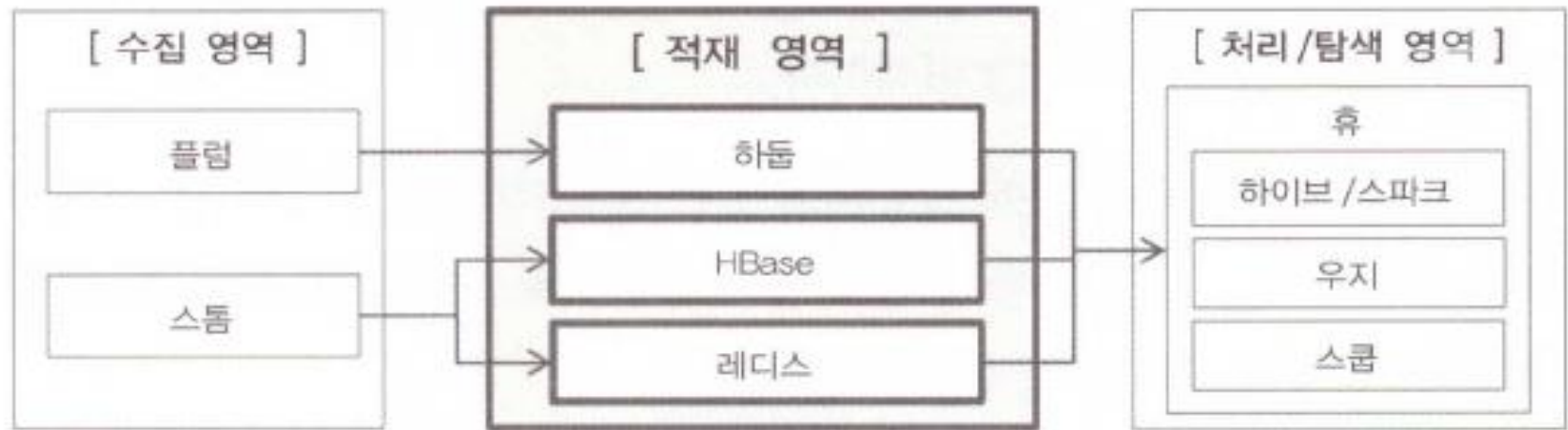
➤ 스톰 / 에스퍼

- 실시간 로그 이벤트 처리.

➤ 카프카

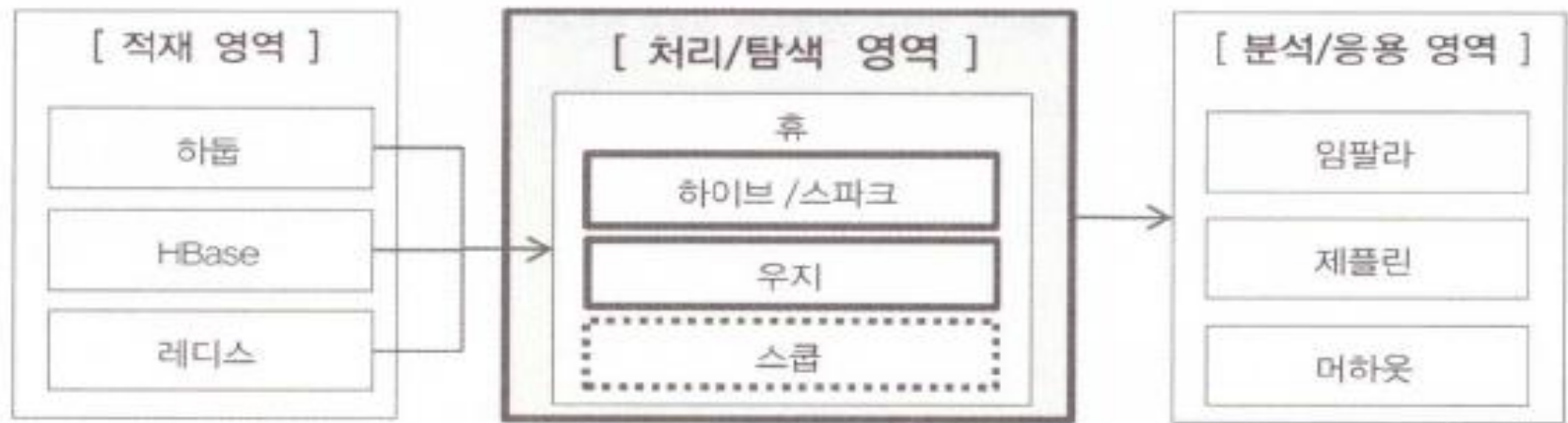
- 데이터의 안정적인 수집을 위한 버퍼링 및 트랜잭션 처리 담당.

적재 Layer



- 요구 사항1(스마트카 상태 정보 대용량 데이터)
 - 차량의 로그 수집 -> 플럼 -> 하둡으로 적재.
- 요구 사항2(스마트가 운전자 운행 정보 대용량 데이터)
 - 실시간 데이터 -> 플럼 -> 카프카 -> 스톰 -> Hbase / 레디스로 적재.
 - 스톰을 통해 실시간 이벤트 분석 수행하고, 분석된 결과에 따라 Hbase와 레디스로 나누어 적재.

처리/탐색 Layer



➤ 하이브 / 스파크

- 하둡에 적재된 데이터를 정제/변형/통합/분리/탐색 등의 작업 수행
- 데이터를 정형화된 구조로 정규화해 데이터 마트를 만들.

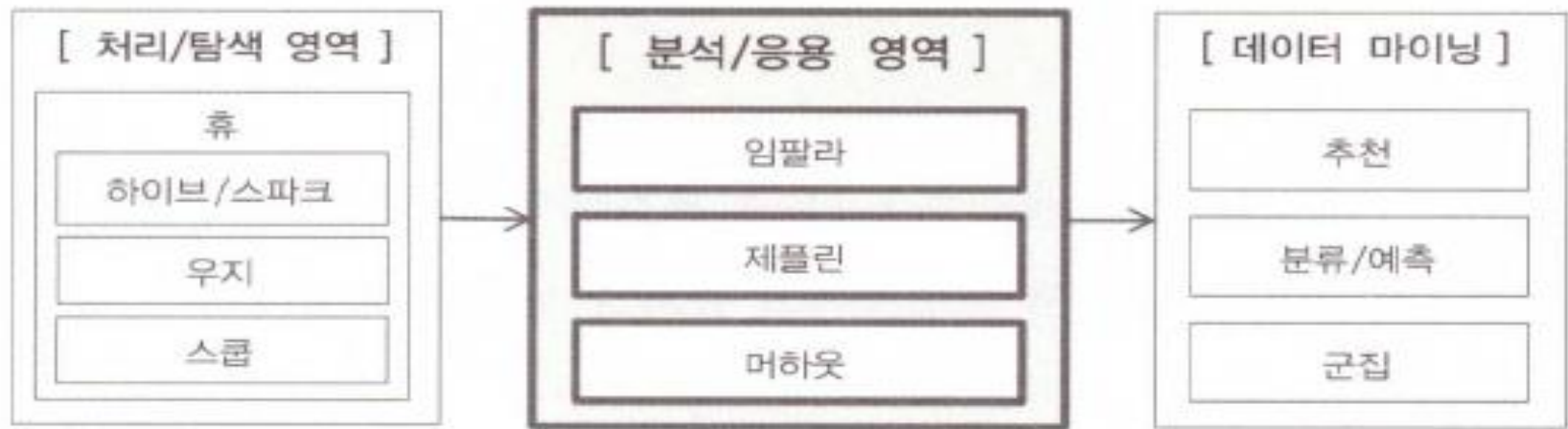
➤ 스크

- 가공/분석된 데이터를 외부로 제공하며, 필요 시 분석/응용 단계에서도 사용.

➤ 우지

- 데이터의 품질을 높이는 단계, 과정이 길고 복잡해지기 마련 .
- 프로세스를 구성해 복잡도를 낮추고, 자동화.

분석/응용 Layer



➤ 임팔라 / 제플린

- 요구사항1/2의 스마트카 상태 점검과 운전자의 운행 패턴을 빠르게 분석.

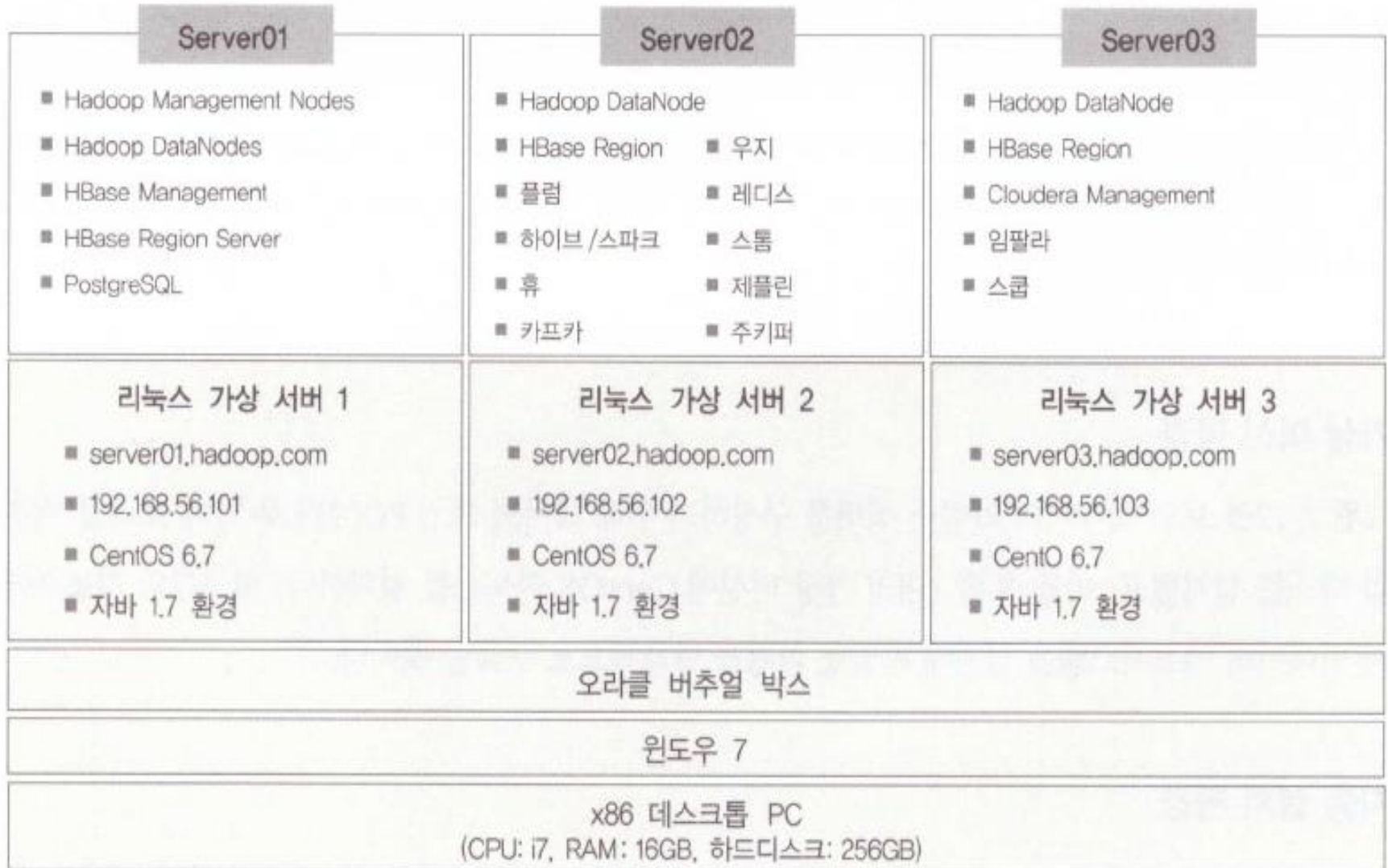
➤ 머하웃

- 하둡에 적재된 데이터를 이용해 스마트카에 대한 군집 분석, 분류/예측 분석 등을 통해 다양한 운행 패턴을 찾아 볼 수 있음.

빅데이터 기술 접근법

- 플랫폼 전문가
 - 하둡 에코시스템 설치 및 구성
- 수집/적재 전문가
 - 대규모 데이터 연동 및 통합
- 처리/탐색 전문가
 - 데이터 모델 설계 및 처리
- 분석/응용 전문가
 - 도메인 분석 및 인사이트 도출

파일럿 프로젝트의 하드웨어 구성도



파일럿 프로젝트 구축 환경

