

# 빅데이터 플랫폼 요구사항 명세서 및 보고서

빅데이터 플랫폼 요구사항 정의 및 검증하기

작성자: 이슬이

## 내용

빅데이터 플랫폼 요구사항 정의하기 .....	3
서론 .....	3
작업내용 .....	3
1) 요구사항1에 해당될 수 있는 요구사항 명세서를 작성하시오.....	3
2) 요구사항2에 해당될 수 있는 요구사항 명세서를 작성하시오.....	5
빅데이터 플랫폼 요구사항 검증하기 .....	7
작업내용 .....	7
위 빅데이터 플랫폼 요구사항에 따라 빅데이터 플랫폼을 구축하시오.....	7

# 빅데이터 플랫폼 요구사항 정의하기

## 서론

파일럿 프로젝트는 스마트카에서 발생하는 수많은 데이터로부터 가치와 통찰력을 찾기 위한 빅데이터 플랫폼 구축과 실습을 진행한다.

요구사항은 다음과 같다.

**요구사항 1** – 차량의 다양한 장치로부터 발생하는 로그파일을 수집해서 기능별 상태를 점검한다.

**요구사항 2** – 운전자의 운행 정보가 담긴 로그를 실시간으로 수집해서 주행 패턴을 분석한다.

## 작업내용

- 1) 요구사항1에 해당될 수 있는 요구사항 명세서를 작성하시오.

요구 사항명	고유 번호	분류	요구 사항 정의 및 세부 내용	산출정보	관련요구사항
시스템 구성 환경 (H/W, S/W)	ECR- 001	시스템 장비	DB 관리 <ul style="list-style-type: none"><li>- 3-tier 구조로 구축할 것</li><li>- 서비스에 필요한 모든 소프트웨어에 대한 라이선스 및 설치 기술료 등 포함 함</li><li>- 오픈 소스 기반 기술을 사용할 경우, 지속적인 업그레이드 및 관리 안을 제시하여야 함</li></ul>		
빅데이 터 플 랫폼	SFR- 001	기능 - 수집	통합 데이터베이스 구축 <ul style="list-style-type: none"><li>- 기간계 데이터의 데이터베이스화: 1MB/1대 (1일 수집규모: 약 100MB/100대)</li><li>- 기간계 데이터의 주기적 수집 및 저장: 3초 / 24시간</li><li>- 수집 데이터의 정제하기 위한 데이터 베이스 영역 개발</li><li>- 기간계 데이터 수집, 표준화, 중복 제거, 배포 위한 정책</li></ul>		데이터베이스, 표준 가이드 라인
성능 검사 실시	PER- 001	성능	성능 검사 실시 <ul style="list-style-type: none"><li>- 측정 대상: 관련 대상 서버</li><li>- 테스트 결과를 보고하고, 보완 사항을 조치할 것</li></ul>		개발 문서, ECR-001, ECR- 002
시스템	UIR-	비기능-인터페	시스템 인터페이스 일반		

인터페이스 일반	001	이스	<ul style="list-style-type: none"> <li>- 시스템 성능에 미치는 영향을 고려하여 갱신 주기 등을 설정하여야 함.</li> <li>- 정보 증가 시에도 안정적인 성능을 제공하여야 하며, 추가 라이선스 비용 등이 발생하지 않아야 함</li> </ul>
데이터 분석, 수집, 입력 요구 사항	DAR-001	비기능-데이터	<p>데이터 분석, 수집, 입력 요구 사항</p> <ul style="list-style-type: none"> <li>- 데이터 현황 자료 산출 및 정보 분석을 지원(수시)하여야 함: 100대의 시범 운영 차량</li> <li>- 데이터 타입 및 처리유형: CSV 반정형 파일 / 배치</li> </ul>
테스트 방안	TER-001	비기능-테스트	<p>테스트</p> <ul style="list-style-type: none"> <li>- 빅데이터 플랫폼이 제공하기로 한 요구 사항을 모두 제공하여, 초기 의한 요구 사항에서 변경 관리 절차를 통해 승인을 획득한 구 사항을 최종 베이스라인(baseline)으로 간주함.</li> <li>- 제공하기로 한 요구 사항을 제공하는지 여부는 각 기능 요구 사항의 검증(테스트) 활동을 통해 예상된 결과가 도출되었을 경우를 기준으로 평가함.</li> </ul>
보안 지침 준수	SER-001	비기능-보안	<p>기관 및 기업의 보안 규정 및 지침을 준수하여 개발이 수행되어야 함</p>
기능 구현 정확성	QAR-001	품질	<p>기능 구현 정확성</p> <ul style="list-style-type: none"> <li>- 빅데이터 플랫폼은 제공하기로 한 요구 사항을 모두 제공하며, 초기 협의한 요구 사항에서 변경 관리 절차를 통해 승인을 획득한 구 사항을 최종 베이스라인(baseline)으로 간주함.</li> <li>- 제공하기로 한 요구 사항을 제공하는지 여부는 각 기능 요구 사항의 검증(테스트) 활동을 통해 예상된 결과가 도출되었을 경우 구 사항을 제공한 것으로 평가함.</li> </ul>
업무 모듈화 및 자 원 활 용 방 안	COR-001	제약 사항	<p>업무 모듈화 및 자원 활용 방안</p> <ul style="list-style-type: none"> <li>- 유연성, 확장성을 확보할 수 있도록 모듈화 개발 전략을 반영함.</li> </ul>

2) 요구사항2에 해당될 수 있는 요구사항 명세서를 작성하시오.

요구 사항명	고유 번호	분류	요구 사항 정의 및 세부 내용	산출정보	관련요구사항
DB 관 리 소 프트웨 어	ECR- 002	시스템 장비	DB 관리 <ul style="list-style-type: none"> <li>- 장비 품목: DBMS</li> <li>- 장비 수량: 3식</li> <li>- 장비 기능: DB 관리</li> <li>- 장비 성능 및 특징: 개방형(Open) 기 반의 DB 처리 능력 지원 및 신 표준 SQL 규격 지원</li> <li>- 오픈 소스 기반 기술을 사용할 경우, 지속적인 업그레이드 및 관리 안을 제시하여야 함</li> </ul>		
빅데이 터 플 랫폼	SFR- 002	기능 - 수집	통합 데이터베이스 구축 <ul style="list-style-type: none"> <li>- 기간계 데이터의 데이터베이스화: 기 간계 데이터의 데이터베이스화: 4KB/1대 (초당 수집규모 : 약 400KB/100대)</li> <li>- 기간계 데이터의 주기적 수집 및 저 장: 1초 / 실시간</li> <li>- 수집 데이터의 정제하기 위한 데이터 베이스 영역 개발</li> <li>- 기간계 데이터 수집, 표준화, 중복 제 거, 배포 위한 정책</li> </ul>		데이터베이스, 표준 가이드 라인
성능 검사 실시	PER- 002	성능	성능 검사 실시 <ul style="list-style-type: none"> <li>- 측정 대상: 관련 대상 서버</li> <li>- 테스트 결과를 보고하고, 보완 사항 을 조치할 것</li> </ul>		
시스템 인터페 이스 일반	UIR- 002	비기능-인터페 이스	시스템 인터페이스 일반 <ul style="list-style-type: none"> <li>- 시스템 성능에 미치는 영향을 고려하 여 갱신 주기 등을 설정하여야 함.</li> <li>- 정보 증가 시에도 안정적인 성능을 제공하여야 하며, 추가 라이선스 비 용 등이 발생하지 않아야 함</li> </ul>		
데이터 분석, 수집, 입력 요구 사항	DAR- 002	비기능-데이터	데이터 분석, 수집, 입력 요구 사항 <ul style="list-style-type: none"> <li>- 데이터 현황 자료 산출 및 정보 분석 을 지원(수시)하여야 함: 100대의 시 범 운영 차량</li> <li>- 데이터 타입 및 처리유형: CSV 반정 형 파일 / 배치</li> </ul>		
테스트 방안	TER- 002	비기능-테스트	테스트 <ul style="list-style-type: none"> <li>- 빅데이터 플랫폼이 제공하기로 한 요</li> </ul>		

			<p>구 사항을 모두 제공하여, 초기 의한 요구 사항에서 변경 관리 절차를 통해 승인을 획득한 구 사항을 최종 베이스라인(baseline)으로 간주함.</p> <ul style="list-style-type: none"> <li>- 제공하기로 한 요구 사항을 제공하는지 여부는 각 기능 요구 사항의 검증(테스트) 활동을 통해 예상된 결과가 도출되었을 경우를 기준으로 평가함.</li> </ul>
보안 지침 준수	SER-002	비기능-보안	<p>기관 및 기업의 보안 규정 및 지침을 준수하여 개발이 수행되어야 함</p>
기능 구현 정확성	QAR-002	품질	<p>기능 구현 정확성</p> <ul style="list-style-type: none"> <li>- 빅데이터 플랫폼은 제공하기로 한 요구 사항을 모두 제공하며, 초기 협의한 요구 사항에서 변경 관리 절차를 통해 승인을 획득한 구 사항을 최종 베이스라인(baseline)으로 간주함.</li> <li>- 제공하기로 한 요구 사항을 제공하는지 여부는 각 기능 요구 사항의 검증(테스트) 활동을 통해 예상된 결과가 도출되었을 경우 구 사항을 제공한 것으로 평가함.</li> </ul>
시스템 구조 설계	COR-002	제약 사항	<p>시스템 구조 설계</p> <ul style="list-style-type: none"> <li>- 기 구축되어 있는 현행 시스템 구조 및 전체 표준과 호환성, 시스템 분산 설계, 데이터 유형, 프로세스 환경 유형, 사용자 유형, 시스템 토폴로지(topology)를 고려하여 구조 설계까지 이루어져야 함.</li> </ul>

# 빅데이터 플랫폼 요구사항 검증하기

## 작업내용

위 빅데이터 플랫폼 요구사항에 따라 빅데이터 플랫폼을 구축하시오.

### 1) 가상 머신 환경구성

빅데이터 파일럿 서버를 구성하기 위해 VirtualBox를 설치하고, 이를 통해 3대의 가상머신을 생성한다. 이 3대의 가상머신에 CentOS 리눅스를 설치한다. 가상머신의 CPU와 메모리 리소스 설정을 파일럿 프로젝트에 맞게 설정한다. 메모리의 경우 아래 표를 참고해서 본인의 PC 환경에 맞춰 재설정한다.

#### 1-1 가상 머신 생성하기

Step01. VirtualBox 프로그램으로 'cm201', 'cm202', 'cm203' 이름으로 총 3대의 가상 머신 생성

Step02. 3대의 가상머신에 HDD 용량 모두 40GB 설정 아래 표를 참고하여 메모리를 할당 · 가상머신 메모리 할당

설치 PC 환경	cm201	cm202	cm203
저사양 PC CPU:듀얼코어, RAM:8GB	4096MB(4GB)	2048MB(2GB)	N/A
고사양 PC CPU:i5 이상, RAM:16GB	10240MB(10GB)	1024MB(1GB)	1024MB(1GB)

### 2) CentOS 설치 및 설정

#### 2-1 CentOS 설치하기

Step01. CentOS 7버전, 최소설치, 기본 파티션, Kdump 해제

Step02. 최고관리자 root 비밀번호 설정, 'bigdata' 관리자 계정추가

#### 2-2 CentOS 설정하기

Step01. 가상머신 고정 IP 할당(모든 시스템 동일)

```
#cd /etc/sysconfig/network-scripts/
#cp ifcfg-enp0s3 ifcfg-enp0s8
#vi ifcfg-enp0s8
- 수정 : BOOTPROTO="none"
- 수정 : NAME="enp0s8"
- 수정 : DEVICE="enp0s8"
- 추가 : IPADDR="192.168.56.xxx"
- 추가 : NETMASK="255.255.255.0"
```

Step02. 네트워크 재시작, dhclient 실행, ping 테스트

```
#systemctl restart network
#dhclient
#ping google.com (핑테스트 종료 Ctrl+c)
```

Step03. 기본 업데이트 진행, 필수 패키지 설치(모든 시스템 동일)

```
#yum -y update
#yum install -y vim wget net-tools java-1.8*
```

Step04. SELINUX 해제(모든 시스템 동일)

```
#vi /etc/selinux/config
7라인 : SELINUX=disabled
```

Step05. NetworkManager 끄기(모든 시스템 동일)

```
#systemctl status NetworkManager
#systemctl stop NetworkManager
#systemctl disable NetworkManager
```

Step06. 방화벽 해제(모든 시스템 동일)

```
#systemctl status firewalld
#systemctl stop firewalld
#systemctl disable firewalld
```

Step07. host 설정(모든 시스템 동일)

```
#vi /etc/hostname // 아래 3개 이름 중 하나만 입력
cm201 // 만약 192.168.xxx.201
cm202 // 만약 192.168.xxx.202
cm203 // 만약 192.168.xxx.203
```

Step08. hosts 설정(모든 시스템 동일)

```
#vi /etc/hosts // 기존 내용 모두 삭제후 아래 내용 그대로 추가
192.168.xxx.201 cm201
192.168.xxx.202 cm202
192.168.xxx.203 cm203
```

Step09. 재부팅 후 putty 접속

```
#reboot
```



### 3) Cloudera Manager 설치

CM(Cloudera Manager)은 빅데이터 에코시스템을 쉽게 설치하고 관리해주는 클라우드라의 강력한 빅데이터 시스템 솔루션이다. 파일럿 프로젝트를 진행하면서 필요한 소프트웨어들을 CM을 통해 설치 및 설정 한다.

#### 3-1 CentOS 설치하기

##### Step01. CM 설치하기(cm201 서버만 적용)

```
#wget https://archive.cloudera.com/cm7/7.1.4/cloudera-manager-installer.bin
#chmod u+x cloudera-manager-installer.bin - 권한을 부여한다
#sudo ./cloudera-manager-installer.bin - 파일을 실행한다
```

```
[root@cm201 ~]# wget https://archive.cloudera.com/cm7/7.1.4/cloudera-manager-installer.bin
--2021-02-17 11:37:21-- https://archive.cloudera.com/cm7/7.1.4/cloudera-manager-installer.bin
Resolving archive.cloudera.com (archive.cloudera.com)... 151.101.108.167
Connecting to archive.cloudera.com (archive.cloudera.com)|151.101.108.167|:443..
. connected.
HTTP request sent, awaiting response... 200 OK
Length: 830310 (811K) [application/octet-stream]
Saving to: 'cloudera-manager-installer.bin'

100%[=====>] 830,310      2.32MB/s   in 0.3s

2021-02-17 11:37:22 (2.32 MB/s) - 'cloudera-manager-installer.bin' saved [830310/830310]

[root@cm201 ~]# ll
합 계 816
-rw-----. 1 root root 1387 2월 15 15:02 anaconda-ks.cfg
-rw-r--r-- 1 root root 830310 10월 15 06:03 cloudera-manager-installer.bin
[root@cm201 ~]# chmod u+x cloudera-manager-installer.bin
[root@cm201 ~]# ll
합 계 816
-rw-----. 1 root root 1387 2월 15 15:02 anaconda-ks.cfg
-rwxr--r-- 1 root root 830310 10월 15 06:03 cloudera-manager-installer.bin
[root@cm201 ~]# sudo ./cloudera-manager-installer.bin
[root@cm201 ~]# ll
합 계 816
-rw-----. 1 root root 1387 2월 15 15:02 anaconda-ks.cfg
-rwxr--r-- 1 root root 830310 10월 15 06:03 cloudera-manager-installer.bin
```

[그림 1] CM 다운로드 화면

##### Step02. CM PostgreSQL DB 설정

```
#vi /var/lib/cloudera-scm-server-db/data/pg_hba.conf
마지막 밑에서 3번째 줄 'reject' → 'md5' 수정
맨 마지막 줄에 새로 추가 → host all all 0.0.0.0/0 trust
```

```

77 # TYPE DATABASE USER ADDRESS METHOD
78
79 # "local" is for Unix domain socket connections only
80 local all all md5
81 # IPv4 local connections:
82 host all all 127.0.0.1/32 md5
83 # IPv6 local connections:
84 host all all ::1/128 md5
85 # Allow replication connections from localhost, by a user with the
86 # replication privilege.
87 local replication all md5
88 host replication all 127.0.0.1/32 md5
89 host replication all ::1/128 md5
90 # Accept connections from all hosts
91 host amon amon 0.0.0.0/0 md5
92 host rman rman 0.0.0.0/0 md5
93 host nav nav 0.0.0.0/0 md5
94 host navms navms 0.0.0.0/0 md5
95 # block remote access for admin user and scm
96 host all cloudera-scm,scm 0.0.0.0/0 md5
97 # enable remote access for other users
98 host sameuser all 0.0.0.0/0 md5
99 host all all 0.0.0.0/0 trust

```

99,1      바    닉

[그림 2] CM PostgreSQL DB 설정 화면

Step03. CM PostgreSQL DB 변경사항 적용을 위해 재가동 및 CM 실행

```

#systemctl stop cloudera-scm-server
#systemctl restart cloudera-scm-server-db
#systemctl restart cloudera-scm-server

```

```

[root@cm201 ~]# systemctl stop cloudera-scm-server
[root@cm201 ~]# systemctl restart cloudera-scm-server-db
[root@cm201 ~]# systemctl restart cloudera-scm-server
[root@cm201 ~]# █

```

[그림 3] CM 실행 화면

Step04. CM 브라우저 접속확인(CM서버가 기동하는데 2~3분 걸림)

주소 : http://192.168.xxx.201:7180

계정 : admin

암호 : admin

### 3-3 CentOS 설정하기

Step01. CM 로그인하기

주소 : http://192.168.xxx.201:7180

계정 : admin

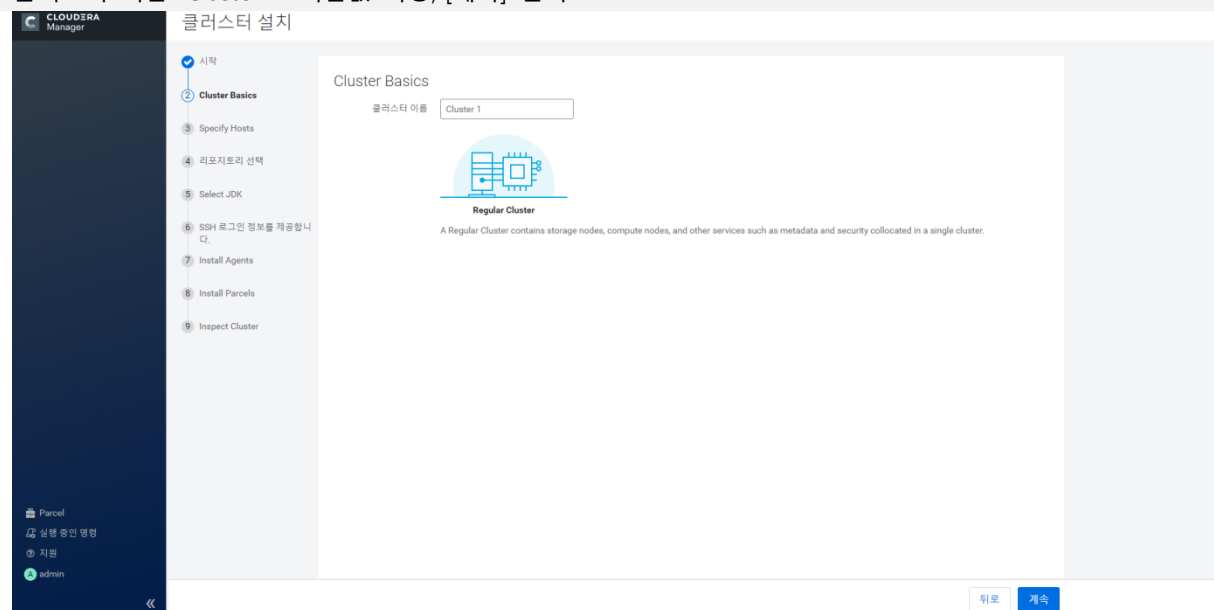
암호 : admin

Step02. CM 동의체크 후 [계속] 클릭

Step03. CM 클러스터 설치 시작 [계속] 클릭

Step04. Cluster Basics

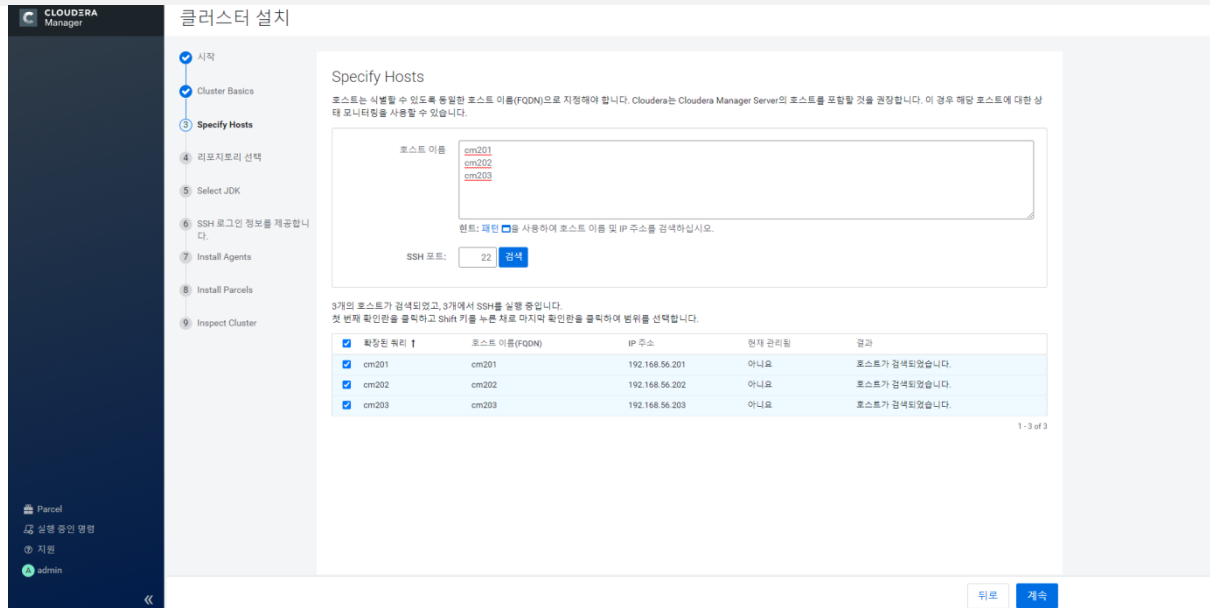
클러스터 이름 'Cluster 1' 기본값 사용, [계속] 클릭



[그림 4] CM 클러스터 설치 화면 – Cluster Basics

## Step05. Specify Host

### 호스트 이름 입력 후 검색



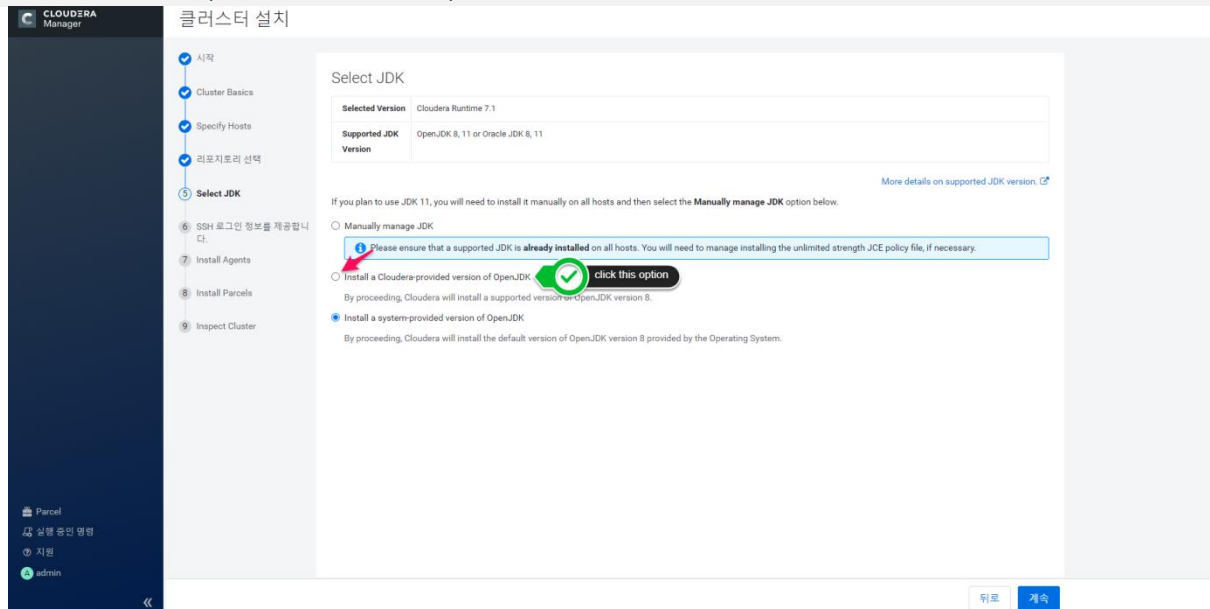
[그림 5] CM 클러스터 설치 화면 – Specify Hosts

## Step06. 리포지토리 선택

### 변경사항 없이 기본값으로 [계속] 클릭

## Step07. JDK 설치옵션

### Install a Cloudera-provided version of OpenJDK를 선택 후 [계속] 클릭



[그림 6] CM 클러스터 설치 화면 – Select JDK

## Step08. SSH 로그인 정보 제공 설정

리눅스 root 비밀번호 입력 후 [계속] 클릭



클러스터 설치

시작

Cluster Basics

Specify Hosts

리포지토리 선택

Select JDK

SSH 로그인 정보를 제공합니다.

Cloudera 패키지를 설치하려면 호스트에 대한 루트 액세스가 필요합니다. 이 설치 관리자는 SSH를 통해 호스트에 연결하고 루트로 직접 로그인하거나 암호 없이 sudo/pbrun 권한을 가진 다른 사용자로 로그인하여 후드가 됩니다.

모든 호스트를 다음으로 로그인: ☒ root ☐ 다른 사용자

위에서 선택한 사용자에 대한 암호 또는 공용 키 인증을 통해 연결할 수 있습니다.

인증 방법: ☒ 모든 호스트가 동일한 암호 허용 ☐ 모든 호스트가 동일한 개인 키 허용

암호 입력:

암호 확인:

SSH 포트:

동시에 진행하는 설치 수:   
(많은 설치를 한꺼번에 실행하면 많은 양의 네트워크 대역폭 및 다른 시스템 리소스가 소모됩니다.)

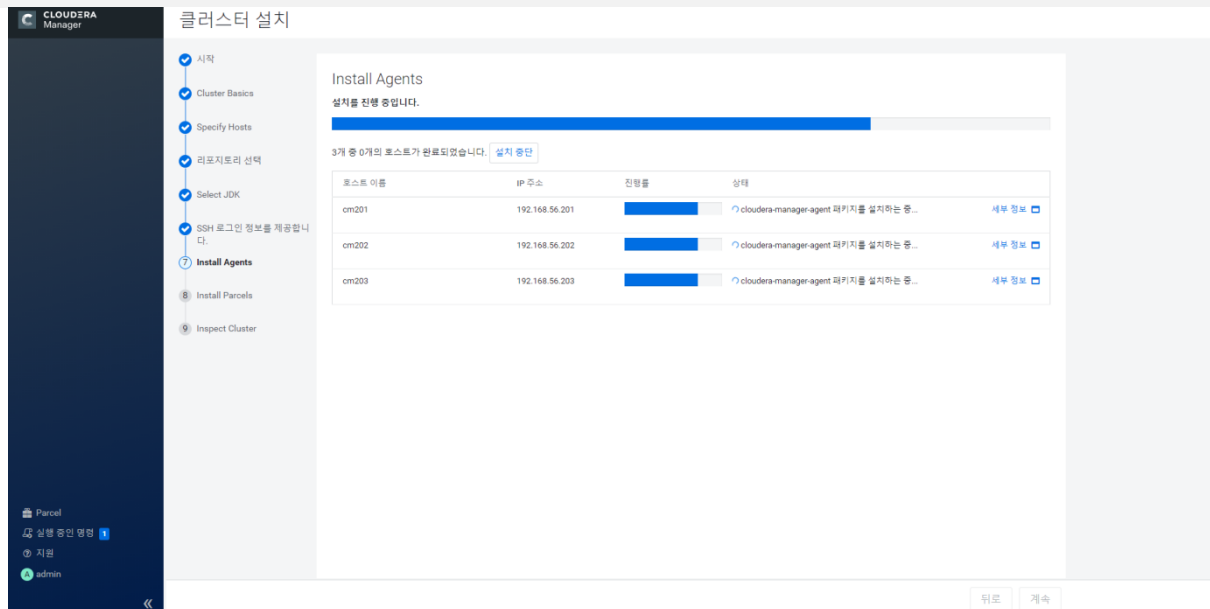
뒤로 계속

[그림 7] CM 클러스터 설치 화면 – SSH 로그인 정보 제공

## Step09. Install Agents

Cloudera Manager 관련 패키지 설치 진행, 설치 완료 후 [계속] 클릭

최소 20분 소요, 네트워크 상태에 따라 최대 40분 소요



클러스터 설치

시작

Cluster Basics

Specify Hosts

리포지토리 선택

Select JDK

SSH 로그인 정보를 제공합니다.

Install Agents

설치를 진행 중입니다.

3개 중 0개의 호스트가 완료되었습니다. 설치 중단

호스트 이름	IP 주소	진행률	상태
cm201	192.168.56.201	<div></div>	cloudera-manager-agent 패키지를 설치하는 중... 세부 정보
cm202	192.168.56.202	<div></div>	cloudera-manager-agent 패키지를 설치하는 중... 세부 정보
cm203	192.168.56.203	<div></div>	cloudera-manager-agent 패키지를 설치하는 중... 세부 정보

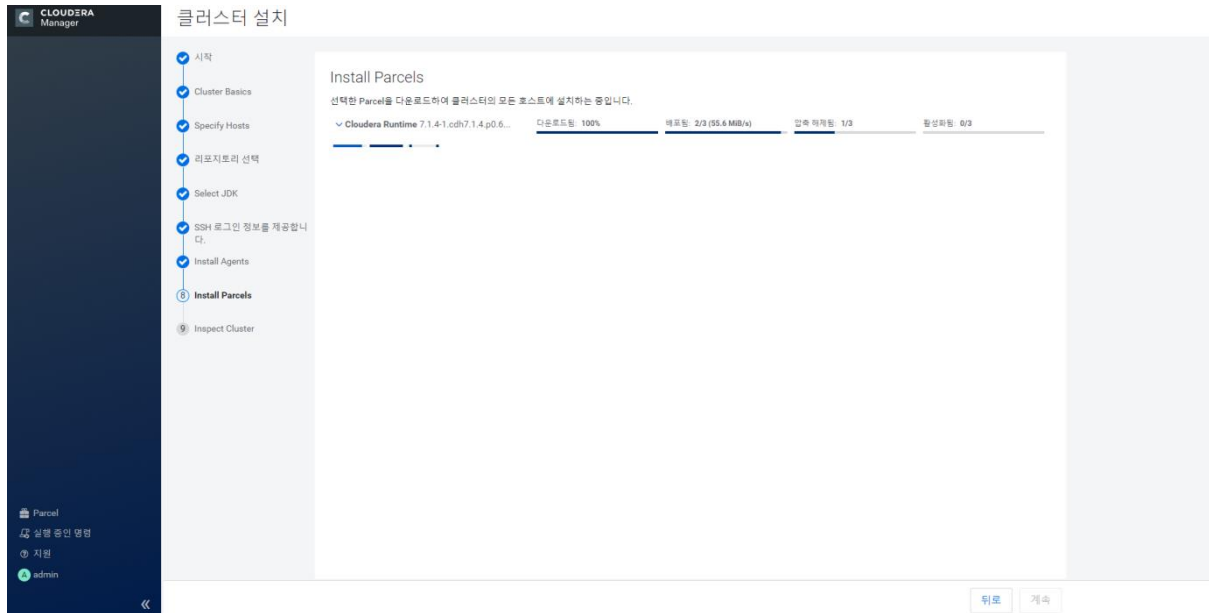
뒤로 계속

[그림 8] CM 클러스터 설치 화면 – Install Agents

## Step10. Install Parcels

Cloudera Manager 추가 패키지 설치 진행, 설치 완료 후 [계속] 클릭

최소 20분 소요, 네트워크 상태에 따라 최대 40분 소요

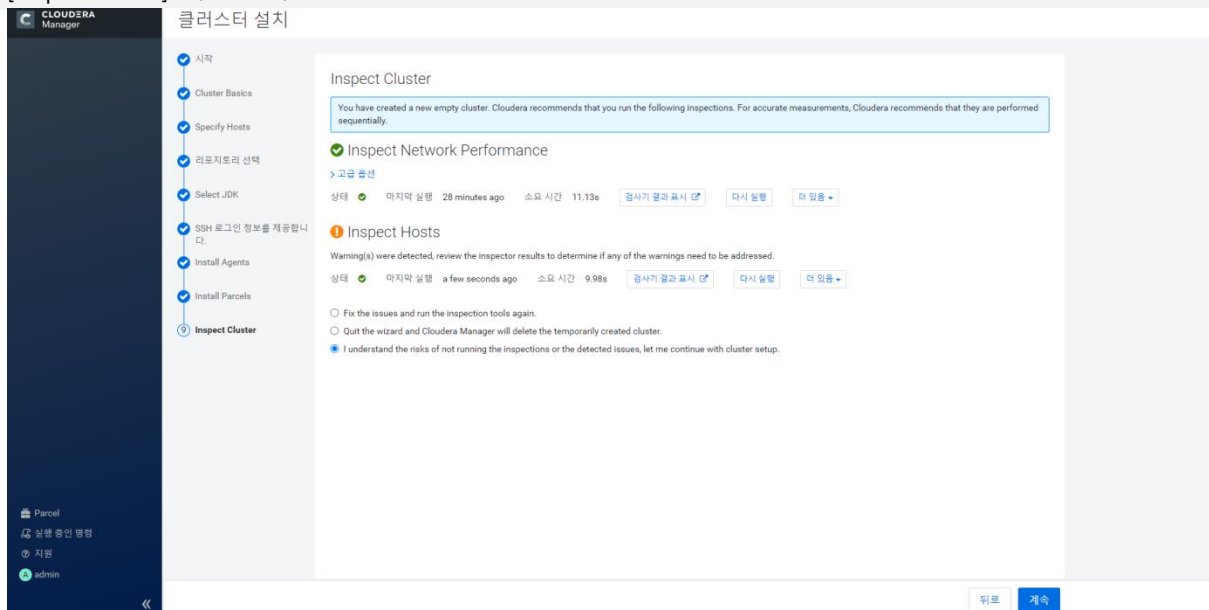


[그림 9] CM 클러스터 설치 화면 – Install Parcels

## Step11. Install Cluster

[Inspect Network Performance] 버튼 클릭 점검

[Inspect Hosts] 버튼 클릭 점검



[그림 10] CM 클러스터 설치 화면 – Install Cluster

Inspect Hosts Warning 발생, 검사기 결과표시 클릭해서 내용 확인

시스템 Swap 메모리 설정(모든 시스템 적용)

```
#sysctl -w vm.swappiness=10
```

```
#vi /etc/sysctl.conf
```

마지막 줄에

```
vm.swappiness=10 ← 추가
```

### 시스템 kernel hugePage 설정(모든 시스템 적용)

```
#echo never > /sys/kernel/mm/transparent_hugepage/defrag
#echo never > /sys/kernel/mm/transparent_hugepage/enabled
#vi /etc/rc.local
마지막 줄에 아래 2줄 추가
echo never > /sys/kernel/mm/transparent_hugepage/defrag
echo never > /sys/kernel/mm/transparent_hugepage/enabled
```

### Step12. 클러스터 서비스 선택 및 역할 할당 사용자 지정

#### [HBase]

- Master : cm201 선택
- HBase REST Server : 미설치
- HBase Thrift Server : cm201 선택
- RegionServer : DataNode(으)로 저장

#### [HDFS]

- NameNode : cm201 선택
- SecondaryNameNode : cm201 선택
- Balancer : cm201 선택
- HttpFS : 미설치
- NFS Gateway : 미설치
- DataNode : 모든 호스트

#### [Hive]

- Gateway : cm201 선택
- Hive Metastore Server : cm201 선택
- WebHCat Server : 미설치
- HiveServer2 : cm201 선택

#### [Hue]

- Hue Server : cm201 선택
- Load Balancer : 미설치

#### [Cloudera Management Service]

- Service Monitor : cm201 선택
- Activity Monitor : 미설치
- Host Monitor : cm201 선택
- Event Monitor : cm201 선택
- Alert Publisher : cm201 선택

#### [Oozie]

- Oozie Server : cm201 선택

#### [Spark]

- History Server : cm201 선택
- Gateway : cm201 선택

#### [YARN (MR2 Included)]

- ResourceManager : cm201 선택

- JobHistory Server : cm201 선택
- NodeManager : 'DataNode(으)로 저장' 선택  
[ZooKeeper]
- Server : cm201 선택

#### Step13. 데이터베이스 설정

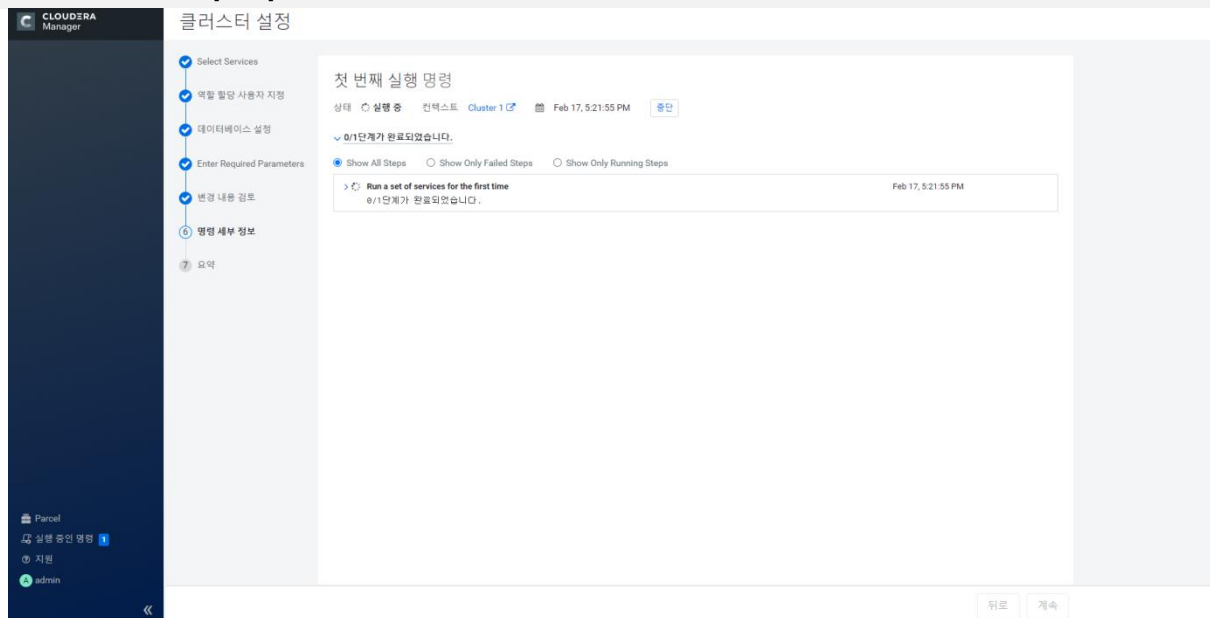
변경사항 없음, [테스트 연결] 클릭 후 [계속] 클릭

#### Step14. 변경 내용 검토

변경사항 없음, [계속] 클릭

#### Step15. 첫 번째 실행 명령

설치가 완료된 후 [계속] 클릭



[그림 11] CM 클러스터 설치 화면 - 명령 세부 정보

#### Step16. 요약 [완료] 클릭



[그림 12] CM 클러스터 설치 화면 - 완료