

2.5 빅데이터 클러스터 구성

DataNode 추가 및 환경설정

DataNode 추가




상태	이름	IP	역할
 ●	server01.hadoop.com	192.168.56.101	<div>▼ 4 Role(s)</div> <ul style="list-style-type: none"> ● HDFS NameNode ● HDFS SecondaryNameNode ● YARN (MR2 Included) JobHistory Server ● YARN (MR2 Included) ResourceManager
 ●	server02.hadoop.com	192.168.56.102	<div>▼ 3 Role(s)</div> <ul style="list-style-type: none"> ● HDFS DataNode ● YARN (MR2 Included) NodeManager ● ZooKeeper Server
 ●	server03.hadoop.com	192.168.56.103	<div>▼ 7 Role(s)</div> <ul style="list-style-type: none"> ● HDFS DataNode ● Cloudera Management Service Alert Publisher ● Cloudera Management Service Event Server

그림 2.89 CM의 호스트 정보 보기

01. CM 홈에서 [HDFS] 메뉴 우측의 콤보박스를 클릭한 후 [역할 인스턴스 추가] 메뉴를 선택한다.

저사양 파일럿 환경: 다음의 데이터 노드 추가 작업을 생략한다.

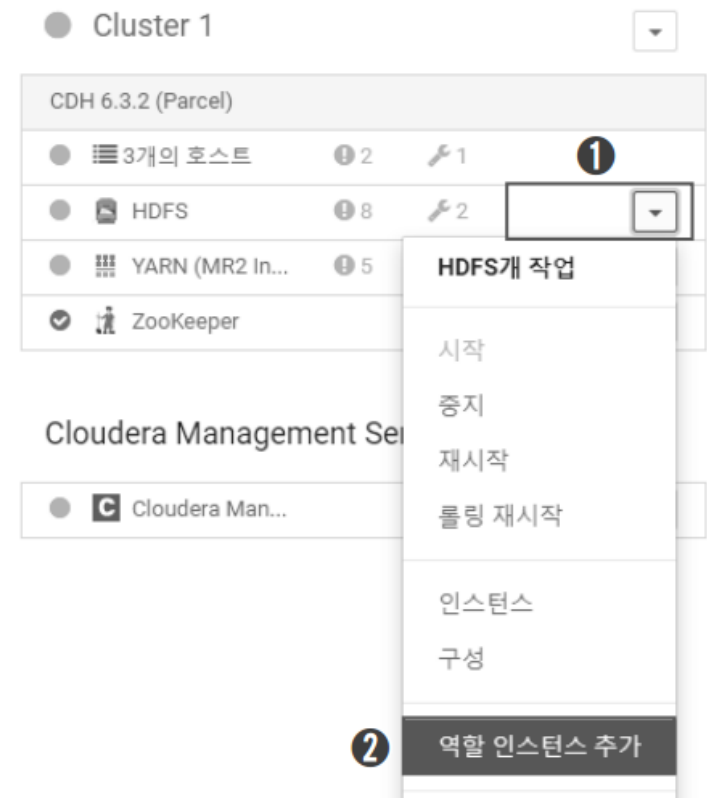


그림 2.90 하둡 데이터노드 추가 – 역할 인스턴스 추가

역할 할당 사용자 지정

여기에서 새 역할에 대한 역할 할당을 사용자 지정할 수 있습니다.

역할 할당을 호스트별로 볼 수도 있습니다.

호스트별로 보기

Gateway

호스트 선택

HttpFS

호스트 선택

SecondaryNameNode x 1

호스트 선택

NFS Gateway

호스트 선택

JournalNode

호스트 선택

NameNode x 1

호스트 선택

Balancer

호스트 선택

Failover Controller

호스트 선택

DataNode x 2

호스트 선택

모든 호스트

사용자 지정...

1

<input checked="" type="checkbox"/> 호스트 이름	IP 주소	랙	코어	물리적 메모리	기존 역할	추가된 역할
<input checked="" type="checkbox"/> server01.hadoop.com	192.168.56.101	/default	1	4.9 GiB	NN SNN JHS RM	DN
<input checked="" type="checkbox"/> server02.hadoop.com	192.168.56.102	/default	1	4.9 GiB	DN NM S	
<input checked="" type="checkbox"/> server03.hadoop.com	192.168.56.103	/default	1	2.9 GiB	DN AP ES HM RM SM NM	

2

Displaying 1 - 3 of 3

취소

3

확인

그림 2.91 하둡 데이터노드 추가 – 설치할 호스트 위치 지정

3 시작

중지

재시작

롤링 재시작

서비스 해제

서비스 재할당

Keytab 재생성

유지관리 모드 진입

2 선택된 작업 (1) ▼ 역할

인스턴스 추가

역할 그룹

	역할 유형	역할	역할 상태	서버	서비스 할당 상태	역할 그룹
1	<input checked="" type="checkbox"/> DataNode	중지됨	server02.hadoop.com	서비스 할당됨	DataNode Default Group	
	<input checked="" type="checkbox"/> DataNode	중지됨	server01.hadoop.com	서비스 할당됨	DataNode Default Group	
	<input type="checkbox"/> DataNode	시작됨	server03.hadoop.com	서비스 할당됨	DataNode Group 1	
	<input type="checkbox"/> NameNode (활성)	시작됨	server01.hadoop.com	서비스 할당됨	NameNode Default Group	
	<input type="checkbox"/> SecondaryNameNode	시작됨	server01.hadoop.com	서비스 할당됨	SecondaryNameNode Default Group	

그림 2.92 하둡 데이터노드 추가 – 추가한 데이터노드 재시작

2.5 빅데이터 클러스터 구성

 *DataNode* 추가 및 환경설정

환경설정

HDFS 복제 계수 설정

저사양 파일럿 환경: 복제 계수를 1로만 설정한다.

HDFS

작업▼ 1

상태

인스턴스

구성

명령

파일 찾아보기

차트 라이브러리

캐시 통계

복제 계수

필터

범위

복제 계수

dfs.replication

3

HDFS(서비스 전체) ↻

2

1 Edited Value

파일럿 환경 구성

4

변경 내용 저장

그림 2.93 하둡의 HDFS 구성 정보 설정 - 복제 계수 변경

- 복제 계수 변경: 3 → 2

Tip _ 복제 계수 증가로 분석 성능 향상

앞서 복제 계수를 줄이는 것과는 반대로 하둡에 이미 저장돼 있는 특정 파일에 대해 복제 계수를 강제로 증가시킬 수 있다. 그러면 여러 데이터노드에 복제 파일이 분산 저장되어 분석 작업 시 로컬리티와 병렬성을 높여 성능을 극대화할 수 있다.

예를 들어, 다음 명령을 통해 /user/hadoop 하위 경로의 모든 파일의 복제계수를 10으로 조정 할 수 있다.

- `hadoop fs -setrep 10 -R /user/hadoop/`

HDFS 접근 권한 해제

- HDFS 권한 검사: 해제

HDFS 블록 크기 변경

- HDFS 블록 크기 변경: 128 → 64

YARN 스케줄러와 리소스매니저의 메모리 설정

- YARN 스케줄러 메모리 변경: 1 → 1.5
- 리소스매니저 메모리 변경: 1 → 5

YARN 스케줄러 변경

- 변경 전: `org.apache.hadoop.yarn.server.resourcemanager.scheduler.fair.FairScheduler`
 - 변경 후: `org.apache.hadoop.yarn.server.resourcemanager.scheduler.fifo.FifoScheduler`
-

Cluster 1

CDH 6.3.2 (Parcel)

3개의 호스트
2
3
1

HDFS
11
2

YARN (MR2 In...
5

ZooKeeper

Cloudera Management Service

Cloudera Man...

파일: hadoop-conf/hdfs-site.xml
hdfs(1) 표시

```

...    ... @@ -23,13 +23,13 @@
23    23      <value>server01.hadoop.com:9870</value>
24    24      </property>
25    25      <property>
26    26        <name>dfs.replication</name>
27    -      <value>3</value>
27    +      <value>1</value>
28    28      </property>
29    29      <property>
30    30        <name>dfs.blocksize</name>
31    -      <value>134217728</value>
31    +      <value>67108864</value>
32    32      </property>
33    33      <property>
34    34        <name>dfs.client.use.datanode.hostname</name>
35    35      <value>>false</value>

```

2
클라이언트 구성 배포

그림 2.94 클라이언트 구성 배포

2.5 빅데이터 클러스터 구성

DataNode 추가 및 환경설정

DataNode 추가 및 환경설정 (실습)