1. \*\*Hadoop 설치\*\*:

- Hadoop 버전 선택

- Hadoop 다운로드 및 설치

- 환경 변수 설정 (예: `HADOOP\_HOME`)

2. \*\*하둡 클러스터 구성\*\*:

- 단일 노드 (로컬 모드) 설정

- 의사 분산 모드 (Pseudo-Distributed Mode) 설정

- 완전 분산 모드 (Fully-Distributed Mode) 설정

3. \*\*핵심 구성 파일 수정\*\*:

- `core-site.xml`: 핵심 설정

- `hdfs-site.xml`: HDFS 설정

- `mapred-site.xml`: MapReduce 설정

- `yarn-site.xml`: YARN 설정

4. \*\*환경 설정 파일 수정\*\*:

- `hadoop-env.sh`: 환경 변수 설정

5. \*\*HDFS 초기화 및 시작\*\*:

- 네임노드(NameNode) 및 데이터노드(DataNode) 설정

- HDFS 포맷

- HDFS 시작

6. \*\*YARN 설정 및 시작\*\*:

- 리소스 매니저(Resource Manager) 및 노드 매니저(Node Manager) 설정

- YARN 시작

7. \*\*네트워크 설정\*\*:

- 호스트명 및 IP 주소 설정

- `etc/hosts` 파일 수정

8. \*\*보안 설정\*\*:

- 사용자 및 그룹 설정

- 액세스 제어 목록 (ACL) 설정

- Kerberos 인증 설정

9. \*\*로깅 및 모니터링 설정\*\*:

- 로그 파일 위치 및 로그 레벨 설정

- Hadoop 관리 인터페이스 설정

10. \*\*백업 및 복구 설정\*\*:

- HDFS 스냅샷 설정

- 데이터 복구 옵션 설정

11. \*\*성능 최적화 및 튜닝\*\*:

- Hadoop 설정 파라미터 튜닝

- 하드웨어 최적화

12. \*\*문제 해결 및 디버깅\*\*:

- 로그 분석

- 일반적인 문제 해결 방법

13. \*\*추가 유틸리티 및 도구 설정\*\*:

- Hadoop 에코시스템 도구 설정 (예: Hive, Pig, HBase 등)

각 설정 단계에서는 정확한 설정 값을 정의하고, 설정의 영향을 이해하는 것이 중요하다.

**하둡 구성.**

**사용 가능한 자원**.  
: Desk top 1대. 램 10GB, 저장용량 SSD 100GB HDD 1TB  
: Notebook 1대. 램 4GB, 저장용량 SSD 50GB

**사용 예정.**  
Desk top 1대. 램 4GB 저장용량 50GB worker\_01(secondary namenode),   
 램 2GB 저장용량 50GB worker\_02

Notebook 1대. 램 4GB 저장용량 50GB master