



빅데이터 처리 계획과 시스템 구성

[요약 보충학습]

1. 빅데이터 처리 시스템

빅데이터 처리 시스템이란, 대량의 데이터를 수집, 관리, 유통, 분석을 처리하는 일련의 분산 병렬 처리 프레임워크를 말합니다. 빅데이터 처리를 통해 유용한 정보와 데이터 내면에 있는 지식을 찾아낼 수 있습니다. 이러한 정보를 찾기 위한 데이터의 가공 및 분석 과정의 전반을 지원하는 것이 빅데이터의 처리 기술입니다.

2. 빅데이터 처리 기술의 필요성

- ① 데이터의 볼륨 증가
- ② 데이터 발생 속도의 증가
- ③ 데이터 포맷의 다양성 증가
- ④ 새로운 처리 기술의 필요성

3. 결함 허용 시스템(FTS)의 운영 목적

모든 시스템은 장애 발생의 위험이 있으며 이는 애플리케이션 문제 또는 하드웨어 리소스의 문제로 인해 발생할 수 있습니다. 결함 허용 시스템(FTS: Fault Tolerance System)은 이러한 장애가 발생하더라도 시스템이 대체 시스템이나 고장 대응 체계를 통해 시스템 운영을 지속하는 것을 말합니다.

[요약 보충학습]

4. 빅데이터 처리의 인프라 및 소프트웨어의 구성

빅데이터 처리 시스템은 다양한 저장 시스템과 연계됩니다. 다양한 형태의 데이터를 실시간, 배치(Batch), 인메모리(In-Memory) 처리 방식에 처리 소프트웨어(플랫폼)를 구성해야 합니다. 그리고 처리 시스템을 구축하기 위한 하드웨어 군(클러스터)을 설치하는 데 개별적인 시스템 구축 방식과 클라우드 컴퓨팅을 사용한 방식으로 구성할 수 있습니다.

5. 하드웨어 노드의 구성

-작업 노드: 데이터 노드(Data Node), 노드 매니저(Node Manager)와 임팔라 서버 데몬(Impala Server Daemons) 등이 수행됩니다.

-마스터 노드: 네임 노드(Name Node), 스탠바이 네임 노드(Standby Name Node), 세컨더리 네임 노드(Secondary Name Node), 리소스 관리자(Resource Manager) 데몬들이 수행됩니다.