

**문항1. 데이터 분석과 빅데이터 분석의 차이점에 4가지에 대해서 설명하시오. (48점)**

빅 데이터는 많은 양의 복잡한 데이터로 기존의 데이터 처리 응용 프로그램 소프트웨어를 사용하여 처리하기가 어렵습니다. 데이터 분석은 유용한 정보를 발견하고 의사 결정을 지원한다는 목표로 데이터를 검사, 정리, 변형 및 모델링하는 프로세스입니다.

이러한 분석과 빅데이터 분석의 차이점을 보면,

1) 빅 데이터와 데이터 분석의 중요한 차이점은 사용법입니다. 빅 데이터의 사용은 시스템 병목 현상, 대규모 데이터 처리 시스템 및 확장성이 뛰어난 분산 시스템을 식별하는 것입니다. 데이터 분석의 사용은 결론을 내리고 결정을 내리고 중요한 비즈니스 통찰력을 얻는 것입니다.

2) 빅 데이터는 빅 데이터 전문가가 처리하는 반면 데이터 분석은 데이터 분석가가 처리합니다.

3) 빅 데이터 애널리스트는 프로그래밍, NoSQL 데이터베이스, Hadoop과 같은 분산 시스템 및 프레임워크에 대한 지식이 필요합니다. 반면 데이터 애널리스트는 프로그래밍, 통계 및 수학에 대한 지식이 있어야 합니다.

4) 금융 서비스, 통신, 정보 기술 및 소매 업계에서 큰 데이터가 발견되는 반면 데이터 분석은 비즈니스, 과학, 건강 관리, 에너지 관리 및 정보 기술에 사용됩니다.

빅 데이터와 데이터 분석의 차이점은 커다란 데이터는 대량의 복잡한 데이터이며 데이터 분석은 유용한 정보를 인식하고 의사 결정을 지원하기 위해 데이터를 조사, 변환 및 모델링하는 프로세스라는 것입니다. 요약하면, 데이터 분석은 큰 데이터에 적용되어 비즈니스 이득을 향상시키고 위험을 줄일 수 있습니다.

**문항2. 데이터 전처리 기술 3가지가 무엇인지 명칭을 작성하고, 그 개념에 대해 설명하시오.(필요시, 사용되는 세부적인 기술을 설명하시오.) (52점)**

데이터 전처리는 빅데이터 분석 목적에 적합한 데이터 목록작성, 확보 가능 여부를 점검하고 수집한 데이터를 저장소에 적재하기 위한 작업입니다.

전처리 기술을 3가지로 구분하면

- 1) 데이터 유형변환 : 데이터유형을 변환하거나 데이터 분석에 용이한 형태로 변환
  - 2) 데이터 여과 : 오류발견, 보정, 삭제 및 중복성 확인등 데이터 품질 향상
  - 3) 데이터 정제 : 결측치변환, 이상치 제거, 노이즈 데이터 교정, 비정형 데이터를 수집할 때 반드시 수행
- 입니다.