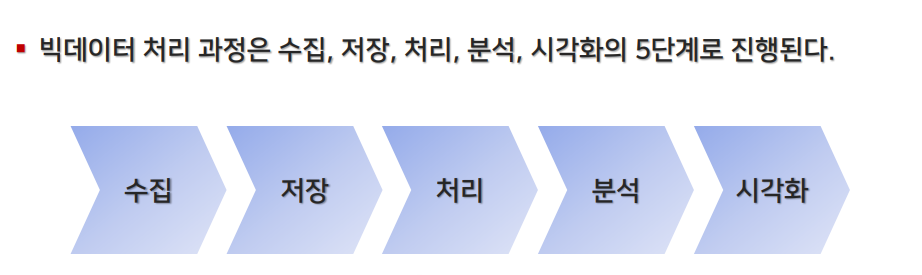
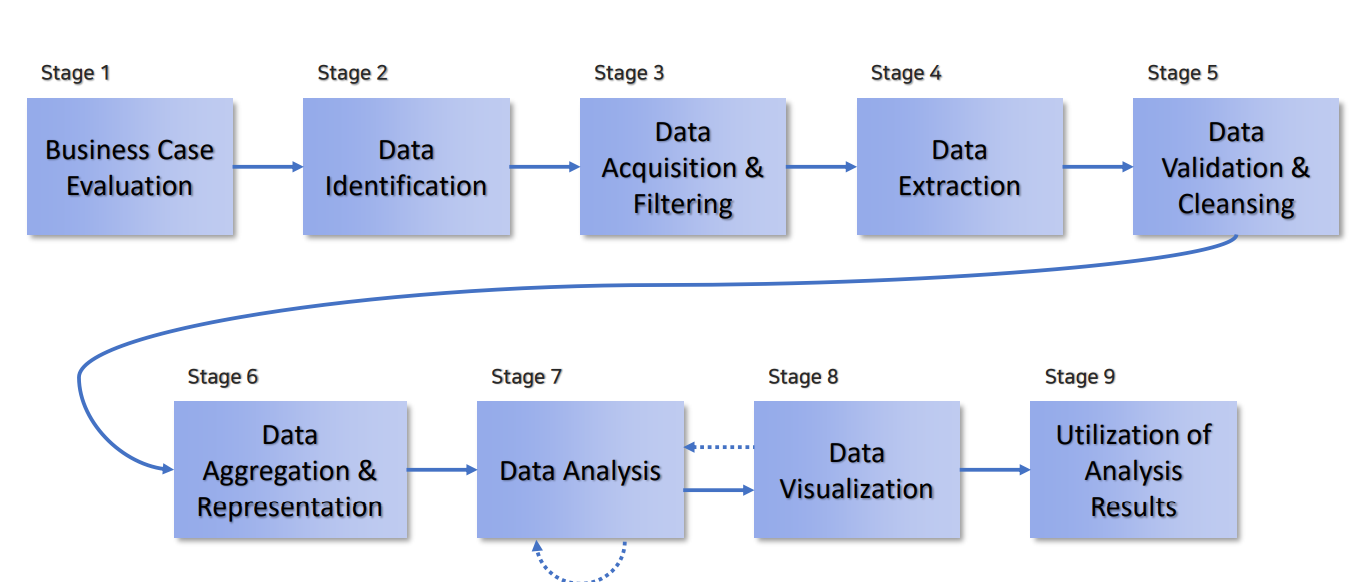
2022 빅데이터 컴퓨팅 (모두 서술형)

1. 빅데이터 규모적 기술적 특징?
   1. 규모적 정의: 전통적인 방법으로 저장, 관리 분석할 수 있는 범위를 초과하는 규모의 데이터 ▪ 기술적 정의: 다양한 종류의 데이터로부터 새로운 가치를 추출하고, 수집, 탐색, 분석을 지원하도록 고안된 차세대 기술
2. 5V란?
   1. Volume: zettabyte의 시대
   2. Variety: 데이터의 종류 증가
   3. Velocity: 데이터 생성 속도의 증가, 실시간 분석 필요
   4. Veracity: 데이터의 불확실성
   5. Value: 유의미한 정보 추출, 가치 창출
3. 빅데이터 처리 5단계 나열하세요
   1. 
4. 데이터사이언스란?
   1. 데이터 과학이란 정형, 비정형 형태를 포함한 다양한 데이터로부터 지식과 인사이트를 추출하는데 과학적 방법론, 프로세스, 알고리즘, 시스템을 동원하는 융합분야 ▪ 데이터 과학은 데이터를 통해 실제 현상을 이해하고 분석하는데 통계학, 데이터 분석, 기계학습과 연관된 방법론을 통합하는 개념으로 정의
5. 데이터의 불확실성이란?
6. Project Manager 역할은 ?
   1. 프로젝트에서 가장 중요한 사람

목표 달성을 위해 이정표를 확인하고 회의를 주관하는 사람

1. Data Sientist 의 역할은?
   1. 분석 기술, 데이터 모델링에 대한 전문 지식 지원 ▪ 전반적인 데이터 분석 목표가 달성되었는지 확인 ▪ 프로젝트에 분석 모델을 제시하고 데이터가 유효한지 확인
2. 빅데이터 분석 수명 주기 9단계 나열 하시오?



1. 3단계 설명
   1. ▪ 3단계: 데이터 습득과 여과 (Data Acquisition and Filtering) ▪ 확인된 다양한 데이터 출처(외부나 내부)에서 데이터를 획득 ▪ 외부의 데이터들은 노이즈가 있을 수 있으므로, 여과해야 한다. (removal) • 정제된 데이터들이 다른 유형의 분석에는 유용할 수 있으므로 사본을 유지 ▪ 메타데이터 추가 • 분류나 쿼리의 성능을 향상 • 메타데이터: 데이터셋의 크기, 구조, 출처, 생성시간, 수집시간, 언어 등의 정보
2. 5단계 설명
   1. 5단계: 데이터 검증 및 정제 (Data Validation and Cleansing) ▪ 빅데이터는 유효하지 않은 값을 가질 수 있음 (잘못된 데이터는 잘못된 결과를 야기) ▪ 업무 전문가들이 정한 규칙으로 검증하여 데이터 값의 유효성 여부를 판단 (valid, invalid) ▪ 검증하여 유효한 데이터 확보
3. 확증적 데이터 분석, 탐색적 데이터 분석 설명하시오
   1. 확증적 데이터 분석 (Confirmatory Data Analysis) • 연역적인 방법 • 가설을 제안하고 데이터 분석으로 가설을 입증(반증) ▪ 탐색적 데이터 분석 (Exploratory Data Analysis) • 귀납적인 방법 • 데이터마이닝과 관련된 방법 • 가설을 미리 만들지 않음 • 패턴이나 이상 현상을 쉽게 발견할 수 있는 일반적인 방향 제