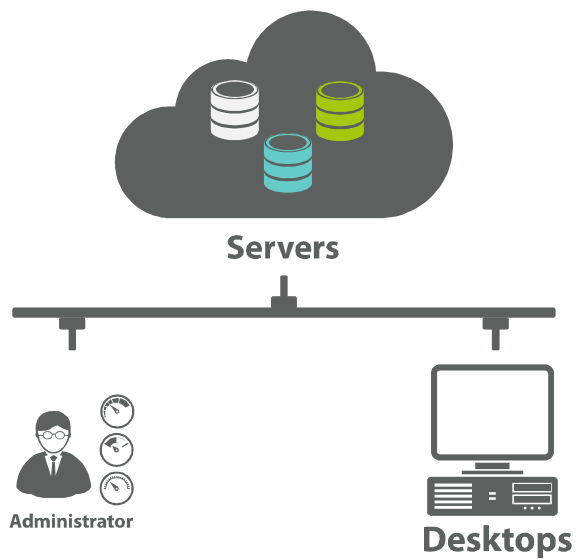




클라우드 기반 데이터레이크 및 분석

Introduction





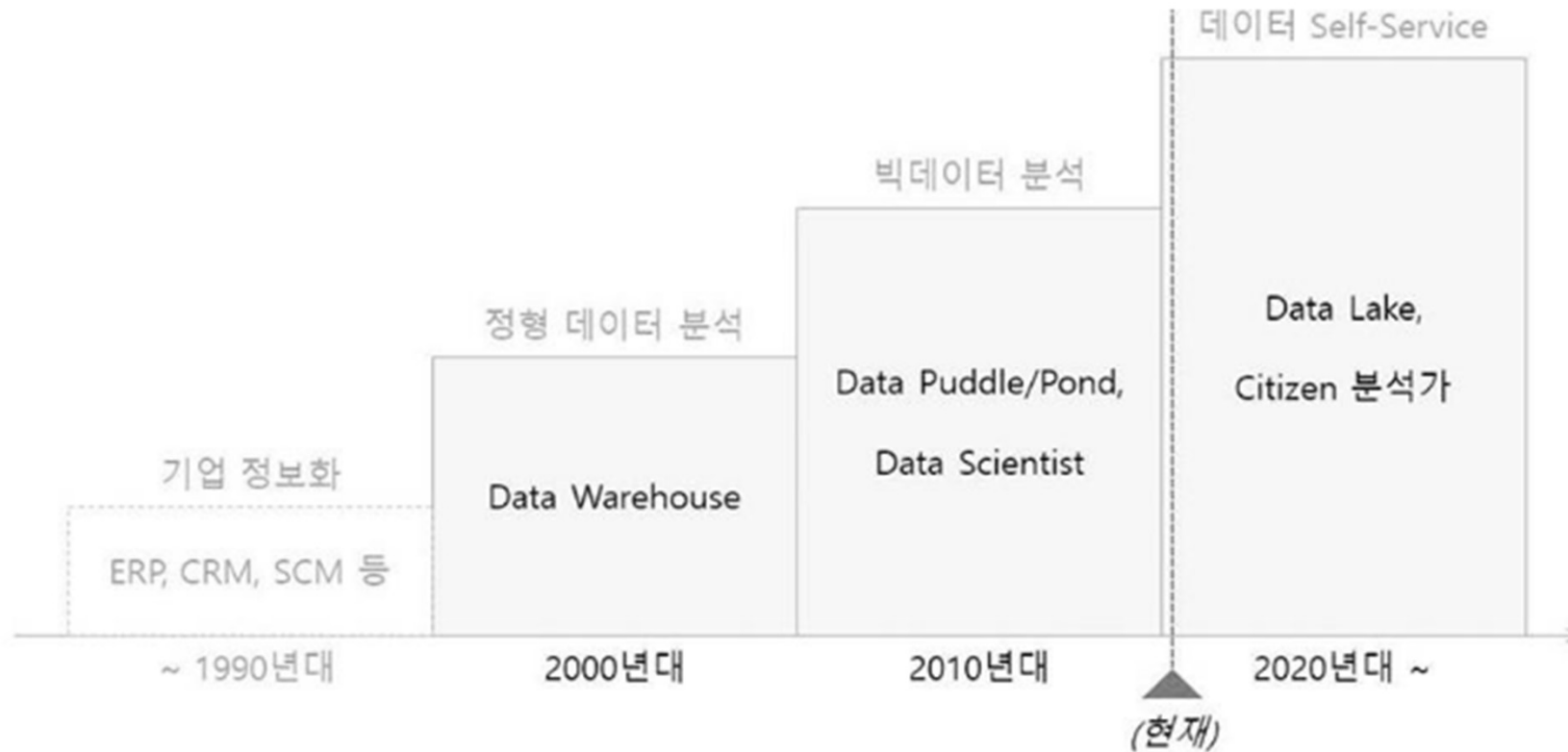
Index

- 01. 데이터 분석 플랫폼 발전 과정
- 02. Data Lake 구축 방식 선정
- 03. Data Lake Components

데이터 분석 플랫폼 발전 과정



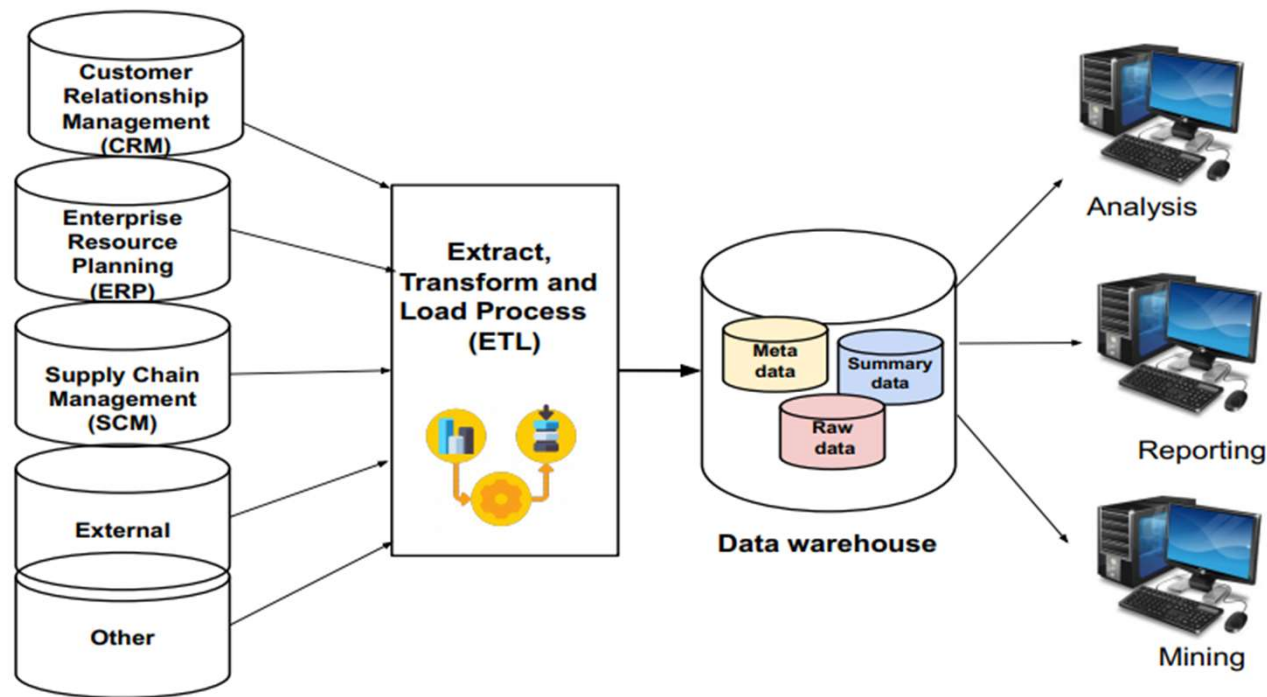
Data Analysis Platform Roadmap



Data Warehouse vs. Data Lake

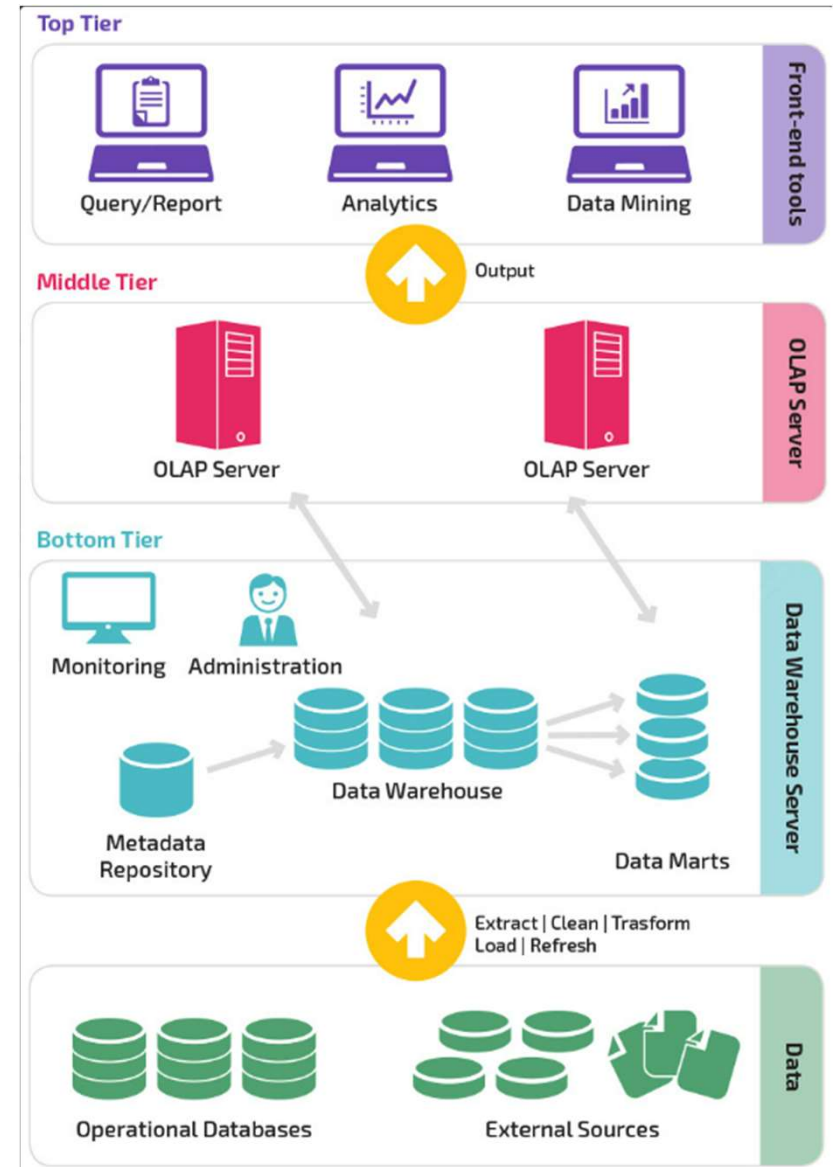
- Data Warehouse

- 정보(Data) + 창고(Warehouse)
- 의사 결정에 도움을 주기 위해 분석가능한 형태로 변환한 데이터들이 저장되어 있는 중앙저장소



Data Warehouse vs. Data Lake

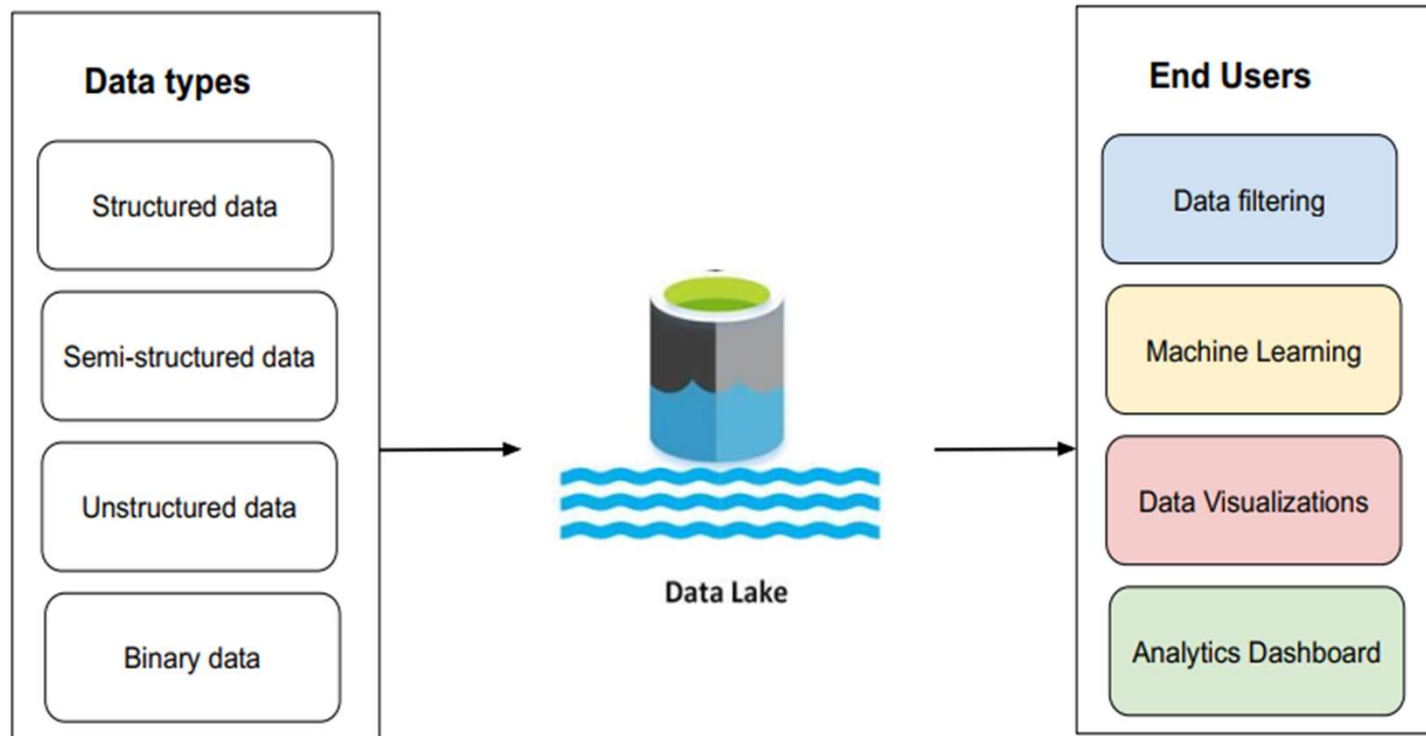
- Raw Data
- ETL
 - Extract: 하나 또는 그 이상의 데이터 원천들로부터 데이터 획득
 - Transform: 데이터 클렌징, 형식 변환 및 표준화, 통합 또는 다수 애플리케이션에 내장된 비즈니스를 적용
 - Load: 변형 단계의 처리가 완료된 데이터를 특정 목표 시스템에 적재
- Data Warehouse Server
- OLAP Server
- Front-end tools



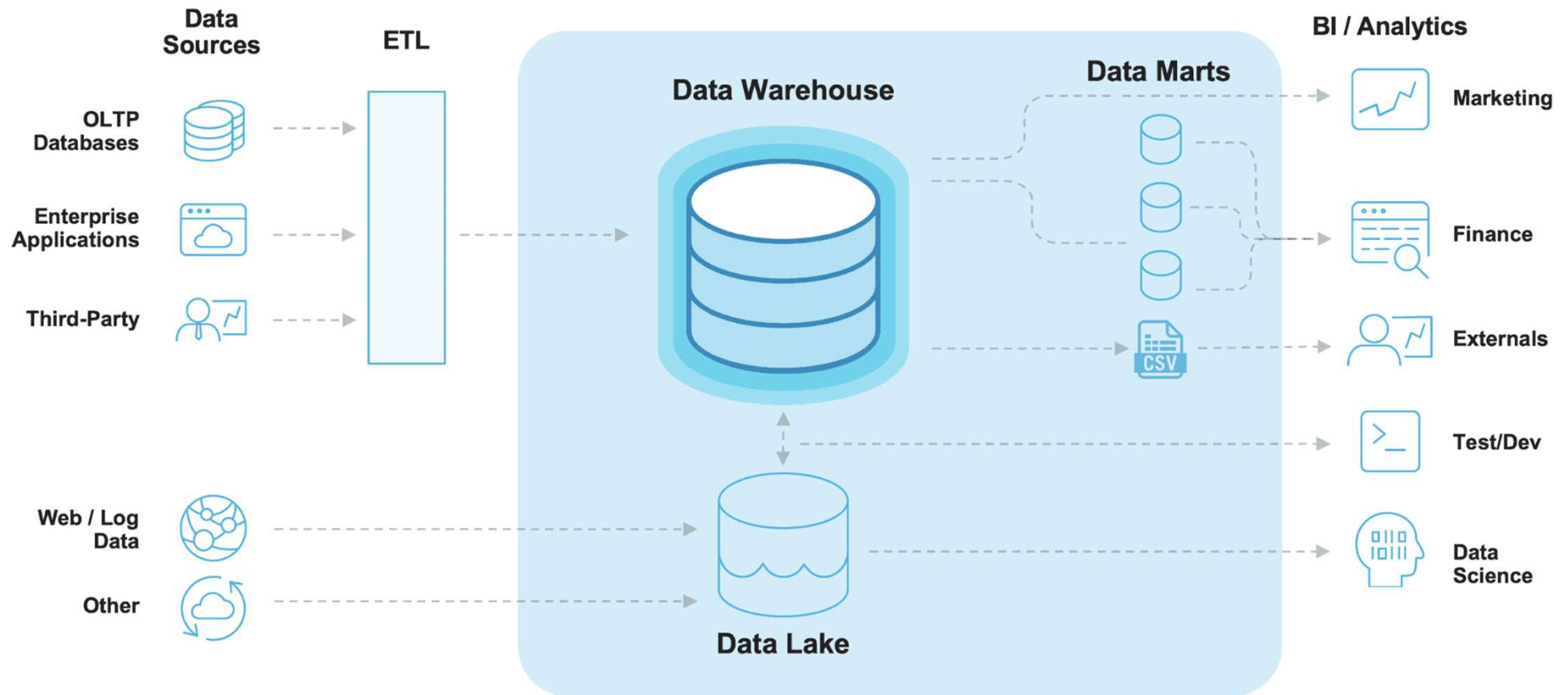
Data Warehouse vs. Data Lake (Cont.)

- Data Lake

- 전사의 주요한 Raw Data 뿐만 아니라 사용자 가공 데이터까지 수집하고 서비스하며, 모든 구성원들이 Self-Service를 통해 손쉽게 데이터를 활용할 수 있도록 서비스하는 플랫폼

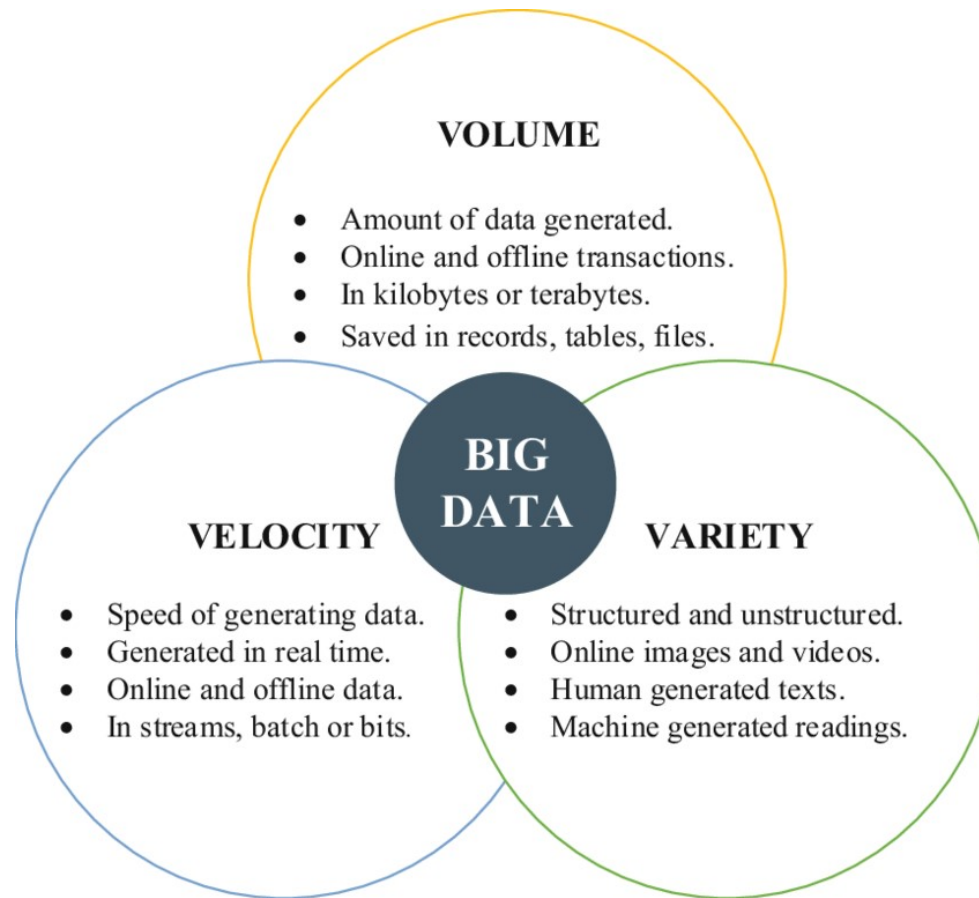


Data Warehouse vs. Data Lake (Cont.)

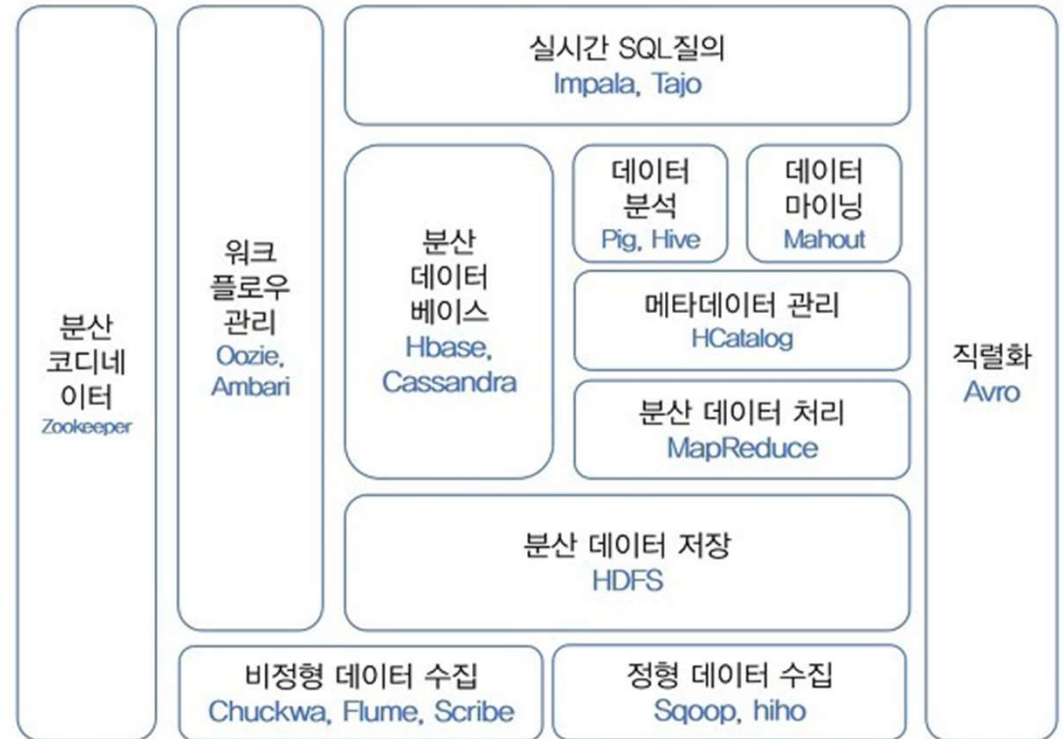
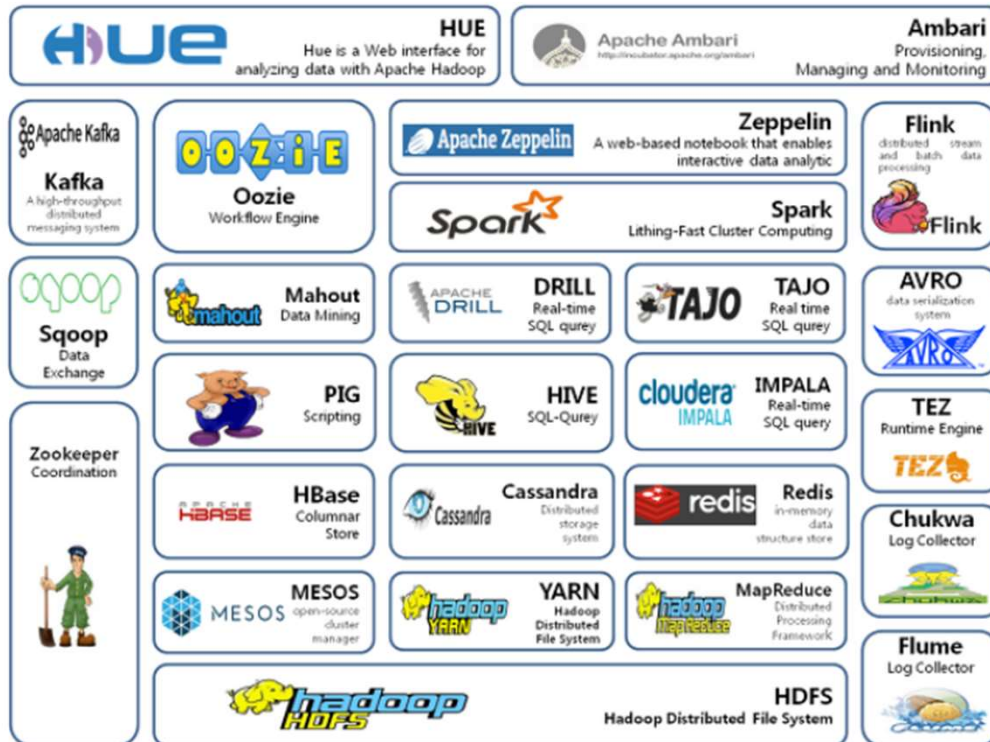


Big Data

- Refer to <https://wikidocs.net/book/2203>



Hadoop Ecosystem



Data Architect (DA)

- 책임과 의무

- 데이터 관리 시스템 및 Data warehousing Solution 설계 및 구현
- 데이터 품질과 정확성 보장
- 타 팀 구성원들과 협력하여 데이터 관련 요구 사항을 파악 및 적용
- 데이터 거버넌스 정책을 수립하고 이를 유지 및 관리
- 데이터 보안 및 개인 정보 보호 책임

- 기술 및 자격 요건

- 데이터베이스 설계 및 관리 지식
- SQL 및 NoSQL 데이터베이스 경험
- 클라우드 기반 데이터 관리 솔루션에 대한 경험
- Data warehousing, 데이터 통합 및 모델링에 대한 이해
- 데이터 보안 및 개인 정보 보호 규정에 대한 지식

- 프로그래밍 언어

- SQL, NoSQL Database, Python, Java, Scala

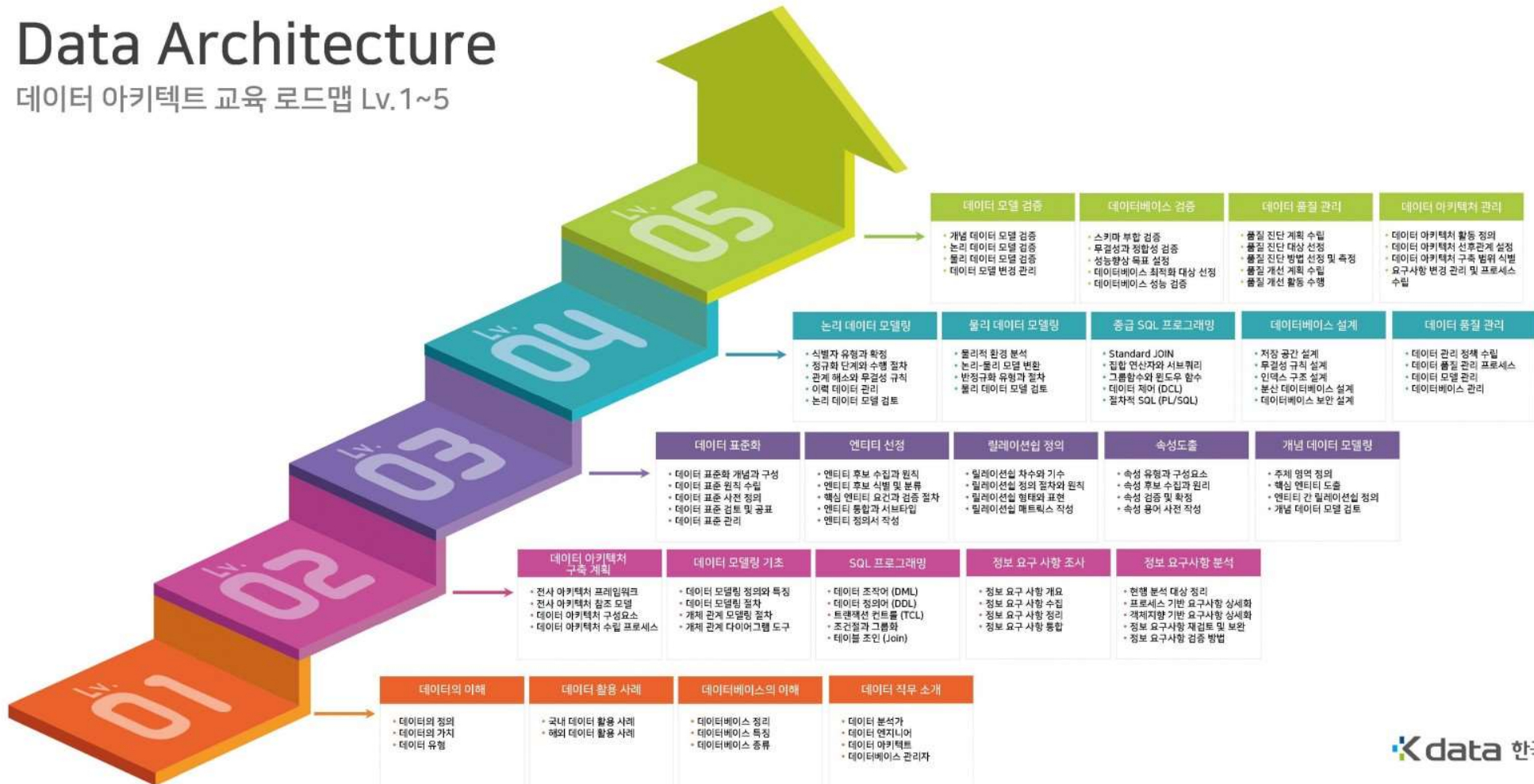
- Refer to

<https://techscene.tistory.com/entry/%EB%8D%B0%EC%9D%B4%ED%84%B0-%EC%A7%81%EA%B5%B0-%EC%95%88%EB%82%B4%EC%84%9C-DA-TA-DE-DS-ML%EC%97%94%EC%A7%80%EB%8B%88%EC%96%B4-BI-%EB%B6%84%EC%84%9D%EA%B0%80?fbclid=IwAR3LAGNjxKjVEj1mc3H9Ci3RPe7hKyKNgwhtoNqBqXEaLdOCg0x-UESKf5s>

Data Architect (DA) (Cont.)

Data Architecture

데이터 아키텍트 교육 로드맵 Lv.1~5



Technical Architect (TA)

- 책임과 의무

- 기술 비전 및 아키텍처를 개발하고 유지
- 시스템 및 소프트웨어 솔루션을 설계 및 구현
- 시스템 및 소프트웨어의 확장성과 성능 책임
- 이해 관계자와 협력하여 기술적 요구사항 이해 및 대응
- 시스템 보안 및 개인 정보 보호 책임

- 기술 및 자격 요건

- 시스템 및 소프트웨어 아키텍처 및 설계 지식
- 여러 프로그래밍 언어에 능숙
- 클라우드 시스템 및 소프트웨어에 대한 경험
- 보안 및 개인 정보 규정에 대한 지식

- 프로그래밍 언어

- Java, Python, C++, C#, Ruby

Data Engineer (DE)

- 책임과 의무

- 데이터 파이프라인 및 데이터 처리 시스템을 설계, 구현 및 운영
- 데이터 품질과 정확성 확보
- 데이터 분석가, Data Scientist 등과 협업하여 데이터 정제 및 전달
- 데이터 거버넌스 정책을 개발 및 유지관리

- 기술 및 자격 요건

- 데이터 관리 및 처리 시스템에 대한 지식
- SQL 및 데이터베이스에 대한 풍부한 경험
- 클라우드 기반 데이터 관리 솔루션 경험
- Data warehousing, 데이터 통합 및 모델링에 대한 이해

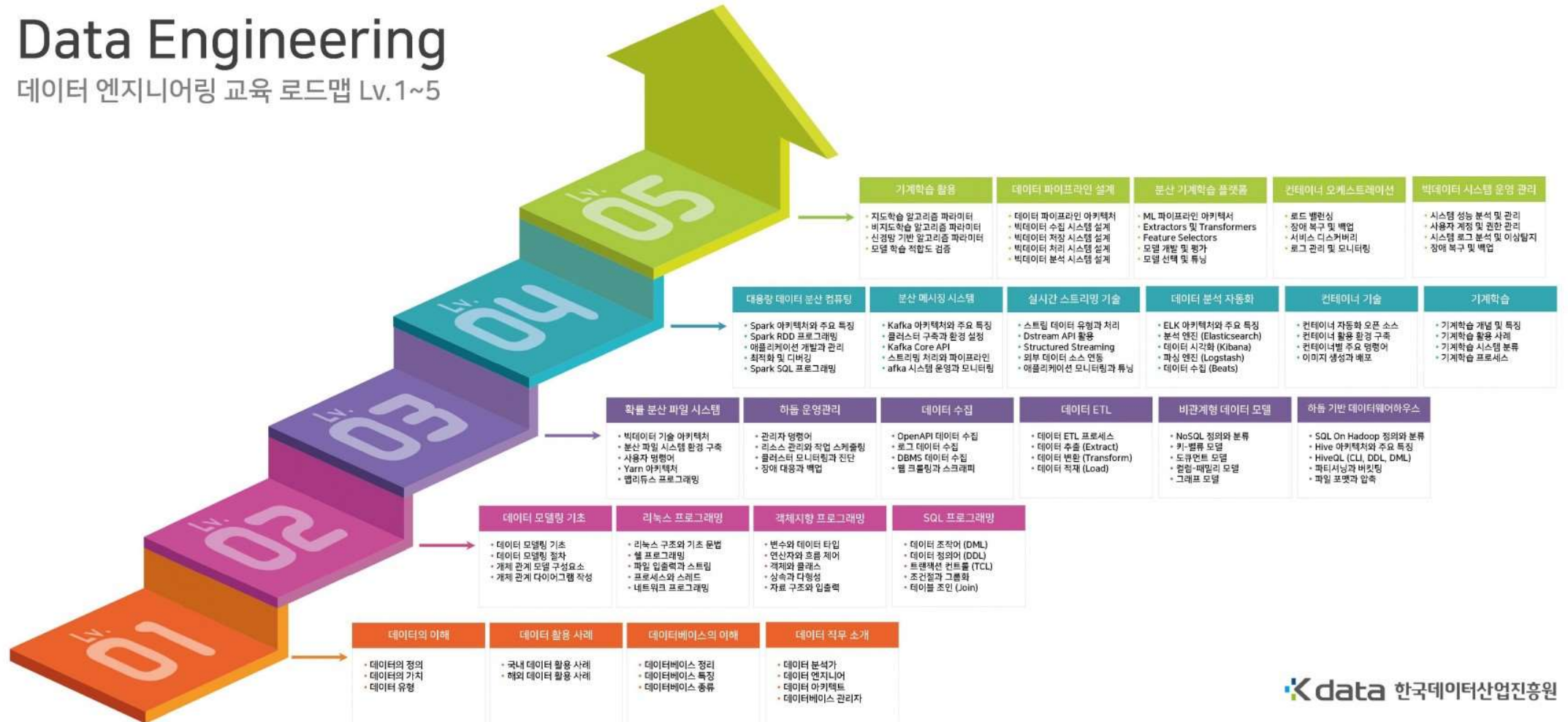
- 프로그래밍 언어

- SQL, NoSQL Database, Python, Java, Scala

Data Engineer (DE) (Cont.)

Data Engineering

데이터 엔지니어링 교육 로드맵 Lv.1~5



Data Scientist (DS)

- 책임과 의무

- 대량의 데이터를 수집, 분석 및 해석
- 통계 기술과 머신러닝 모델을 사용하여 예측하고 추세를 분석
- 데이터 시각화 도구 및 기술을 사용하여 이해 관계자에게 결과 전달
- 개발한 머신러닝 혹은 인공지능 모델의 보안 책임
- ML 엔지니어와 다르게 비즈니스 도메인에 대한 이해가 더 필요하고, 주로 Top Management 와 소통하며 비즈니스 임팩트를 만들어야 함.

- 기술 및 자격 요건

- 데이터 분석 및 시각화 기술에 대한 지식
- 통계 기술 및 머신러닝 알고리즘과 모델링에 능숙
- SQL 및 데이터 시각화 도구(예: Tableau, Power BI)에 능숙
- 클라우드 기반 데이터 분석 솔루션에 대한 경험
- Data warehousing, 데이터 통합 및 데이터 모델링에 대한 높은 이해도

- 프로그래밍 언어

- Python, R, Java, Scala

Machine Learning Engineer (ML Engineer)

- 책임과 의무
 - 기계 학습 모델 및 알고리즘을 설계 및 구현
 - 모델의 정확성과 성능 보장
 - Front-end/Back-end의 담당자와 협업하여 머신러닝 관련 요구사항을 소화
 - 머신러닝 인프라를 개발하고 유지
 - 머신러닝 모델의 보안 책임
- 기술 및 자격 요건
 - 기계 학습 알고리즘 및 모델에 대한 풍부한 지식
 - Python 및 기계 학습 라이브러리(예:TensorFlow, PyTorch)에 능숙
 - 클라우드 기반 머신러닝 솔루션 경험
 - Data warehousing, 데이터 통합 및 모델링에 대한 높은 이해
- 프로그래밍 언어
 - Python, R, Java, Scala

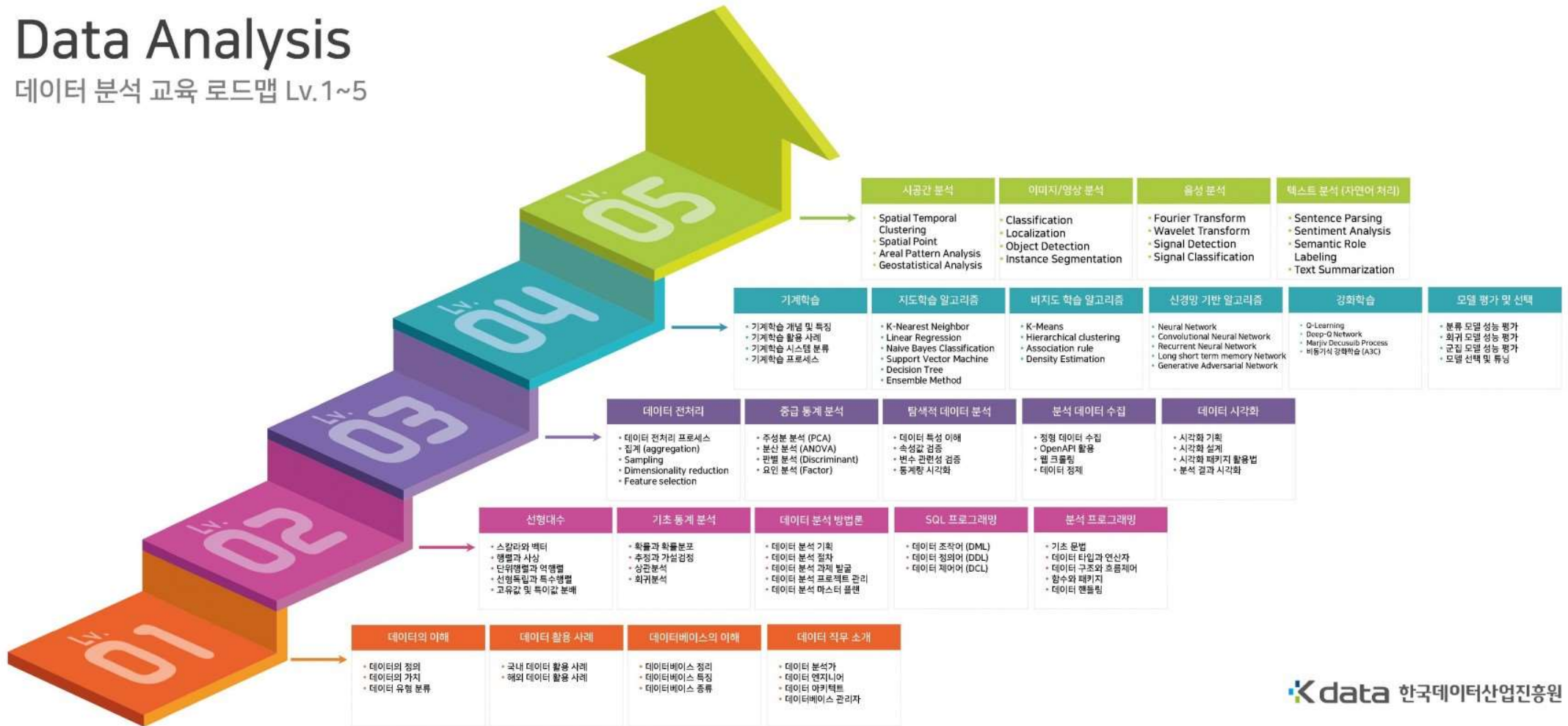
Business Intelligence Analyst (BI Analyst)

- 책임과 의무
 - 심플한 데이터 분석을 빠르게 수행하여 Insight 전달
 - 데이터 시각화 도구 적극 활용
 - 분석 결과를 비즈니스에 어떻게 활용할 것인가에 대한 고민을 통해 Impact 창출
- 기술 및 자격 요건
 - BI 분석 툴 및 시각화에 대한 지식
 - SQL에 기반한 쿼리에 대한 풍부한 지식
 - 클라우드 기반 데이터 관리 솔루션에 대한 경험
- 프로그래밍 언어
 - SQL(그외 추가적으로 Tableau, MSTR, Power BI ...)

Data Analystist

Data Analysis

데이터 분석 교육 로드맵 Lv.1~5



Data Lake 구축 방식 선정



Data Lake 구축 방식 선정 기준

평가 기준	낮음	높음
리소스 소요 변동 폭	On-Premise	Cloud
데이터 민감도	Public Cloud	On-Premise or Private Cloud
실시간 처리 필요성	Cloud	On-Premise
Biz/업무의 특수성	Public Cloud	On-Premise or Private Cloud
보유 리소스 (인력/비용) 수준	Public Cloud	On-Premise or Private Cloud



Data Lake Components



Data Lake Platform 기능 구성도

