

과정 소개

The background features a large, abstract, wavy shape in shades of green and white, resembling a stylized wave or a flowing ribbon. The shape is composed of several overlapping, curved segments that create a sense of movement and depth. The colors transition from a bright, vibrant green to a soft, pale green, and finally to a clean white. The overall effect is modern and dynamic.

목차

1. 데이터 분석 관련 직무
2. 채용 관련 기술 요구사항 통계
3. 주요 업무 및 기술
4. 교육 내용 구성
5. 기타

빅데이터 (분석) 관련 직무

빅데이터, 인공지능 관련 직무

직무	설명
데이터 엔지니어	<ul style="list-style-type: none">▪ 데이터 수집, 저장, 처리, 관리를 위한 데이터 아키텍처 구축 및 유지 보수 업무 수행▪ 데이터 아키텍처 구성, 데이터 파이프라인 설계, 데이터베이스 및 데이터 웨어하우스 설계와 운영, 데이터 추출/변환/로드(ETL) 개발, 데이터 분산 시스템 및 클라우드 서비스 구축 및 유지 보수 등
데이터 애널리스트 (데이터 분석가)	<ul style="list-style-type: none">▪ 데이터를 활용해서 필요한 분석 수행.▪ 데이터베이스 필요한 데이터를 추출하고 추출된 데이터를 다양한 측면에서 분석 및 시각화▪ 분석 결과를 기반으로 인사이트를 도출하여 비즈니스 의사 결정 지원.
데이터 사이언티스트	<ul style="list-style-type: none">▪ 데이터 분석, 통계, 기계 학습, 예측 모델링, 데이터 시각화 등 다양한 분야의 기술을 활용하여 데이터 기반의 문제를 해결하는 전문가.▪ 머신 러닝을 이용해서 미래를 예측 하거나 주어진 데이터에 대한 미지의 세계를 탐험하는 역할 수행.▪ 분석 이외에도 데이터 간 관계, 패턴, 규칙 등을 찾아내 모형화하고 머신 러닝 기법들을 서비스에 도입시킬 방법에 대한 연구 수행

빅데이터, 인공지능 관련 직무 필요 역량

직무	필요 역량
데이터 엔지니어	<ul style="list-style-type: none">▪ 데이터 아키텍처 및 데이터 모델링의 전문지식▪ 데이터베이스 관리 및 성능 튜닝에 대한 전문 지식▪ ETL 도구 및 프로세스 개발 경험▪ 대규모 분산 시스템의 설계, 구축 및 운영 경험▪ 클라우드 플랫폼 (AWS, Azure, Google Cloud 등)의 이해 및 경험▪ 프로그래밍 언어 (Python, Java, Scala 등)에 대한 지식과 경험
데이터 애널리스트 (데이터 분석가)	<ul style="list-style-type: none">▪ 데이터 분석 및 통계 지식▪ 데이터 시각화 및 리포팅 도구 (Tableau, Power BI 등) 사용 능력▪ 데이터베이스 및 SQL 쿼리 작성 능력▪ 프로그래밍 언어 (Python, R 등)에 대한 이해 및 사용 능력▪ 비즈니스 이해력과 문제 해결 능력
데이터 사이언티스트	<ul style="list-style-type: none">▪ 수학 및 통계학 지식▪ 프로그래밍 언어 (Python, R 등)에 대한 능숙한 이해와 활용 능력▪ 머신러닝 및 딥러닝 알고리즘에 대한 지식▪ 빅데이터 처리 기술 (Hadoop, Spark 등)에 대한 이해와 사용 능력▪ 데이터 시각화 및 리포팅 도구 사용 능력▪ 비즈니스 이해력과 문제 해결 능력

빅데이터, 인공지능 관련 직무 필요 역량

	데이터 엔지니어	데이터 분석가	데이터 사이언티스트
주요 업무	데이터 수집, 저장, 처리 관리 시스템 구축 및 유지보수	데이터 분석 및 시각화 비즈니스 인사이트 도출	데이터 모델링 머신러닝 알고리즘 개발 예측 분석
필수 기술 (도구)	SQL, NoSQL, 데이터웨어하우징, 빅데이터기술(Spark, Hadoop), 클라우드 서비스	SQL, Excel, 데이터 시각화 도구 (Tableau, Power BI 등), 통계 분석 도구(R, SAS 등)	Python, R, 머신러닝 라이브러리 (Tensorflow, Scikit-learn, Pytorch 등), 통계, 데이터 모델링
주요 역할	데이터 인프라 설계 및 구현 데이터 파이프라인 개발	데이터 분석 리포트 작성 비즈니스 의사 결정 지원	머신러닝 모델 개발 및 배포 데이터 기반 제품 및 서비스 개발
취업 분야	대기업 IT 부서, 데이터 플랫폼 회사, 클라우드 서비스 제공 회사	금융, 마케팅, 리테일 등 다양한 산업	IT 기업, 연구 기관, 금융 등 다양한 산업

기술 요구사항 통계

구인 통계

617(115)

데이터 엔지니어

547(122)

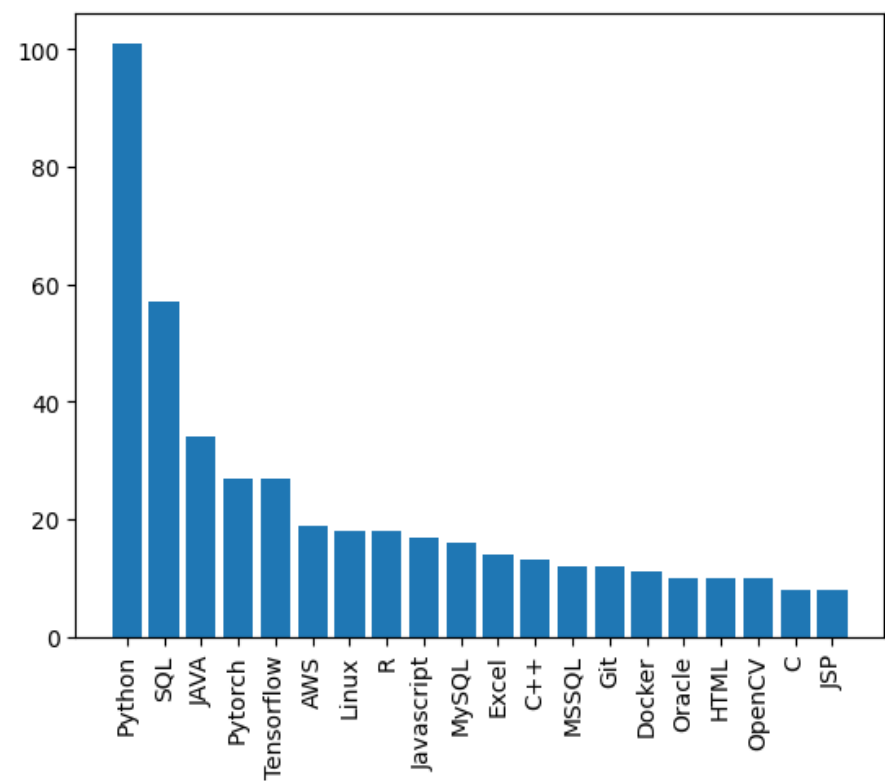
데이터 사이언티스트

703(127)

머신러닝엔지니어

빅데이터, 인공지능 관련 기술 요구 사항

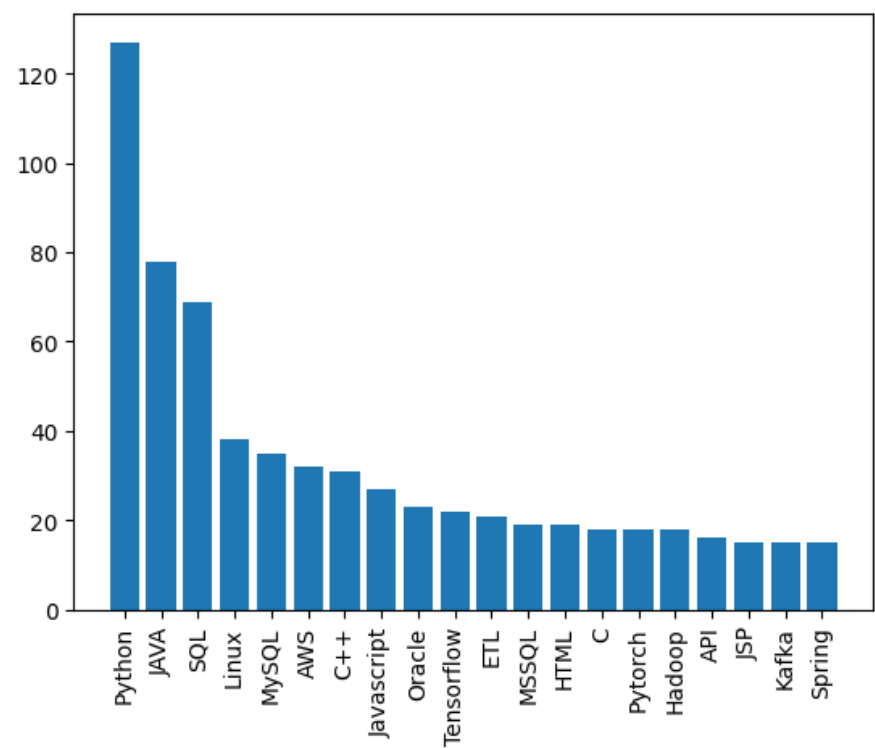
데이터 분석가, 데이터 사이언티스트



	skill	count_required
0	Python	101
1	SQL	57
2	JAVA	34
3	Pytorch	27
4	Tensorflow	27
5	AWS	19
6	Linux	18
7	R	18
8	Javascript	17
9	MySQL	16
10	Excel	14
11	C++	13
12	MSSQL	12
13	Git	12
14	Docker	11
15	Oracle	10
16	HTML	10
17	OpenCV	10
18	C	8
19	JSP	8

빅데이터, 인공지능 관련 기술 요구 사항 통계

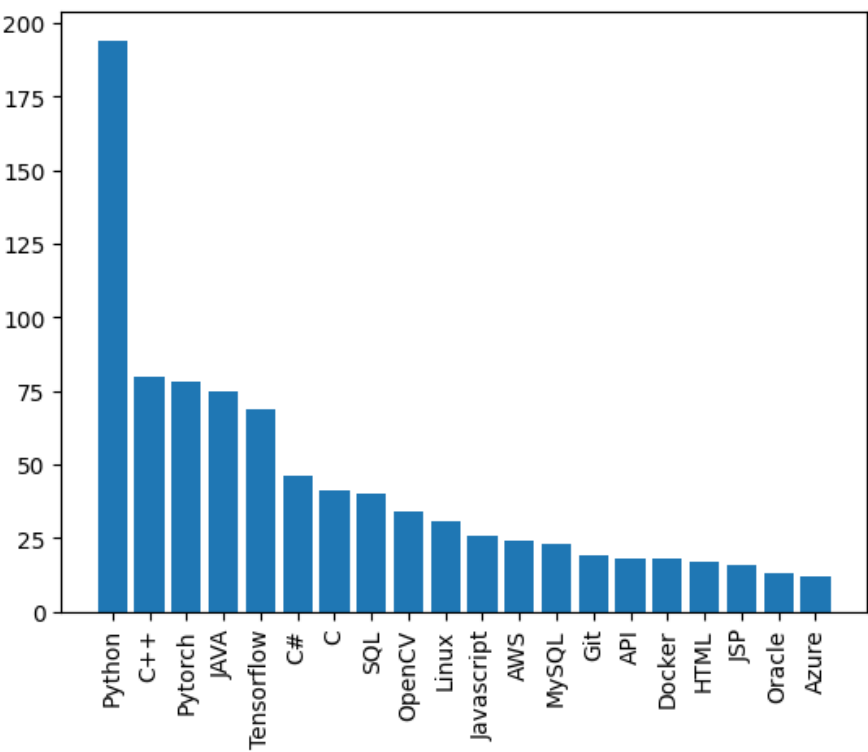
데이터 엔지니어



	skill	count_required
0	Python	127
1	JAVA	78
2	SQL	69
3	Linux	38
4	MySQL	35
5	AWS	32
6	C++	31
7	Javascript	27
8	Oracle	23
9	Tensorflow	22
10	ETL	21
11	MSSQL	19
12	HTML	19
13	C	18
14	Pytorch	18
15	Hadoop	18
16	API	16
17	JSP	15
18	Kafka	15
19	Spring	15

빅데이터, 인공지능 관련 기술 요구 사항 통계

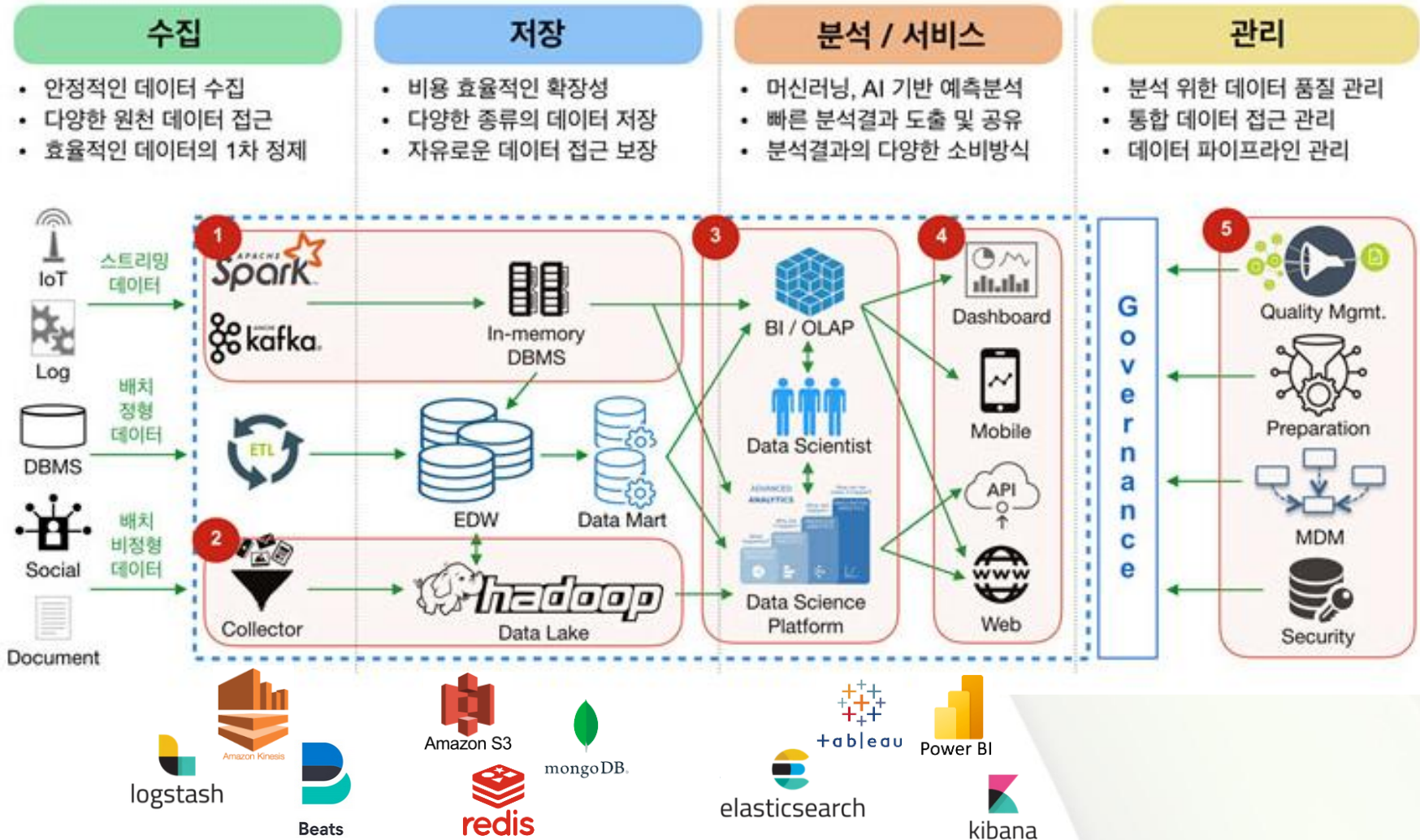
머신러닝 엔지니어



	skill	count_required
0	Python	194
1	C++	80
2	Pytorch	78
3	JAVA	75
4	Tensorflow	69
5	C#	46
6	C	41
7	SQL	40
8	OpenCV	34
9	Linux	31
10	Javascript	26
11	AWS	24
12	MySQL	23
13	Git	19
14	API	18
15	Docker	18
16	HTML	17
17	JSP	16
18	Oracle	13
19	Azure	12

주요 업무 및 기술

빅데이터 처리 시스템 구조



빅데이터 처리 시스템 구조

■ 주요 데이터 수집 기술

제품 (기술)	설명
Apache Kafka	<ul style="list-style-type: none">대용량 실시간 데이터 스트리밍 처리를 위한 분산 메시징 시스템매우 높은 처리량과 확장성을 제공
Apache Spark	<ul style="list-style-type: none">대용량 데이터 처리에 특화된 빠른 클러스터 컴퓨팅 기술메모리 내 데이터 처리를 지원하여 Hadoop MapReduce보다 처리 속도가 훨씬 빠르며, 머신 러닝, 스트리밍 처리, SQL 쿼리 등 다양한 데이터 처리 기능 제공
Amazon Kinesis	<ul style="list-style-type: none">AWS 클라우드 기반 서비스로, 실시간 데이터 스트리밍 처리 지원.대용량 스트림 데이터 수집 및 분석
Beats	<ul style="list-style-type: none">가볍고 단일 목적의 데이터 수집기Filebeat, Metricbeat 등 여러 버전의 Beats 제공특정 유형의 데이터를 수집하고 Elasticsearch 또는 Logstash로 전송
Logstash	<ul style="list-style-type: none">강력한 데이터 처리 파이프라인 제공다양한 입력, 필터, 출력 플러그인 제공데이터 변환과 조작을 위한 다양한 필터를 제공 → 복잡한 데이터 파이프라인 구성 가능

빅데이터 처리 시스템 구조

■ 주요 데이터 저장 기술

제품 (기술)	설명
HDFS	<ul style="list-style-type: none">▪ Apache Hadoop의 핵심 컴포넌트로 대규모 데이터 세트를 저장하기 위한 분산 파일 시스템▪ 데이터를 여러 노드에 분산 저장함으로써 높은 처리량 제공 및 노드 장애에 대비한 고가용성 지원
Amazon S3	<ul style="list-style-type: none">▪ AWS에서 제공하는 객체 저장 서비스로 인터넷을 통해 언제 어디서나 데이터 저장 및 검색 지원▪ 뛰어난 확장성과 데이터 내구성 및 보안 기능을 제공▪ 빅데이터 애플리케이션 및 분석 도구와 통합 용이
MongoDB	<ul style="list-style-type: none">▪ 유연한 스키마를 지원하는 NoSQL 문서 데이터베이스▪ 대용량 데이터를 처리하고 저장하는 데 효율적▪ JSON 형식의 문서를 사용해 데이터 구조의 변화가 자주 일어나는 애플리케이션에 적합.
Redis	<ul style="list-style-type: none">▪ 인메모리 데이터 구조 저장소로 키-값 형식의 데이터베이스▪ 매우 빠른 성능 제공▪ 캐싱, 메시지 브로커, 큐 등 다양한 용도로 활용▪ 빅데이터 애플리케이션에서 실시간 처리를 필요로 할 때 종종 사용

빅데이터 처리 시스템 구조

- 주요 데이터 분석 및 시각화 도구

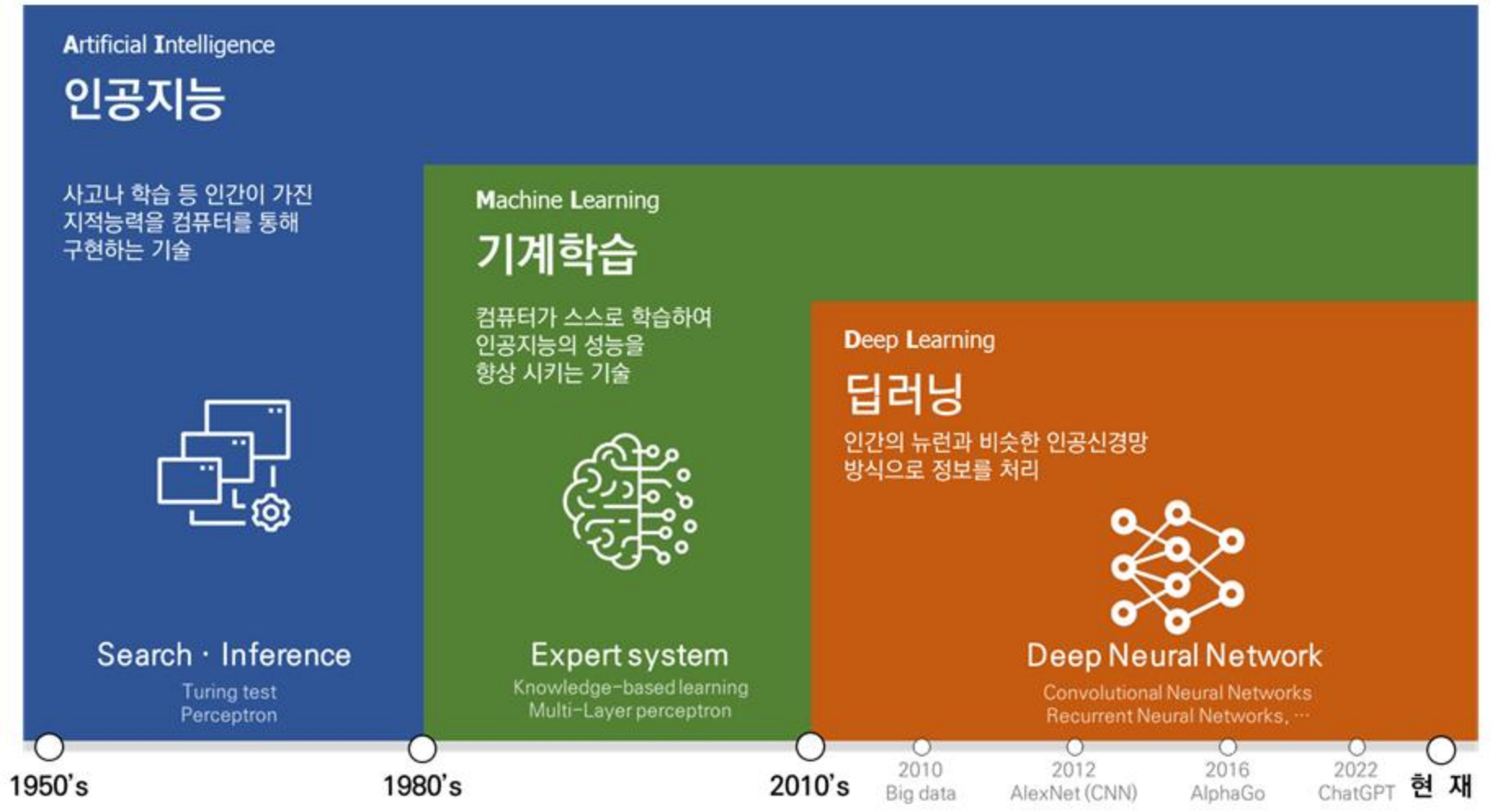
제품 (기술)	설명
Tableau	<ul style="list-style-type: none">데이터 시각화 도구쉬운 데이터 분석 및 인터랙티브한 대시보드 생성 지원복잡한 데이터 세트에 대한 시각적 탐색 → 다양한 관점에서 인사이트 도출
Microsoft Power BI	<ul style="list-style-type: none">데이터 분석 및 비즈니스 인텔리전스를 위한 도구다양한 소스의 데이터에 대한 통합, 분석 및 시각화 지원클라우드 기반 서비스로 제공되어 접근성과 편의성이 좋음

빅데이터 처리 시스템 구조

■ 데이터 사이언스 플랫폼

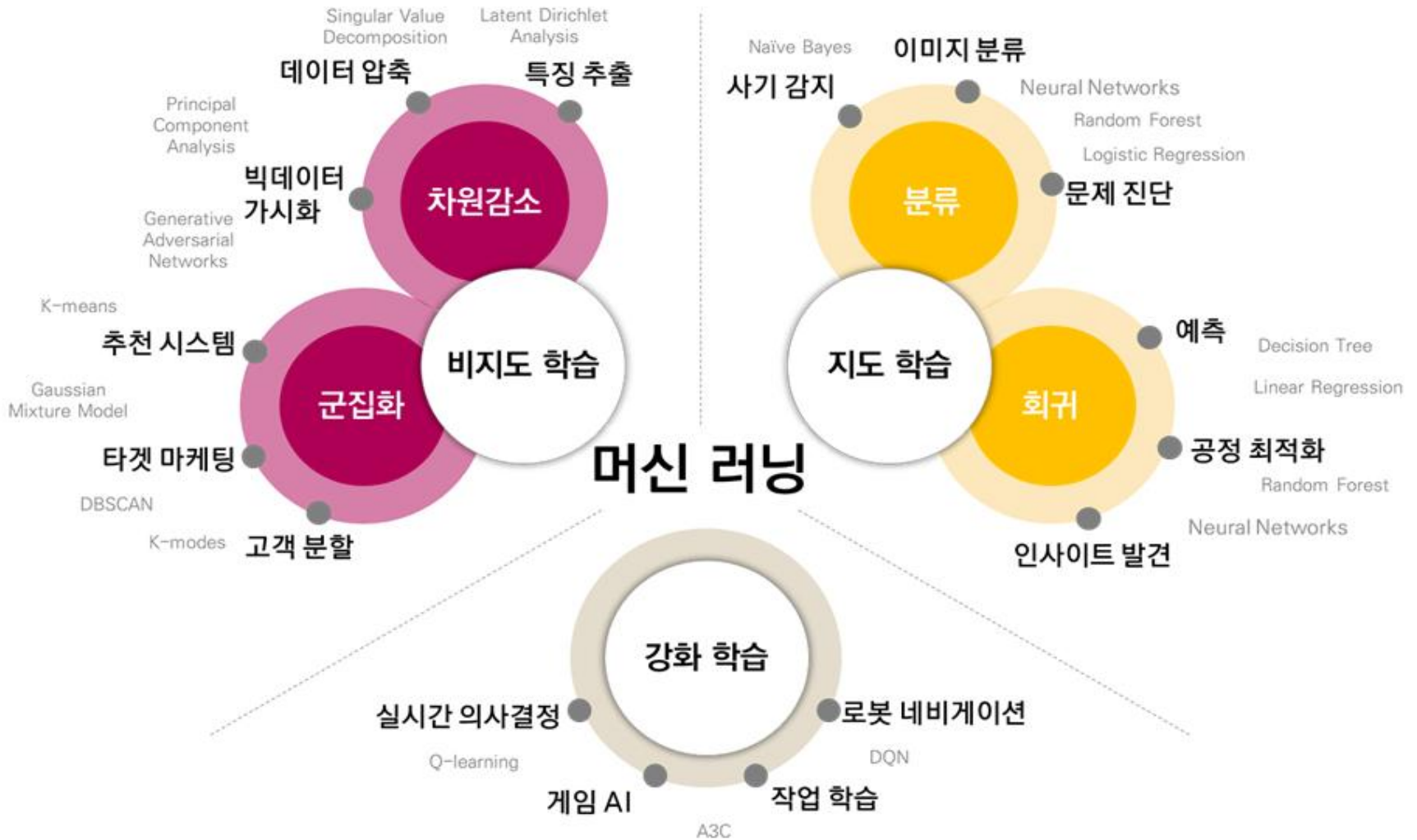
제품 (기술)	설명
Google Cloud Platform (GCP)	<ul style="list-style-type: none">▪ BigQuery, Google AI Platform, Dataflow 등 다양한 데이터 처리와 머신러닝 서비스 제공▪ 강력한 인프라와 함께 뛰어난 스케일링 및 통합 서비스 제공
Microsoft Azure Machine Learning	<ul style="list-style-type: none">▪ 데이터 분석 및 비즈니스 인텔리전스를 위한 도구▪ 클라우드 기반의 ML 서비스로 모델 구축, 훈련, 배포를 위한 도구 제공▪ 다양한 데이터 소스와의 통합과 자동화된 ML 기능 및 시각화 도구 지원
AWS SageMaker	<ul style="list-style-type: none">▪ 데이터 사이언티스트와 개발자가 머신러닝 모델을 쉽게 생성, 훈련 및 배포할 수 있는 완전 관리형 서비스 제공▪ 주피터 노트북 인스턴스를 포함하여 다양한 ML 도구 지원
Databricks	<ul style="list-style-type: none">▪ Apache Spark 기반의 클라우드 서비스▪ 빅데이터 분석, 머신러닝, 스트리밍 및 그래프 처리 지원▪ 대규모 데이터셋의 처리와 협업에 강점
IBM Watson Studio	<ul style="list-style-type: none">▪ 데이터 사이언스 및 머신러닝 프로젝트를 위한 강력한 플랫폼▪ 데이터 분석, 시각화, 모델링을 위한 툴 제공▪ Watson AI의 기능을 통해 향상된 분석 기능 제공

인공지능 개념 체계



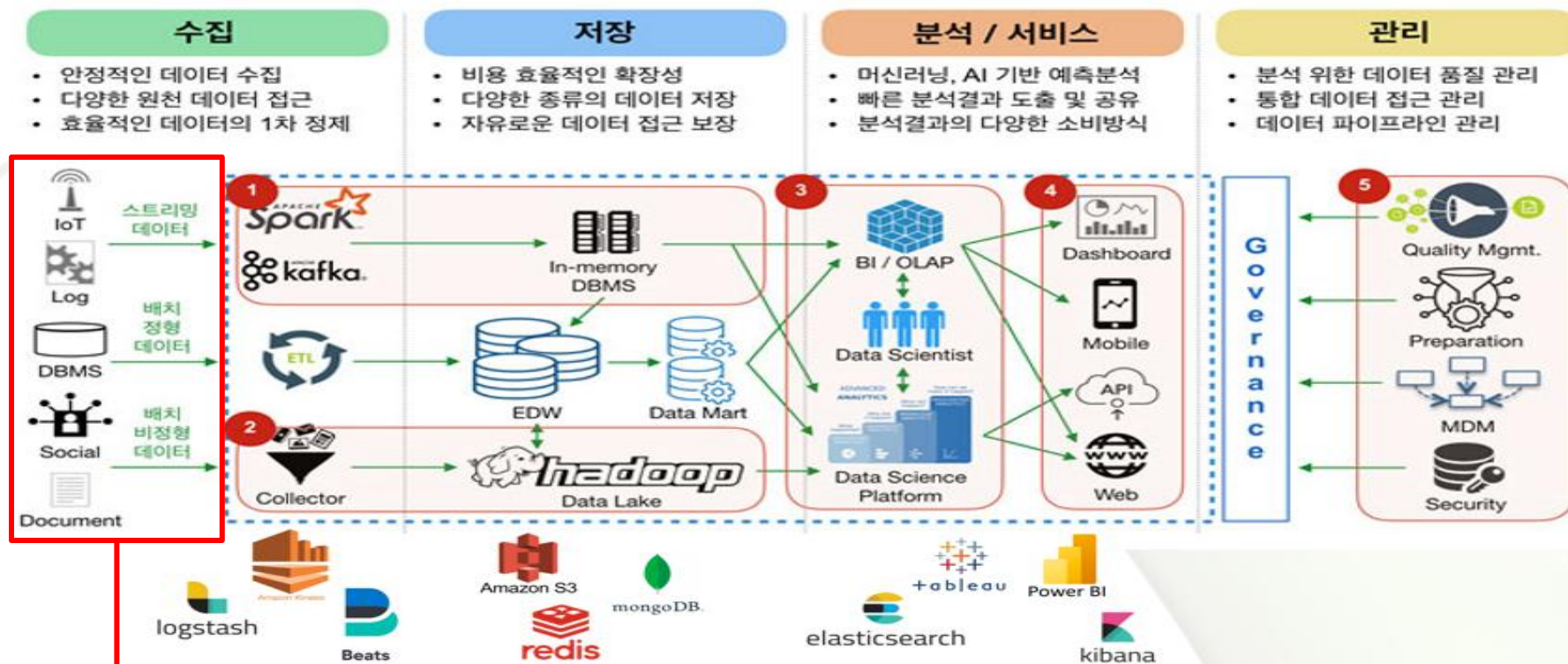
출처) CAD와 AI의 현재, 그리고 미래, 한국 CDE 학회 웹진, https://webzine.cde.or.kr/webzine/sub.html?md_no=41

머신러닝, 딥러닝



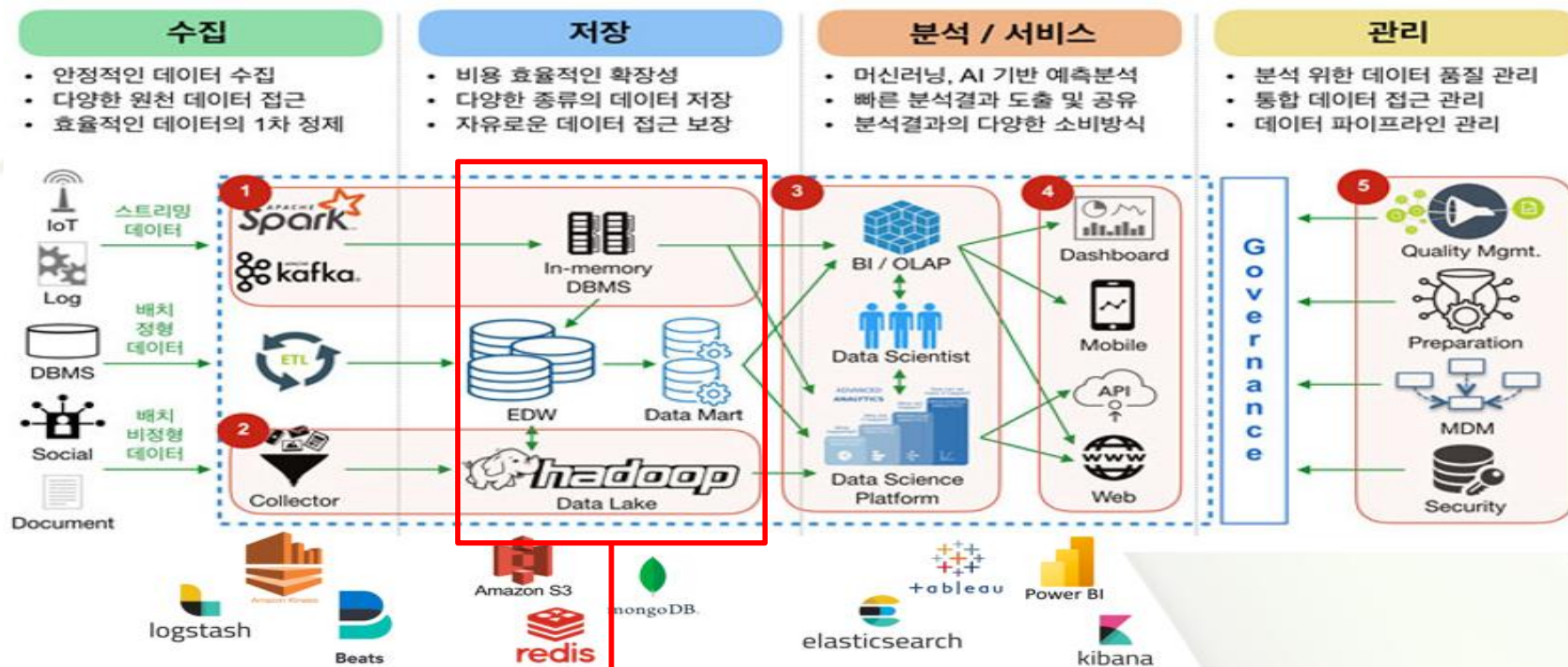
교육 내용 구성

데이터 수집 교육 내용



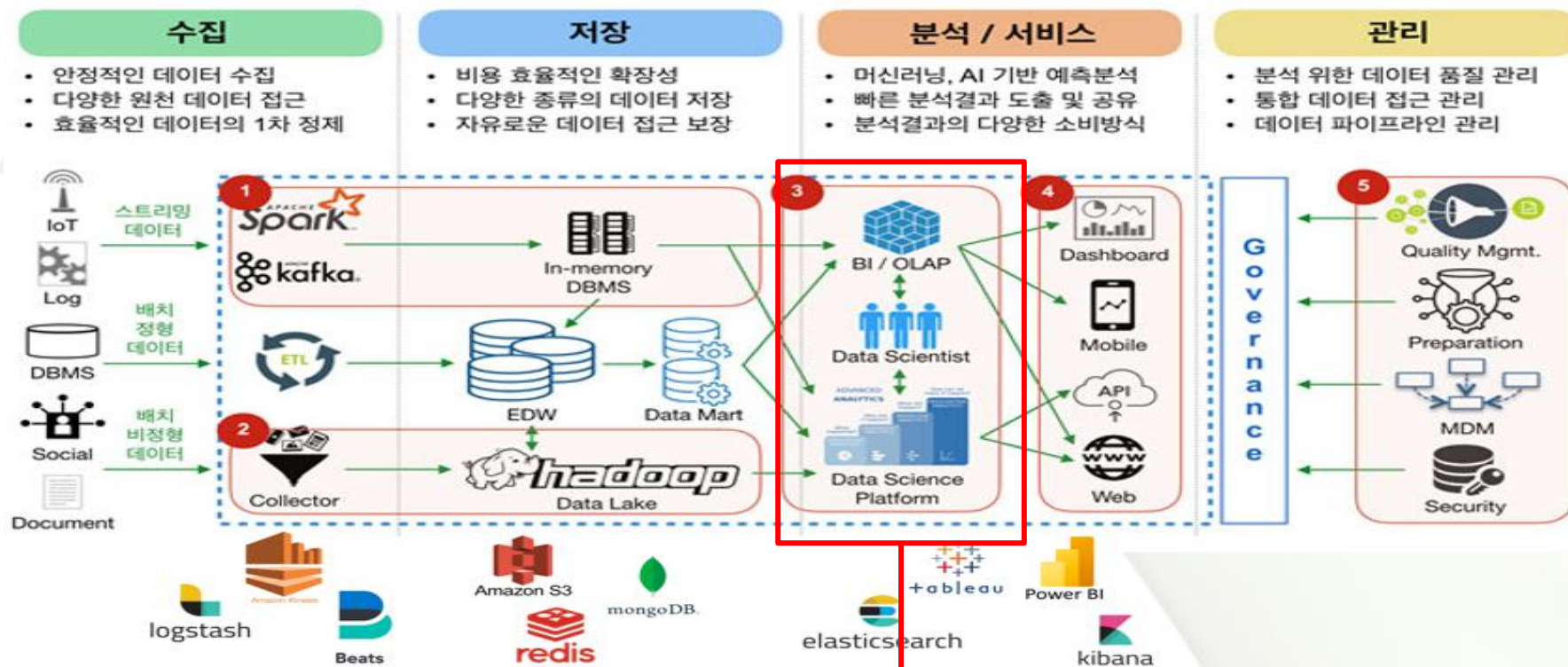
- 데이터베이스 → MySQL (MariaDB) 등의 데이터베이스 데이터 활용
- 파일 → xml, csv, json 등의 파일의 데이터 활용
- 웹크롤링 → Open API 또는 웹페이지에서 데이터 추출

데이터 저장 교육 내용



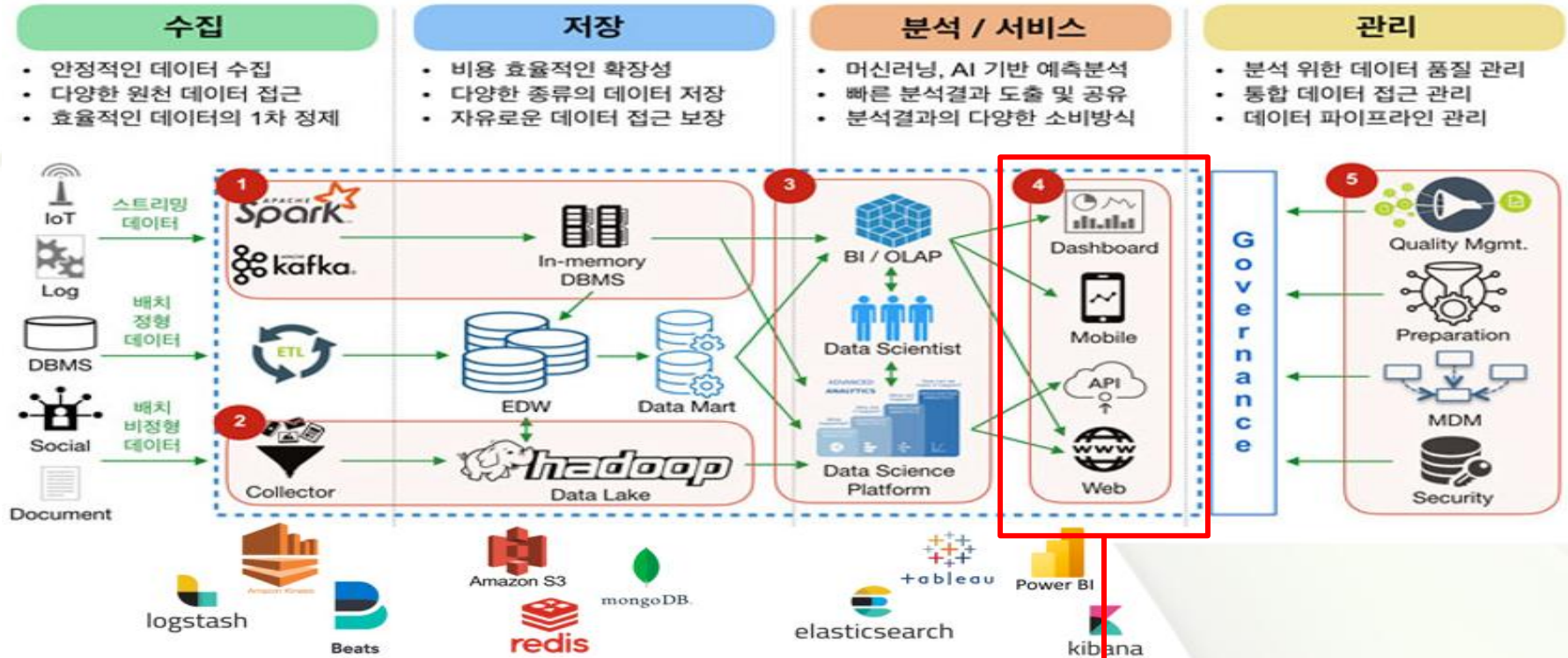
- 데이터베이스 → MySQL (MariaDB) 등의 데이터베이스 데이터 활용
- 파일 → xml, csv, json 등의 파일의 데이터 활용
- NoSQL 데이터베이스 → MongoDB 등의 NoSQL 데이터베이스 데이터 활용

데이터 분석 / 서비스 교육 내용



- 데이터 분석 → 통계 기반 데이터 분석, 탐색적 데이터 분석
- 머신러닝 → 다양한 머신러닝 알고리즘 활용 예측 모델 개발
- 딥러닝 → ANN, CNN, RNN, Transformers 등 알고리즘 활용 모델 개발
- 생성형 AI → LLM 기반 생성형 AI 서비스 활용

데이터 분석 / 서비스 교육 내용



- 모델 서빙 → Flask 기반 웹서비스 또는 API 서비스 개발



기타

참고

- Google i/o developer keynote recap
 - https://www.youtube.com/watch?v=xKmEOXZsU_0
- Microsoft build keynote recap
 - <https://www.youtube.com/watch?v=vBCrJaLseQc&list=PLFPUGjQjckXE-DsPj61-5kKIhJ3194fON&index=1>
- WWDC keynote recap
 - <https://www.youtube.com/watch?v=LtIWE4CCX5M>
 - <https://www.youtube.com/watch?v=sBXdyUA6A88> (13분 50초 ~)
- 기타
 - <https://www.youtube.com/watch?v=Vq2rgjejxHU> (4분 42초 ~ 6분 14초)