

1. 데이터 문해력이란?

데이터 팀의 역할과 데이터 조직 구성원
데이터 문해력의 정의와 중요성

Contents

1. 데이터란?
2. 데이터 팀의 미션과 발전 단계
3. 클라우드란?
4. 데이터 조직 구성원
5. 데이터 문해력의 정의와 중요성
6. 데이터 일을 할 때 기억할 점



데이터란?

도대체 데이터란 무엇인가?

◆ 데이터는 우리 생활 모든 곳에 존재

❖ 데이터는 우리가 일상생활에서 관찰할 수 있는 모든 것

- 온도, 풍향, 소리, 움직임, ...
- 데이터를 바탕으로 의미있는 정보의 도출이 가능

❖ 시작은 이런 데이터를 기록하고 수집하는 것!

- 이를 보통 **Digitization**이라고 부름
- 데이터의 수집이 가장 쉬운 환경은 바로 온라인 환경

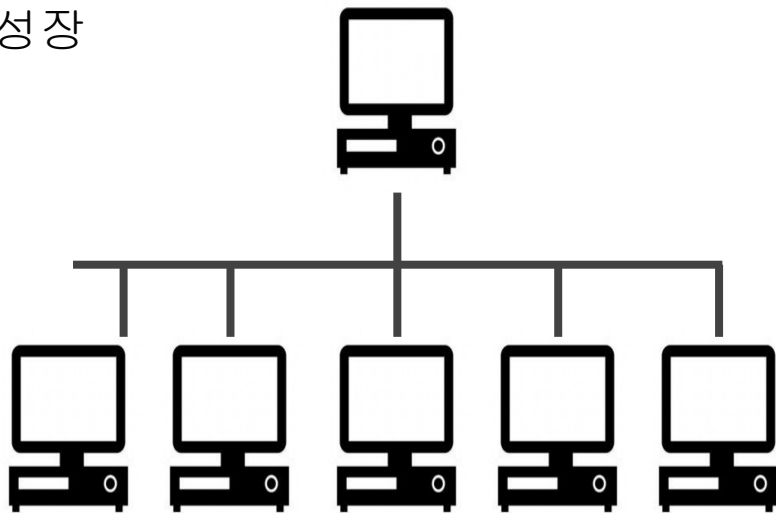
◆ 데이터의 크기 변화

❖ 웹과 모바일 폰 사용의 보편화

- 데이터 크기의 폭발적 성장

❖ 큰 데이터를 처리할 수 있는 기술적 진보

- 빅데이터 기술과 클라우드의 성장



분산처리 시스템의 등장

◆ 빅데이터 예 - 디바이스 데이터

- ❖ 모바일 디바이스
 - 위치정보
- ❖ 스마트 TV
- ❖ 각종 센서 데이터 (IoT 센서)
- ❖ 네트워킹 디바이스
- ❖ ...



◆ 빅데이터 예 - 웹

- ❖ 수십 조개 이상의 웹 페이지 존재 -> 온갖 종류의 지식의 바다
- ❖ 웹 검색엔진 개발은 진정한 대용량 데이터 처리
 - 웹 페이지를 크롤하여 중요한 페이지를 찾아내고 (페이지 랭크) 인덱싱하고 서빙
 - 구글이 빅데이터 기술의 발전에 지대한 공헌
- ❖ 사용자 검색어와 클릭 정보 자체도 대용량
 - 이를 마이닝하여 개인화 혹은 별도 서비스 개발이 가능
 - 검색어를 바탕으로한 트렌드 파악, 통계 기반 번역, ...
- ❖ 요즘은 웹 자체가 **NLP** 거대 모델 개발의 훈련 데이터로 사용되고 있음

데이터 팀의 미션과 발전 단계

데이터 팀의 일반적인 미션과 이상적인 발전 단계를 통해
데이터 팀이 어떻게 회사의 발전/성장에 도움이 되는지
살펴본다

◆ 데이터 조직의 미션은?

◆ 신뢰할 수 있는 데이터를 바탕으로 **부가 가치** 생성

- Data is the new oil
- But data can be a liability
 - 데이터의 잘못된 노출과 사용으로 인한 **위험**을 줄여야 함

◆ 데이터 조직이 하는 일 (1) - Decision Science

- ❖ 고품질 데이터 기반으로 의사 결정권자에게 입력 제공
 - 데이터를 고려한 결정(data informed decisions)을 가능하게 해줌
 - vs. 데이터 기반 결정(data driven decisions)
 - 예를 들면 데이터 기반 지표 정의, 대시보드와 리포트 생성 등을 수행



Data Literacy (데이터 문해력)

◆ 데이터 조직이 하는 일 (2) - Product Science

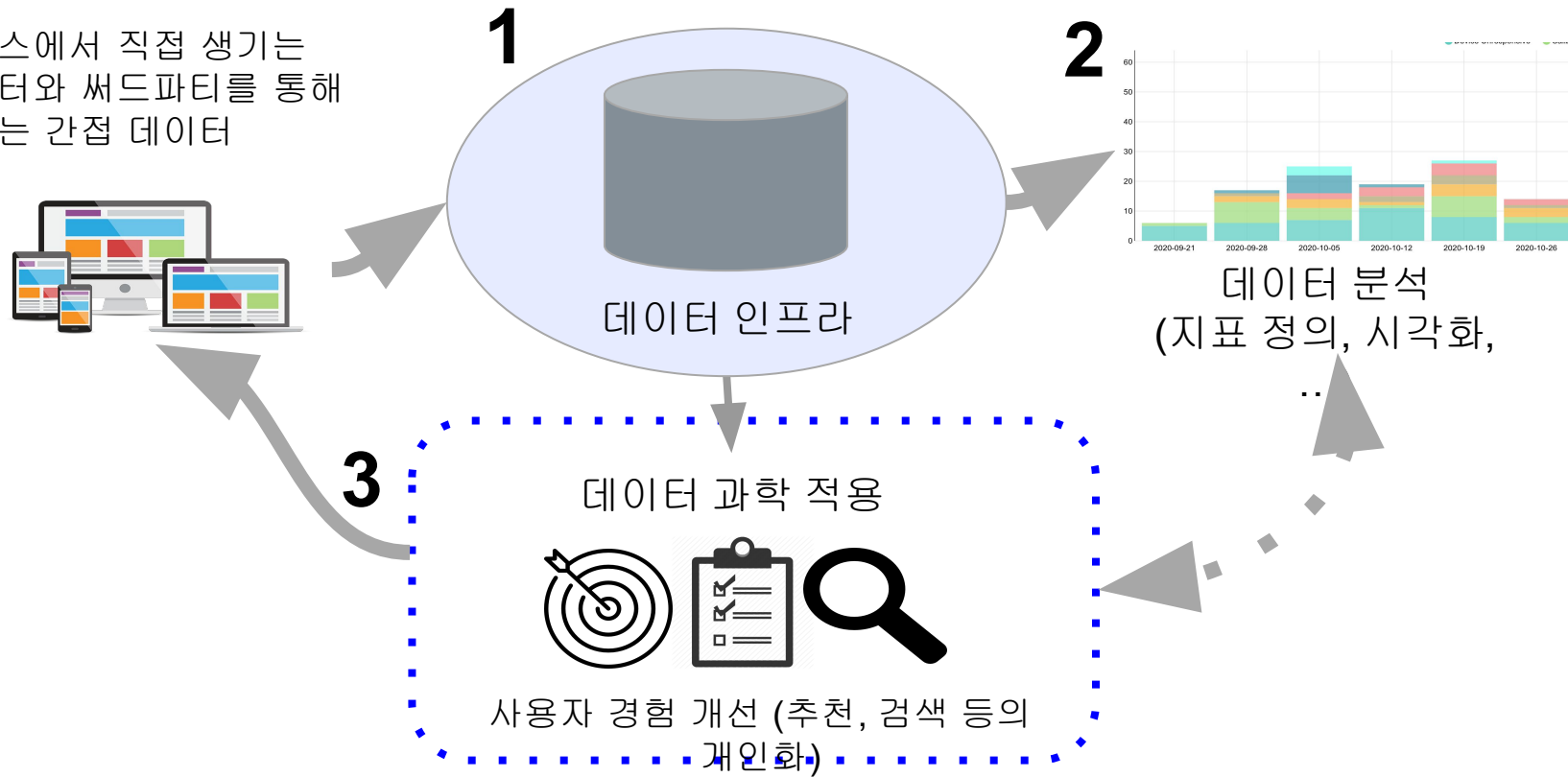
❖ 고품질 데이터를 기반으로 사용자 서비스 경험 개선 혹은 프로세스 최적화

- 머신 러닝과 같은 알고리즘을 통해 사용자의 서비스 경험을 개선
 - 예) 개인화를 바탕으로한 추천과 검색 기능 제공
- 공장이라면 공정 과정에서 오류를 최소화 혹은 기기 고장 예측등을 수행

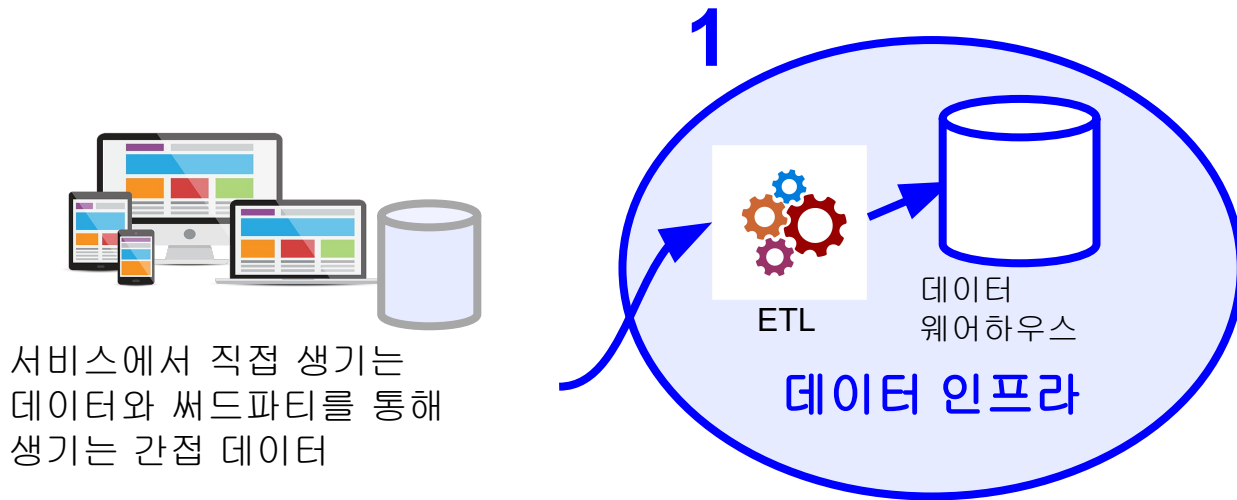


◆ 데이터의 흐름과 데이터 팀의 발전 단계

서비스에서 직접 생기는
데이터와 써드파티를 통해
생기는 간접 데이터



◆ 데이터 팀의 발전 - 1. 데이터 인프라 구축



데이터 인프라의 구축은 **데이터 엔지니어**가 수행함

ETL: Extract/Transform/Load

다루는 데이터의 크기가 커지면 빅데이터 처리 기술 습득이 필요 (Spark)

◆ 데이터 웨어하우스

- ❖ 회사에 필요한 모든 데이터를 모아놓은 중앙 데이터베이스 (SQL 데이터베이스)
 - 보통 다음 중 하나를 선택 (이 모두 SQL을 지원)
 - 클라우드 옵션: AWS Redshift, 구글 클라우드의 BigQuery, Snowflake 등등
 - 오픈소스 기반의 Hive/Presto
- ❖ 중요 포인트는 프로덕션용 데이터베이스와 별개의 데이터베이스여야 한다는 점
- ❖ 데이터 웨어하우스의 구축이 진정한 데이터 조직이 되는 첫 번째 스텝
- ❖ 클라우드를 사용하는 것이 일반적
 - 클라우드에 대해 뒤에서 별도 설명

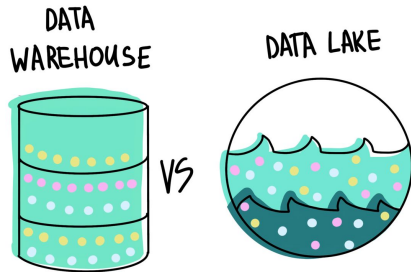
◆ 용어 설명: 데이터 레이크 vs. 데이터 웨어하우스

❖ 데이터 레이크 (Data Lake)

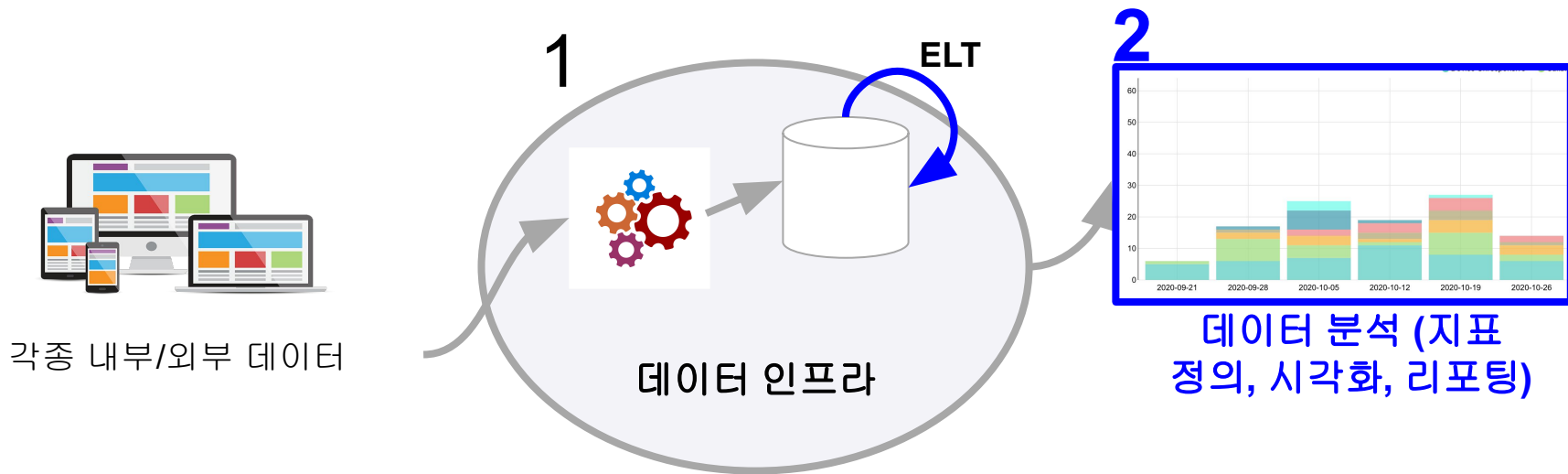
- 구조화 데이터 + 비구조화 데이터
- 보존 기한이 없는 모든 데이터를 원래 형태로 보존하는 스토리지에 가까움
- 보통은 데이터 웨어하우스보다 몇배는 더 큰 스토리지

❖ 데이터 웨어하우스 (Data Warehouse)

- 보존 기한이 있는 구조화된 데이터를 저장하고 처리하는 스토리지
- 보통 BI 툴들(룩커, 태블로, 수퍼셋, ...)은 데이터 웨어하우스를 백엔드로 사용함



◆ 데이터 팀의 발전 - 2. 데이터 분석 수행



이는 **데이터 분석가 (Data Analyst)**가 맡는 일임
ETL된 데이터를 조합하여 새로운 정보 생성 (ELT)
좋은 지표 정의, 대시보드 생성/관리, 데이터 기반 리포트 작성

◆KPI(Key Performance Indicator)란?

❖ 조직내에서 달성하고자 하는 중요한 목표

- 보통 정량적인 숫자가 선호됨
- 예를 들면 매출액 혹은 유료 회원의 수/비율 (정의가 *중요*함)

❖ KPI의 수는 적을수록 좋음

❖ 잘 정의된 KPI -> 현재 상황을 알고 더 나은 계획 가능

- 정량적이기에 시간에 따른 성과를 추적하는 것이 가능
- OKR(Objectives and Key Results)과 같은 목표 설정 프레임웍의 중요한 포인트

◆ 시각화 대시보드란?

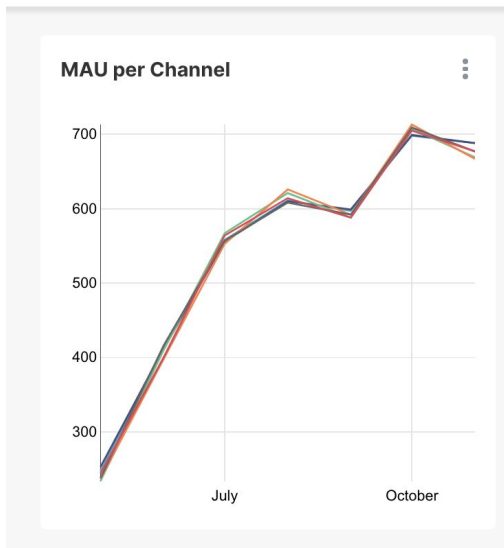
❖ 보통 중요한 지표를 시간의 흐름과 함께 보여주는 것이 일반적

- 지표의 경우 3A(Accessible, Actionable, Auditable)가 중요
- 중요 지표의 예: 매출액, 월간/주간 액티브 사용자수, ...

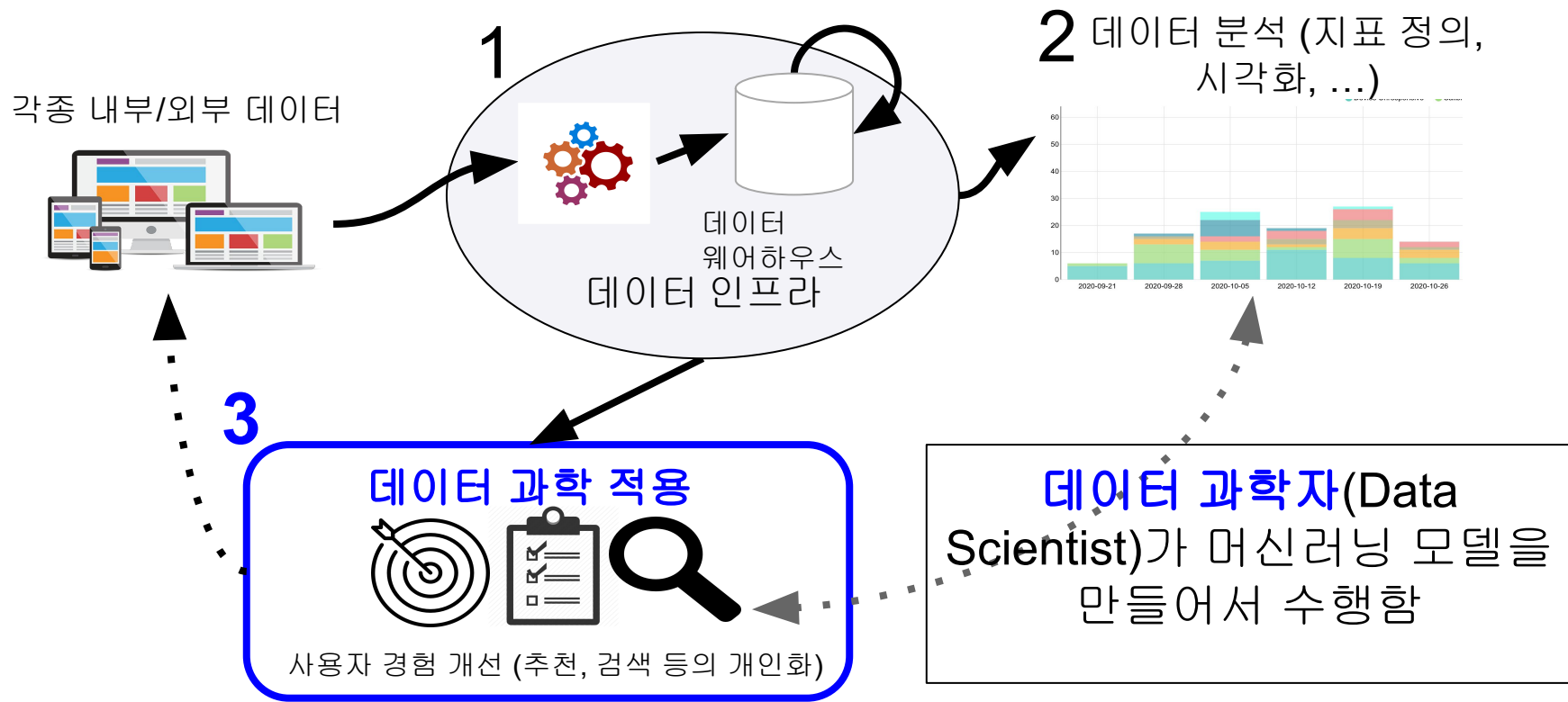
❖ 가장 널리 사용되는 대시보드:

- 세일즈포스의 태블로 (Tableau)
- 마이크로소프트의 파워 BI(Power BI)
- 구글 클라우드의 룩커(Looker)

Key Metrics Draft ☆

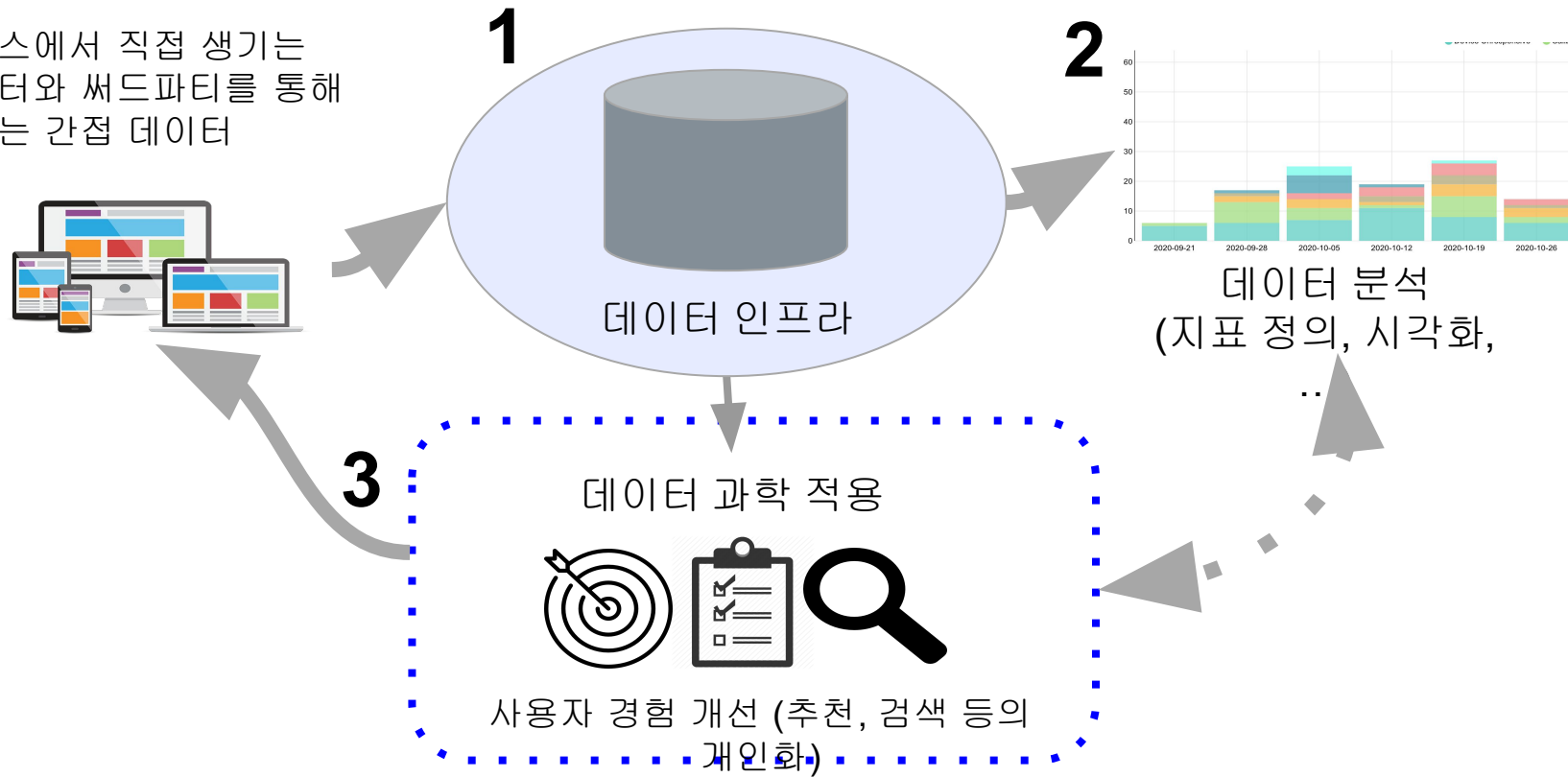


◆ 데이터 팀의 발전 - 3. 데이터 과학 적용



◆ 데이터의 흐름과 데이터 팀의 발전 단계

서비스에서 직접 생기는
데이터와 써드파티를 통해
생기는 간접 데이터





클라우드란?

클라우드가 무엇인지 알아보고 대표적인 클라우드인
AWS에 대해 알아보자

◆ 클라우드의 정의

- ❖ 컴퓨팅 자원(하드웨어, 소프트웨어)을 네트워크를 통해 서비스 형태로 사용하는 것
- ❖ 키워드:
 - “No Provisioning”
 - “Pay As You Go”
- ❖ 자원(예: 서버)을 필요한만큼 (거의) 실시간으로 지불
 - 탄력적으로 필요한만큼의 자원을 유지하는 것이



◆ 클라우드 컴퓨팅이 없었다면?

- ❖ 서버/네트워크/스토리지 구매와 설정을 직접 수행함
- ❖ 데이터센터 공간을 직접 확보 (Co-location)
 - 확장이 필요한 경우 공간을 먼저 더 확보해야함
- ❖ 그 공간에 서버를 구매하여 설치하고 네트워크 설정
 - 보통 서버를 구매해서 설치하는데 적어도 두세달은 걸림
- ❖ Peak time 기준으로 Capacity planning을 해야함!
 - 놓고 있는 자원들이 높게 되는 현상 발생
- ❖ 직접 운영비용 vs. 클라우드 비용
 - 기회비용!



source: <http://discountlowvoltage.blogspot.com/>

◆ 클라우드의 장점

- ❖ 초기 투자 비용이 크게 줄어듦
 - CAPEX (Capital Expenditure) vs. OPEX (Operating Expense)
- ❖ 리소스 준비를 위한 대기시간 대폭 감소
 - Shorter Time to Market
- ❖ 노는 리소스 제거로 비용 감소
- ❖ 글로벌 확장 용이
- ❖ 소프트웨어 개발 시간 단축
 - Managed Service (SaaS) 이용

of Servers: 100

CLICK

◆ AWS란?

❖ Amazon Web Services

❖ 가장 큰 클라우드 컴퓨팅 서비스 업체

❖ 2002년 아마존의 상품데이터를 **API**로 제공하면서 시작

- 현재 100여개의 서비스를 전세계 15개의 지역에서 제공.
- 대부분의 서비스들이 오픈소스 프로젝트들을 기반으로 함
- 최근 들어 ML/AI 관련 서비스들도 내놓기 시작

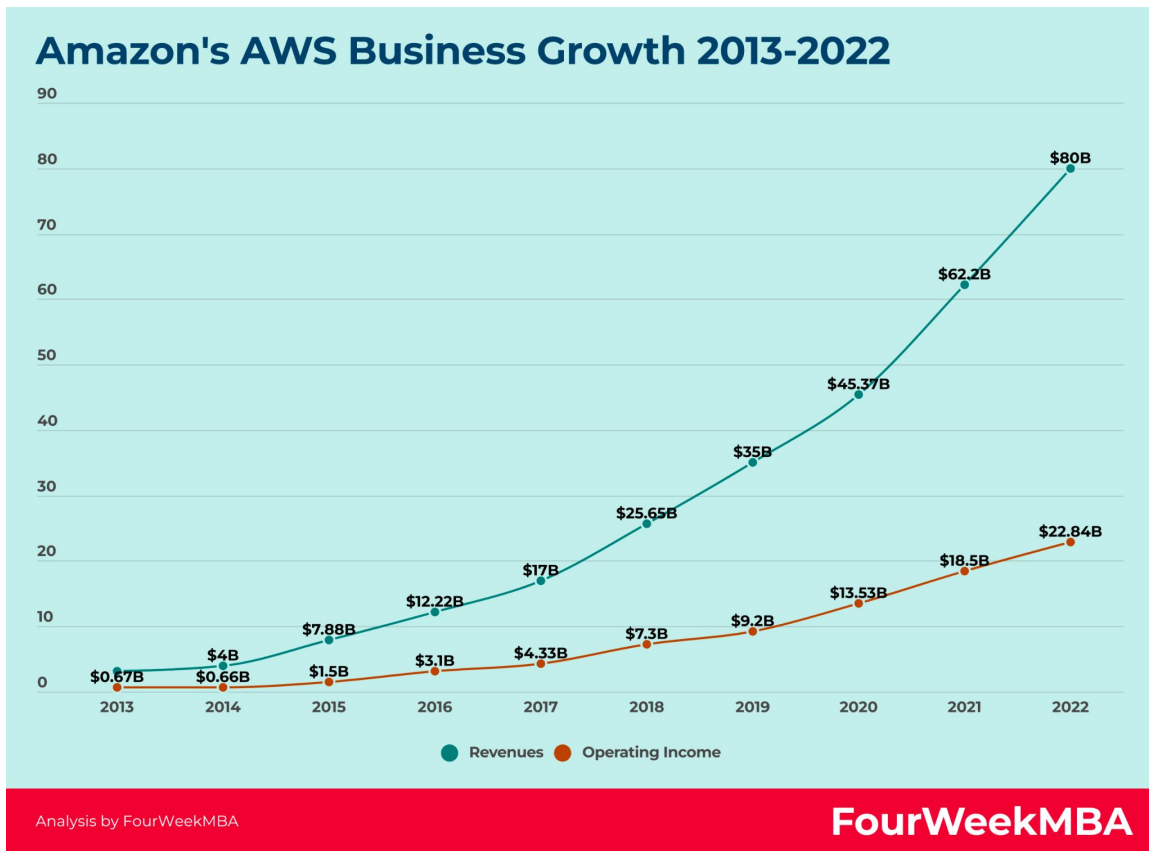
❖ 다양한 종류의 소프트웨어/플랫폼 서비스를 제

- AWS의 서비스만으로 쉽게 온라인서비스 생성
- 뒤에서 일부 서비스를 따로 설명.

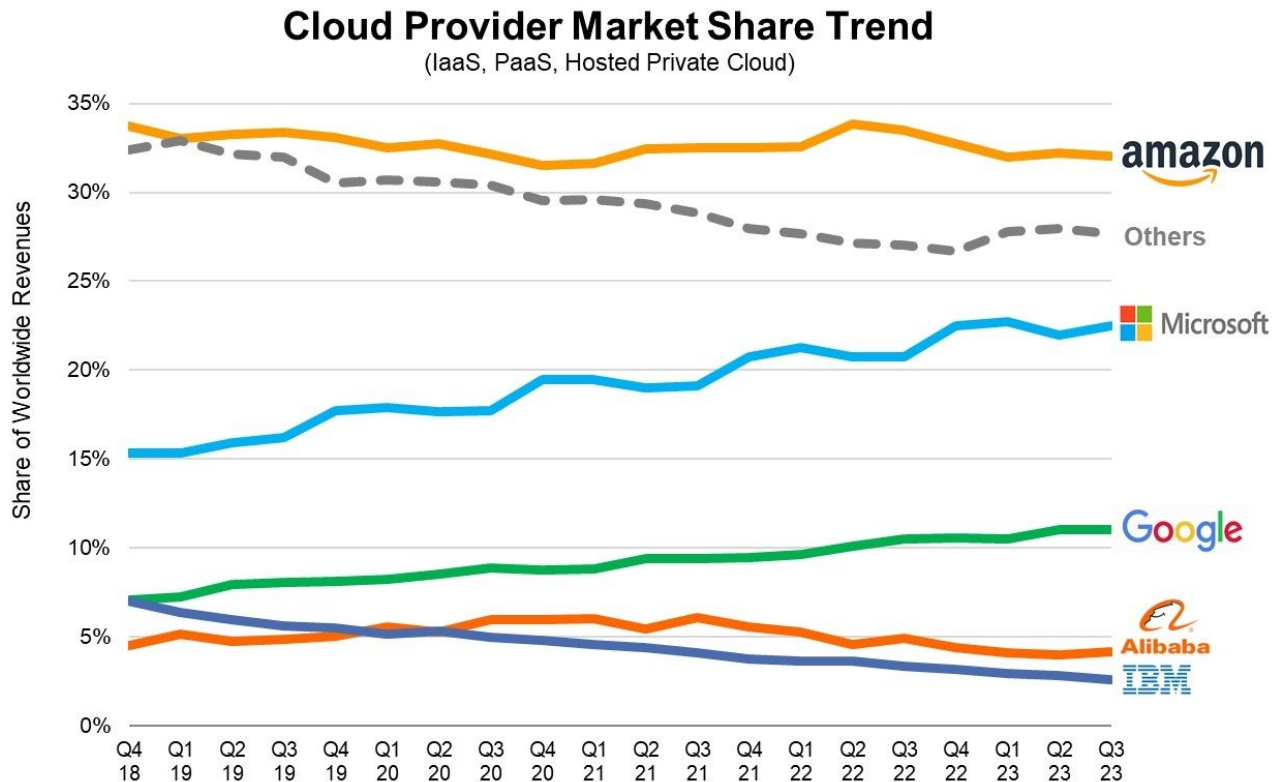


◆ AWS의 매출 트렌드

- ❖ 2022년 매출: \$80B
- ❖ 2022년 순익: \$22.84B



◆ 글로벌 클라우드 업체 순위




Source: Synergy Research Group

◆ AWS 진출 지역

Regions	Name
US East (Virginia)	us-east-1
US East (Ohio)	us-east-2
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Asia Pacific (Mumbai)	ap-south-1
Asia Pacific (Osaka)	ap-northeast-3
Asia Pacific (Seoul)	ap-northeast-2
Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
Canada (Central)	ca-central-1
China (Beijing)	cn-north-1
China (Ningxia)	cn-northwest-1
EU (Frankfurt)	eu-central-1
EU (Ireland)	eu-west-1
EU (Paris)	eu-west-3
EU (Stockholm)	eu-north-1
EU (London)	eu-west-2
South America (Sao Paulo)	sa-east-1
AWS GovCloud (US, US-East)	us-gov-west-1, us-gov-east-1


◆ AWS 제공 서비스들

Group




Compute

- EC2
- EC2 Container Service
- Lightsail [↗](#)
- Elastic Beanstalk
- Lambda
- Batch




Storage

- S3
- EFS
- Glacier
- Storage Gateway




Database

- RDS
- DynamoDB
- ElastiCache
- Redshift




Networking & Content Delivery

- VPC
- CloudFront
- Direct Connect
- Route 53




Developer Tools

- CodeCommit
- CodeBuild
- CodeDeploy
- CodePipeline




Management Tools

- CloudWatch
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog
- Trusted Advisor
- Managed Services
- Application Discovery Service




Security, Identity & Compliance

- IAM
- Inspector
- Certificate Manager
- Directory Service
- WAF & Shield
- Compliance Reports




Analytics

- Athena
- EMR
- CloudSearch
- Elasticsearch Service
- Kinesis
- Data Pipeline
- QuickSight [↗](#)




Artificial Intelligence

- Lex
- Polly
- Rekognition
- Machine Learning




Internet Of Things

- AWS IoT




Game Development

- GameLift




Mobile Services

- Mobile Hub
- Cognito
- Device Farm




Application Services

- Step Functions
- SWF
- API Gateway
- Elastic Transcoder




Messaging

- SQS
- SNS
- SES



Business Productivity

- WorkDocs
- WorkMail
- Amazon Chime [↗](#)



Desktop & App Streaming

- WorkSpaces
- AppStream 2.0

◆ AWS 제공 서비스 - EC2 (Elastic Compute Cloud)

❖ AWS의 서버 호스팅 서비스

- 리눅스 혹은 윈도우 서버를 론치하고 어카운트를 생성하여 로그인 가능
- 가상 서버들이라 전용서버에 비해 성능이 떨어짐
- Bare-metal 서버도 제공하기 시작

❖ 다양한 종류의 서버 타입 제공: <http://aws.amazon.com/ec2/>

- 예를 들어 미국 동부에서 스몰타입(t2.small)의 무료 리눅스 서버를 하나 할당시
 - 시간당 2.3 센트의 비용지불.
 - 2GB 메모리, 1 가상코어, 160GB 하드디스크
 - 2012년에는 8.5 센트였음
 - 타입별 지역별 가격을 알고 싶다면 [여기](#)를 방문
- Incoming network bandwidth는 공짜이지만 outgoing은 유료.

◆ AWS 제공 서비스 - EC2 (Elastic Compute Cloud)

❖ EC2 구매 옵션

구매 옵션	설명
On-Demand	시간당 비용을 지불되며 가장 흔히 사용하는 옵션
Reserved	1년이나 3년간 사용을 보장하고 1/3 정도에서 40% 디스카운트를 받는 옵션
Spot Instance	일종의 경매방식으로 놓고 있는 리소스들을 보다 싼 비용으로 사용할 수 있는 옵션

◆ AWS 제공 서비스 - S3 (Simple Storage Service)

- ❖ <http://aws.amazon.com/s3/>
- ❖ 아마존이 제공하는 대용량 클라우드 스토리지 서비스
- ❖ S3는 데이터 저장관리를 위해 계층적 구조를 제공
- ❖ 글로벌 네임스페이스를 제공하기 때문에 톱레벨 디렉토리 이름 선정에 주의.
- ❖ S3에서는 디렉토리를 버킷(Bucket)이라고 부름
- ❖ 버킷이나 파일별로 액세스 컨트롤 가능

◆ AWS 제공 서비스 - S3 (Simple Storage Service)

❖ <https://aws.amazon.com/ko/s3/pricing/>

❖ Low cost. 1TB per month:

- Standard storage: \$23
 - Infrequent Access storage: \$12.5
 - SLA가 다름
- Glacier storage: \$4

◆ AWS 제공 서비스 - 데이터베이스 관련 서비스들

❖ RDS (Relational Database Service)

- MySQL/MariaDB, PostgreSQL, Aurora
- Oracle, MS SQL Server

❖ DynamoDB

❖ Redshift

❖ ElastiCache

❖ Neptune (Graph database)

❖ ElasticSearch

❖ MongoDB

◆ AWS 제공 서비스 - AI & ML 관련 서비스들

❖ SageMaker

- Deep Learning and Machine Learning end-to-end framework
- MLOps Service!

❖ Lex

- Conversational Interface (Chatbot service)

❖ Polly

- Text to Speech Engine

❖ Rekognition

- Image Recognition Service

❖ Comprehend

- NLP Service => LLM (Large Language Model)



데이터 조직 구성원

데이터 조직을 구성하는 직군들은 무엇이 있는지 살펴보자

◆ 데이터 팀에는 누가 있는가?

- ❖ 조직에 따라 한 사람이 몇 개의 역할을 동시 수행하는 것이 일반적
- ❖ 데이터 엔지니어 (Data Engineer)
- ❖ 데이터 분석가 (Data Analyst)
- ❖ 데이터 과학자 (Data Scientist)
- ❖ ML 엔지니어
- ❖ MLOps 엔지니어
- ❖ 프라이버시 엔지니어
- ❖ ...

데이터 문해력: 문제와 데이터를 연결해서 가치 있는 결론을 내는 사고방식

- 파이썬, 자바, 스칼라 (github)
- SQL, 데이터 웨어하우스

- ETL (Airflow)
- 클라우드 (AWS,GCP,Azure)

- 빅데이터 분산처리 기술 이해
- Spark/Hadoop

- 컨테이너 관련 기술 (Docker,K8S)

- NoSQL (카산드라, 몽고DB, ...)

- 가설 기반 접근법
- 머신러닝 기술 (딥러닝, NLP, ...)
- 데이터 청소, EDA, Feature 엔지니어링
- A/B 테스트 설계, 프레이밍셋, 모델 관리 (알고리즘, 하이퍼 파라미터)
- ML 개발 프레임워크 (SageMaker, MLflow, Kubeflow)

- KPI/지표
- 통계 지식
- A/B 테스트 분석
- 데이터 모델링/데이터 분석
- 비즈니스 도메인에 대한 지식
- BI 툴 (Looker, Tableau, ...)

dbt

개인정보 보안: 위반시 페널티가 커지면서 점점 더 중요한 요소가 되고 있음 (GDPR)

AI 윤리: Trustworthy AI

- 모델 배포, 모델 모니터링, CT/CI/CD(Continuous Training,Integration,Delivery)

AI를 사용한 업무 효율화

◆ 데이터 엔지니어 (Data Engineer)

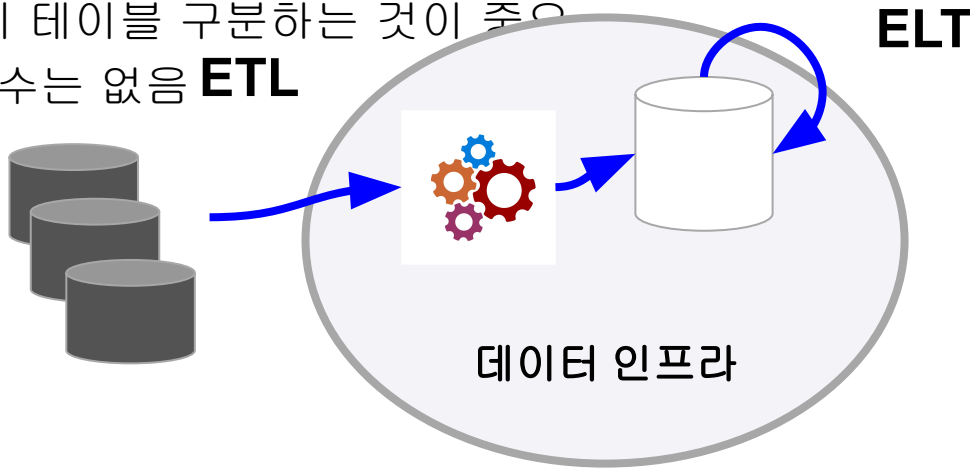
- ❖ 기본적으로 소프트웨어 개발자 (Python, SQL, Airflow, Spark, ...)
- ❖ 데이터 인프라 (데이터 웨어하우스/레이크와 ETL) 구축
- ❖ 내부/외부 데이터를 데이터 웨어하우스로 가져오는 역할을 수행
- ❖ 보통 외부 요청에 의해 새로운 데이터 소스를 추가
 - 비즈니스 오너를 정하는 것이 중요 -> 품질 관리 및 이슈 발생시 노티
 - PII 등의 데이터 분류가 중요 (태그 등 사용)
 - PII: Personal Identifiable Information

◆ 데이터 분석가 (Data Analyst)

- ❖ 데이터 웨어하우스의 데이터를 기반으로 지표를 만들고 시각화 (대시보드)
 - ELT를 수행해서 새로운 데이터 생성 (비즈니스 오너 지정이 중요)
 - DBT와 같은 툴을 사용하는 것이 추세
- ❖ 내부 직원들의 데이터 관련 질문 응답

◆ ETL vs. ELT

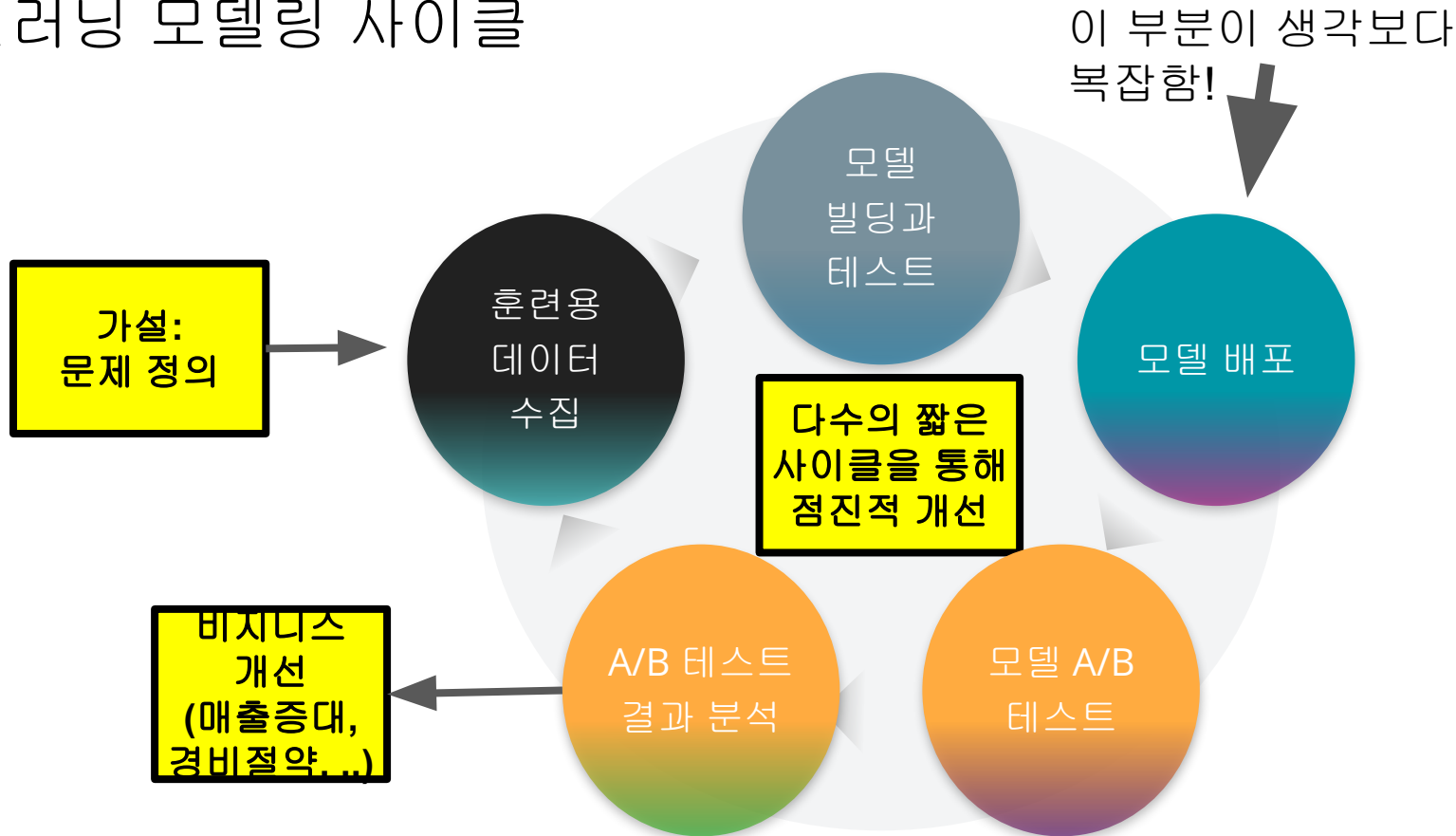
- ❖ ETL: 데이터 시스템 밖에 있는 데이터를 안으로 가져오는 것
- ❖ ELT: 데이터 시스템 안에 있는 데이터를 조합하여 새로운 데이터를 만드는 것
 - 데이터 리니지가 중요해짐
 - Not all tables are equal
 - 중요 테이블과 테스트/임시 테이블 구분하는 것이 중요
 - 모든 테이블 관리를 잘 할 수는 없음



◆ 데이터 과학자 (Data Scientist)

- ❖ 과거 데이터를 기반으로 미래를 예측하는 머신러닝 모델 생성
 - 고객들의 제품/서비스 사용 경험 개선 (개인화 혹은 자동화 혹은 최적화)
- ❖ 데이터 수집에 있어 왜곡이 있는지, 혹시 개인 정보를 사용하고 있는지?
- ❖ 모델의 동작에 대해 설명할 수 있는지?
- ❖ 모델 개발 뿐만 아니라 배포 과정이 자동화되어야 함 => MLOps

◆ 머신러닝 모델링 사이클



◆ A/B 테스트

❖ A/B 테스트 = 실험 (Split Test or Bucket Test)

- Randomized Controlled Trial의 온라인 버전
- 가설이 필요
- 실제 사용자를 대상으로 진행

❖ 다수의 Variant로 구성됨

- 하나의 컨트롤 (기존 버전)과 하나 이상의 테스트





데이터 문해력의 정의와 중요성

데이터를 수집하는 것은 시작일 뿐 데이터를 활용해야 한다

◆ 데이터 문해력(Data Literacy)이란?

- ❖ 보통 문해력이란 글을 읽고 쓰는 능력
- ❖ 데이터 문해력은 데이터를 이해하고 활용할 수 있는 능력
 - 레벨에 따라 요구 조건은 달라짐
 - 기업 도메인과 성숙도에 따라 굉장히 다양한 수준이 존재
- ❖ 결국 데이터를 다음과 같이 활용하는 능력
 - 데이터 기반 의사 결정
 - 데이터 기반 제품 개선
 - 데이터 (GenAI) 기반 생산성 증대



◆ 데이터 문해력 발전 트렌드 (1)

❖ 데이터/IT 조직 → 회사 전체로 문해력 향상/전파 발생

- 결국 데이터는 자산(**Asset**)이라는 인식이 먼저 필요
- 요즘 트렌드는 모든 조직에서 데이터를 활용하거나 활용하는데 관심이 많음 (**Decentralization**)
 - 생산성 증대가 중요해짐

❖ 점점 더 많은 인력들이 데이터 활용 능력을 갖게 됨

- 시민 데이터 분석가 (**Citizen Data Analyst**)
- 시민 데이터 과학자 (**Citizen Data Scientist**)

◆ 데이터 문해력 발전 트렌드 (2)

❖ Gen AI와 같은 툴을 사용한 생산성 증대

- ChatGPT와 같은 AI 툴을 이용한 업무 효율성 증대가 가능해짐

- No Code 혹은 Low Code 툴들이 발전하고 있음

❖ 개발 업무 뿐만 아니라 다양한 업무에 영향을 주고 있음

- 마케팅, CS, 세일즈, ...

◆ 데이터 관리의 중요성

❖ 잘못 관리된 데이터는 커다란 위험 요소

- 2023-01: Database of Over 200m Twitter Users Goes Public
- 2022-12: Slack Code Repositories Compromised
- 2022-12: Google Fined \$57M by Data Protection Watchdog Over GDPR Violations

❖ 개인정보 보호 관련 법안

- GDPR
- CCPA (CPRA)
- HIPAA
- The right to know
- The right to delete
- The right to opt-out
- The right to non-discrimination

◆ 데이터 거버넌스가 필요해짐

❖ 데이터 거버넌스란?

- 구글의 정의: "데이터의 보안, 개인정보 보호, 정확성, 가용성, 사용성을 보장하기 위해 수행하는 모든 작업으로 여기에는 사람들이 취해야 하는 조치, 따라야 하는 프로세스, 데이터 수명 주기 전반에 걸쳐 이를 지원하는 기술이 포함됨"



데이터 일을 할 때 기억할 점

데이터 팀을 운영하면서 배운 교훈을 정리해보자

◆ 데이터를 통해 매출이 생겨야 한다

❖ 어느 조직이건 회사에서의 존재 이유는 매출 창조 혹은 경비 절감

- 데이터 인프라와 데이터 팀원(데이터 과학자)의 몸값은 상대적으로 높음
- 직접적이건 간접적이건 데이터를 통해 회사 수익에 긍정적인 영향을 끼쳐야함

❖ 데이터 조직의 수장의 역할이 아주 중요

- 리더십 팀과 주변 팀들이 데이터 팀으로부터 바라는 기대를 잘 관리
 - 데이터 인프라의 구성이 첫 번째라는 점을 잘 설명하면서 단기적으로 좋은 결과를 낼 방법을 찾아야함
- 회사 중요 지표에 데이터 팀이 끼치는 영향을 객관적으로 신뢰가능하게 챙겨야함
- 이게 가능하면 데이터 조직을 중앙집중 혹은 하이브리드 형태로 운영이 가능함

◆ 데이터 인프라가 첫 번째 스텝!

❖ 데이터 인프라 없이는 데이터 분석이나 모델링은 불가능

- 하지만 아주 작은 회사에서 생존이 더 중요한 문제라 데이터 인프라는 조금더 성장한 뒤에 걱정해도 됨
- 첫 번째 팀원은 인프라 구축 이외에도 약간의 분석/모델링 스킬이 있는 사람이 최적

❖ 고려점

- 클라우드 **vs.** 직접 구성
- 배치 **vs.** 실시간

◆ 데이터의 품질이 아주 중요

❖ Garbage In Garbage Out

❖ 데이터 과학자가 가장 많은 시간을 쏟는 분야는?

- 데이터 청소 작업!
- 모델링에 드는 시간을 100이라고 하면 그중 70은 데이터 클린업에 들어감

❖ 중요 데이터의 경우 좀더 품질 유지에 노력이 필요

- 어디에 데이터가 있는지?
- 이 데이터의 품질에 혹시 문제가 있는지 계속적으로 모니터링

“If you are not thinking about how to keep your data clean from the very beginning, you are fucked. I guarantee it.”

◆ 항상 지표부터 생각

❖ 무슨 일을 하건 그 일의 성공 척도(지표)를 처음부터 생각

- 또한 나름대로 가설을 세우는 것이 인사이트를 키우는데 큰 도움이 됨

❖ 지표의 계산에 있어서 객관성이 중요

- 계산된 지표를 아무도 못 믿는다면 큰 문제
- 지표를 어떻게 계산할 것인지 그리고 이걸 다른 사람들에게 어떻게 설명할지 고려

◆ 가능하면 간단한 솔루션으로 시작

❖ 모든 문제를 딥러닝으로 해결해야 하나?

- IF문 몇개의 간단한 논리로 해결할 수 있는지 부터 고민!
- 실제 회사에서 딥러닝으로 문제를 해결하는 경우는 드뭄. 왜 동작하는지 설명도 힘들고 개발과 론치 모두 시간이 걸림

❖ 반복 기반의 점진적인 개발방식 vs. 한 큐에 모델 만들기

- 후자는 시간만 오래 걸리고 최종 성과는 안 좋을 확률이 높음
- 전자로 가면서 원하는 결과가 나오면 그 때 중단. 더 개선할 필요 없음





퀴즈

이번 강좌에서 배운 내용을 퀴즈로 풀어보자

❖ 데이터 팀의 역할 퀴즈

- 아래 퀴즈를 풀어보세요!
 - <https://forms.gle/nCpS5VCyZ1D2yGT5A>



Q & A

이번 강의에 질문이 있으면 알려주세요!