

introduction

# Fundamentals of database

Sejong Oh

Bio Information Technology Lab.

- Course description

- This course presents the basic theory of database and the standard language SQL.
- Students are expected to learn the notions of database systems, DBMS, Data models, and standard languages. Students are, also, expected to practice these notions using MySQL and apply them in various situations.
- DBMS management tool, ODBC
- DBMS : MySQL
- DB security, ERD

제품화

- Text

- None

- Evaluation
  - Midterm exam : 30%
  - Final exam : 30%
  - Exercise reports, quiz : 30%
  - Attendance : 10%

# Data is important



Welcome to data world !

# Data is important

<https://realestate.daum.net/news/detail/main/20190215060805785>

## 줄줄새는 아파트 관리비도 꿈쩍마!.."빅데이터만 돌렸는데 한집 3만원씩 줄었죠"

[빅데이터로 뚝뚝해지는 행정] ②실제 성공적 적용 사례  
경기도만 쏙단지 감사에 240년.. 빅데이터로 단축 가능  
관리비와 세부 원천데이터 비교.. 152억 부당부과 적발  
표준분석모델, 他지자체 보급.. "전국 적용땐 1.1조 절감"  
행안부 "국민 직접체감 기대.. 모델 확산 적극 나설 것"

<https://yukaichou.com/loyalty/big-data-how-target-knows-you-are-pregnant/>



- 해외 사례: 타깃의 맞춤형 광고

한 남성이 화가 잔뜩 난 채로 대형마트에 들어서더니 “매니저 나와”라고 소리 질렀다. 손에는 이 대형마트가 딸에게 우편으로 보낸 아기 옷과 침대 등 유아용품 할인쿠폰이 들려 있었다. 남성은 “내 딸은 아직 고등학생인데 이런 쿠폰을 보내다니. 임신을 부추기는 거냐”고 따졌다. 영문을 몰랐던 매니저는 우선 남성을 진정시키고 돌려보냈다. 며칠 뒤 재차 사과를 하려고 전화를 건 매니저에게 뜻밖의 반응이 돌아왔다. 이 남성은 “딸이 임신을 한 걸 뒤늦게 알게 됐다”며 사과했다. 미성년 딸이 부모 모르게 아이를 가졌던 것이다. 대형마트 마케팅팀은 부모도 몰랐던 딸의 임신 사실을 어떻게 알아챘을까? 답은 ‘빅데이터 분석’에 있다. 이 얘기는 미국 대형마트 ‘타깃’의 미니애폴리스 점포에서 실제로 일어났던 일이다. <뉴욕 타임스>가 보도한 이 사례는 빅데이터 분석의 힘을 여실히 보여준다. 대형마트는 고객이 가족에게도 알리지 않은 비밀을 찾아내 앞으로 무엇이 필요할지 정확하게 예측해냈다.

# Data is important

타깃의 빅데이터 전문가들은 고객의 25가지 구매 행태를 분석하면 여성의 임신과 출산을 상당히 정확하게 예측할 수 있다는 사실을 확인했다. 예를 들어 향이 나는 로션을 사던 여성이 무향의 로션으로 바꾸거나, 평소 사지 않던 미네랄 영양제를 갑자기 사들이는 경우다. 타깃은 고객 데이터 베이스에 이를 적용했고, 전국적으로 수만명의 임신 추정 여성들을 가려내 관련 할인쿠폰을 보냈다. 이를 통해 고객의 구매 행동을 이끌어낼 수 있기 때문이다.



<https://theamericangenius.com/tech-news/target-remodeling-stores/>



# 1. Data

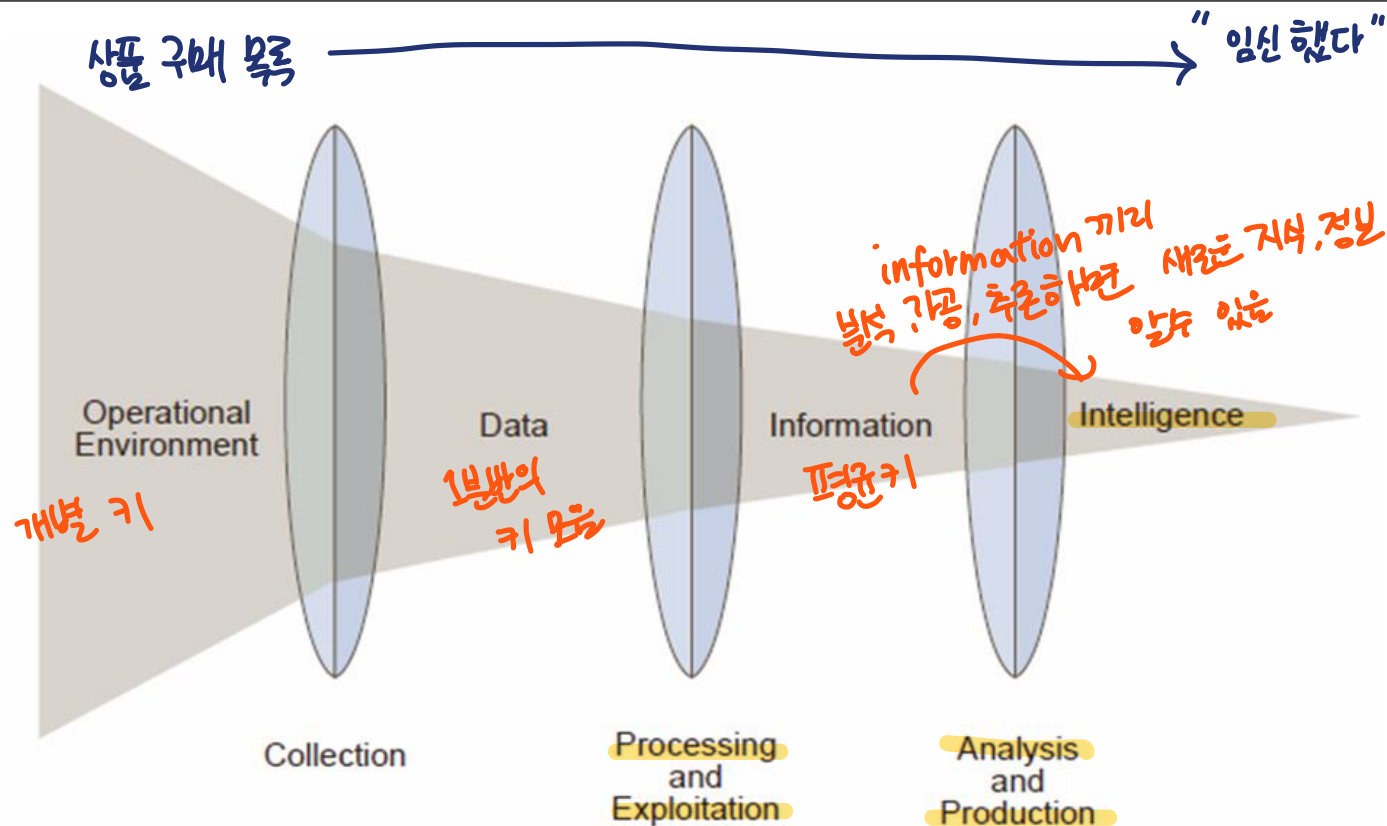
- Definition <sup>한글</sup> <sup>data</sup> <sup>fact</sup>
  - 추정과 추론의 근거가 되는 사실 (옥스포드 사전)
  - A thing given or granted; something known or assumed as fact, and made the basis of reasoning or calculation;
  - Data is a set of values of subjects with respect to qualitative or quantitative variables. (wiki)
    - ↓ \* 하나하나의 값들은  
data 모은 것 ⇒ data
  - data is collected by a huge range of organizations and institutions, including businesses
  - Data is measured, collected and reported, and analyzed, whereupon it can be visualized using graphs, images or other analysis tools



# 1. Data

- Data, information, intelligence

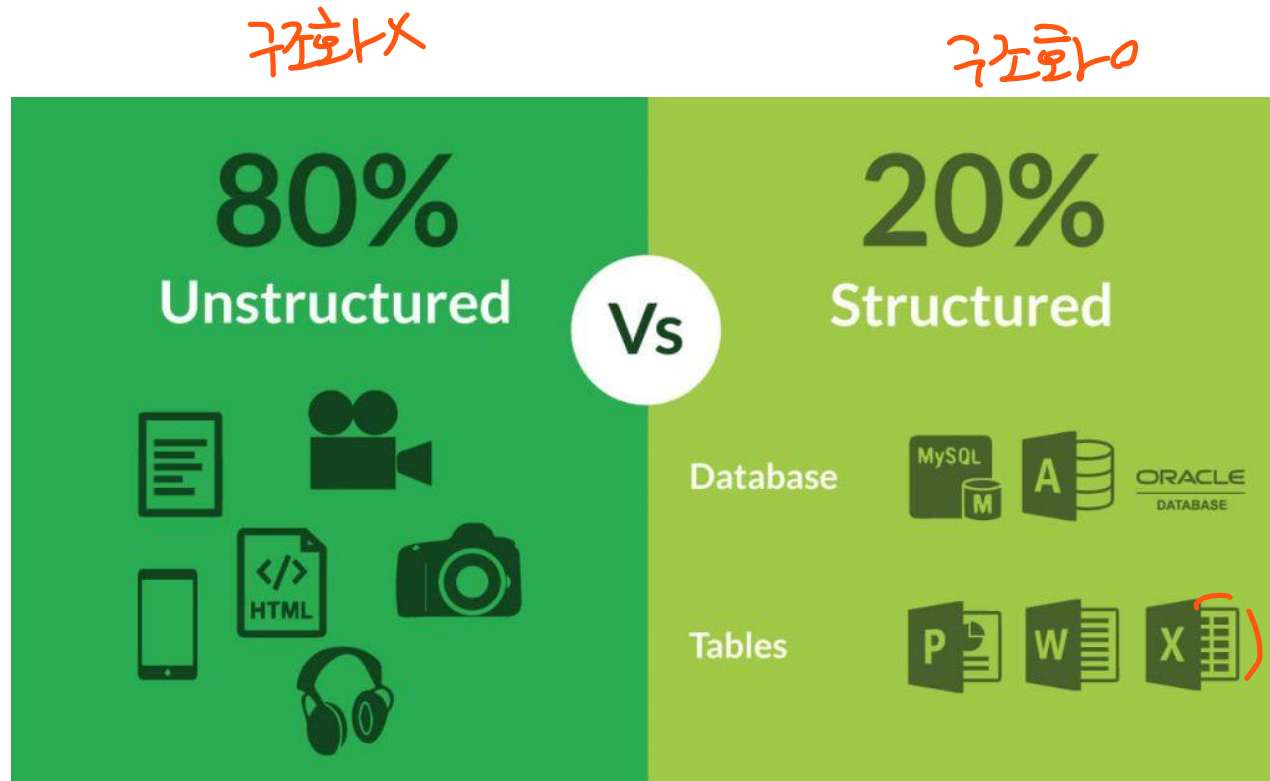
## Relationship of Data, Information and Intelligence



Source: Joint Intelligence / Joint Publication 2-0 (Joint Chiefs of Staff)

# 1. Data

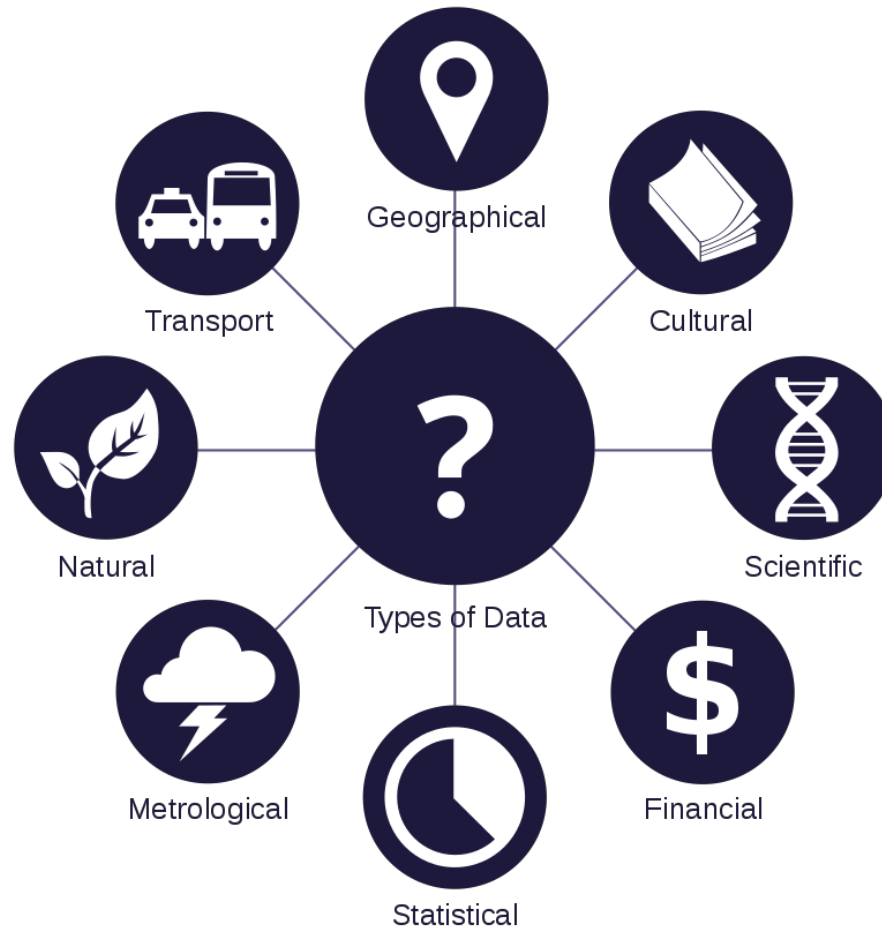
- Types of data



<http://www.ndimensionz.com/kb/volume-velocity-and-variety-three-vs-of-big-data-management/>

# 1. Data

- Types of data



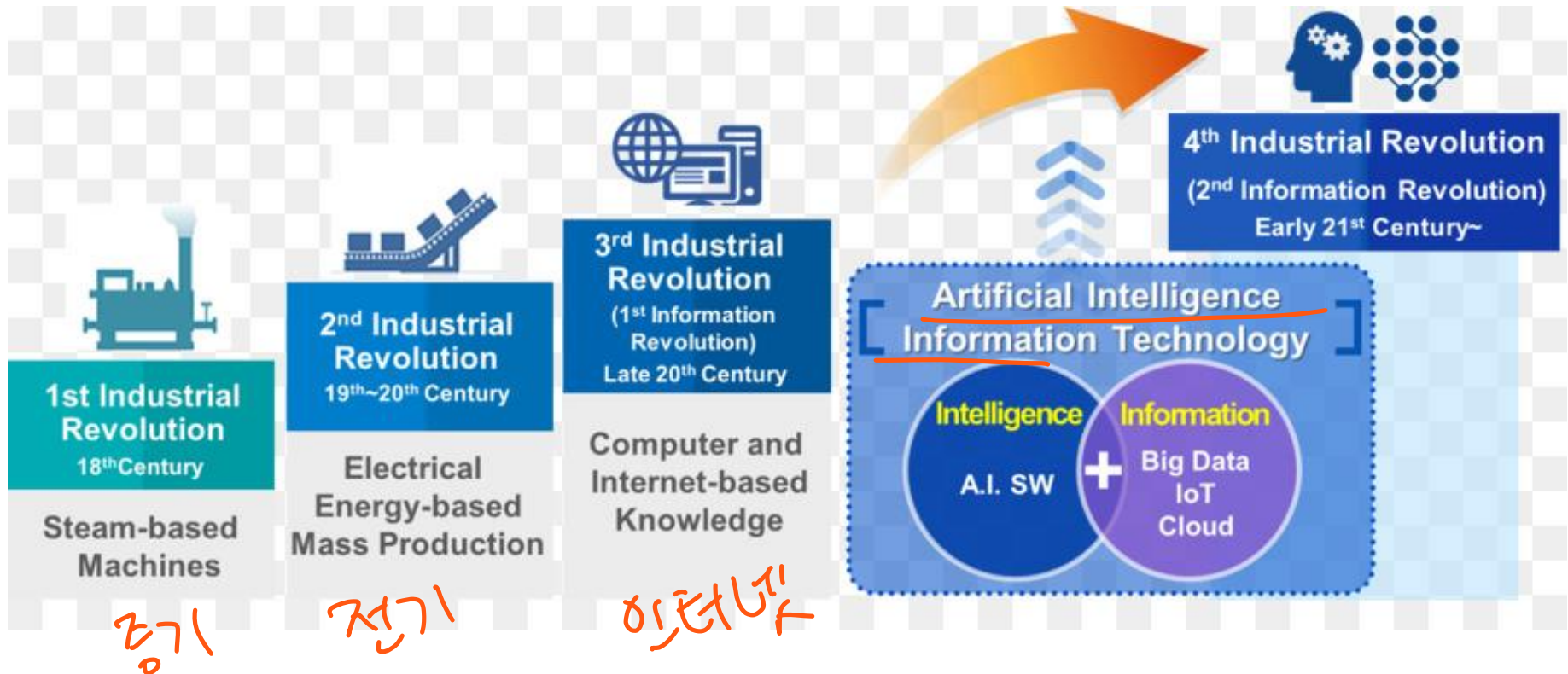
# 1. Data

- The age of data
  - “Data are becoming the new raw material of business” - Craig Mundie, Microsoft
  - “Data tab has been turned on and will never be turned off” – Mike Hoskins, Actian CTO, 2013
  - “You can have data without information, but you cannot have information without data” – Writer
  - “It is a capital mistake to theorize before one has data” – Sherlock Homes
  - “Data is the new oil” – Clive Humby, ANA Senior marketer’s submit



# 1. Data

- 4<sup>th</sup> industrial revolution & data



<https://banner2.kisspng.com/20180510/ahq/kisspng-fourth-industrial-revolution-artificial-intelligen-5af46a3f716053.6382638915259674234644.jpg>

# 1. Data

## MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

### MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

### PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

### DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

### COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



[https://cdn-images-1.medium.com/max/1200/1\\*T5GfsoZ-IWK3rcVkZ7R2bw.png](https://cdn-images-1.medium.com/max/1200/1*T5GfsoZ-IWK3rcVkZ7R2bw.png)

**MarketingDistillery.com** is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization; customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.

**Marketing  
DISTILLERY**

DATA LAB.



## 2. Big Data

- Big data

- Big data refers to data sets that are too large or complex for traditional data-processing application software to adequately deal with.

### 빅 데이터 생성 속도

- 하루 **250경** 바이트 비정형 정보
- 매달 **10억** 여 개 트윗
- 매달 **300억** 여 개 페이스북 메시지
- **1조** 대 이상 모바일 기기로 가속화





## 2. Big Data

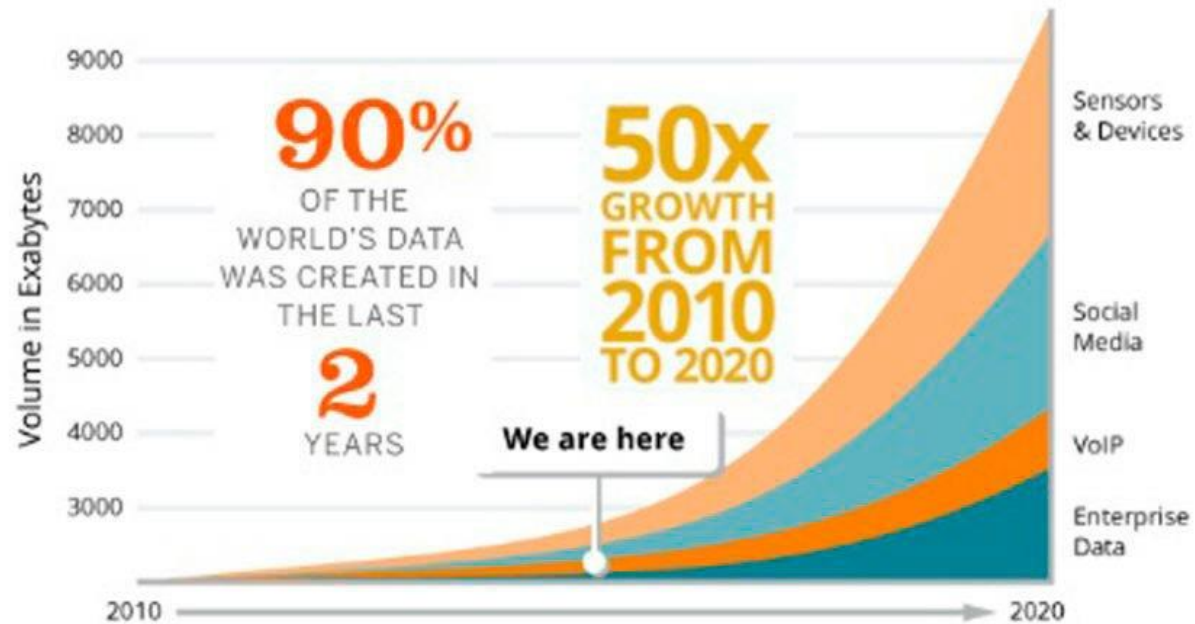
- Grows of Data

### CONTEXT: WHAT'S BIG DATA?

7

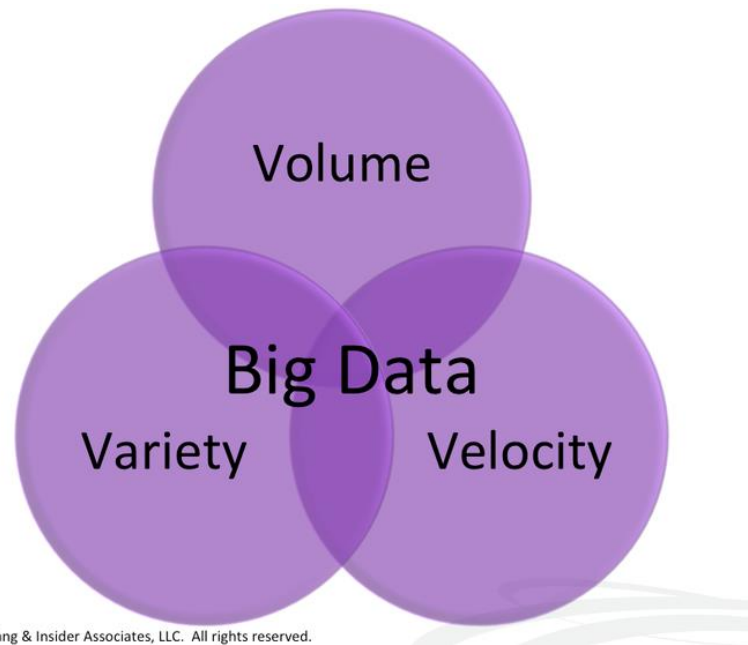
### BIG IN GROWTH, TOO.

1 exabyte (EB) = 1,000,000,000,000,000 bytes



## 2. Big data

- TDWI, BI/DW research institute, defined 3V of big data



## 2. Big data



[출처:채승병 외, 빅데이터: 산업 지각변동의 진원, SERI CEO Information, 2012. 5]

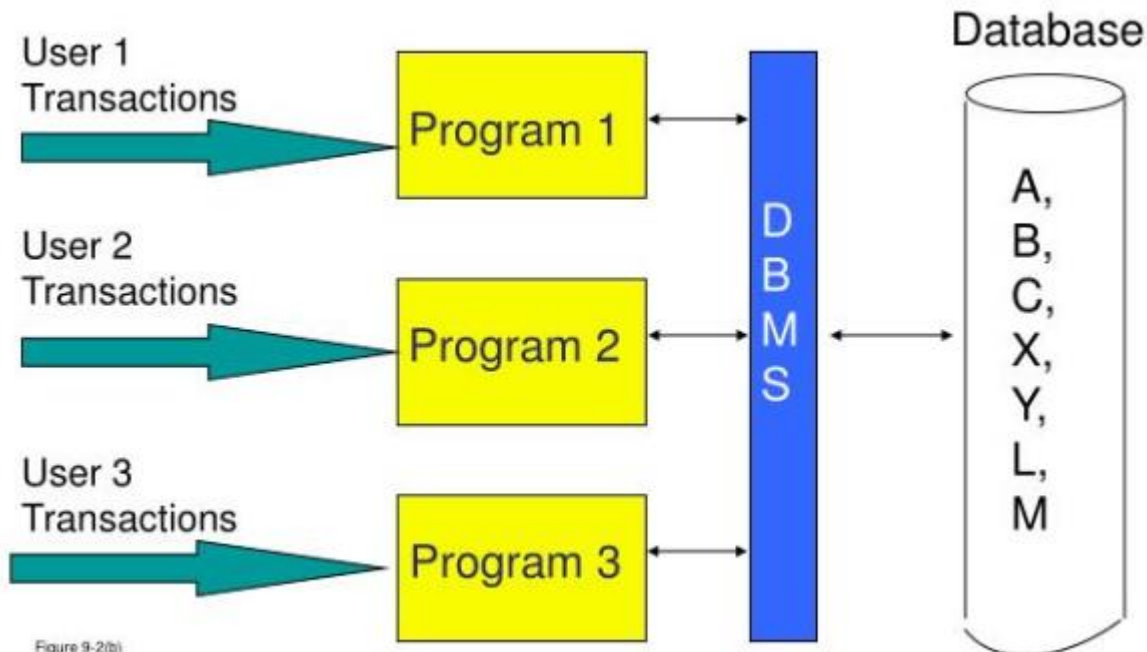
### 3. Database

데이터를 구조화해서 잘 모아놓은 것

- A **database** is an organized collection of data, generally stored and accessed electronically from a computer system. (wiki)

Database 관리 시스템

- The database management system (**DBMS**) is the software that interacts with end users, applications, and the database itself to capture and analyze the data.



# 3. Database

- products



<https://www.novelucent.com/templates/novelucent/images/database-services.jpg>

## 4. Setup practice environment

- Install by your self

- MySQL server DBMS
- MySQL workbench
- MySQL shell

서버와 communicate  
하기 위한

## 4. Setup practice environment

- <https://dev.mysql.com/downloads/installer/>

### MySQL Community Downloads

MySQL Installer


**General Availability (GA) Releases** Archives ⓘ

### MySQL Installer 8.0.23

Select Operating System:  
Microsoft Windows ▼

Looking for previous GA versions?

<b>Windows (x86, 32-bit), MSI Installer</b> (mysql-installer-web-community-8.0.23.0.msi)	8.0.23	2.4M	<b>Download</b>
	MD5: a3af6d91f93e046452b38a1e2589534c   <a href="#">Signature</a>		
<b>Windows (x86, 32-bit), MSI Installer</b> (mysql-installer-community-8.0.23.0.msi)	8.0.23	422.4M	<b>Download</b>
	MD5: 8de85ced955631901829a1a363cdf50   <a href="#">Signature</a>		

 We suggest that you use the [MD5 checksums](#) and [GnuPG signatures](#) to verify the integrity of the packages you download.



## 4. Setup practice environment

### MySQL Community Downloads

Login Now or Sign Up for a free account.

An Oracle Web Account provides you with the following advantages:

- Fast access to MySQL software downloads
- Download technical White Papers and Presentations
- Post messages in the MySQL Discussion Forums
- Report and track bugs in the MySQL bug system

**Login »**

using my Oracle Web account

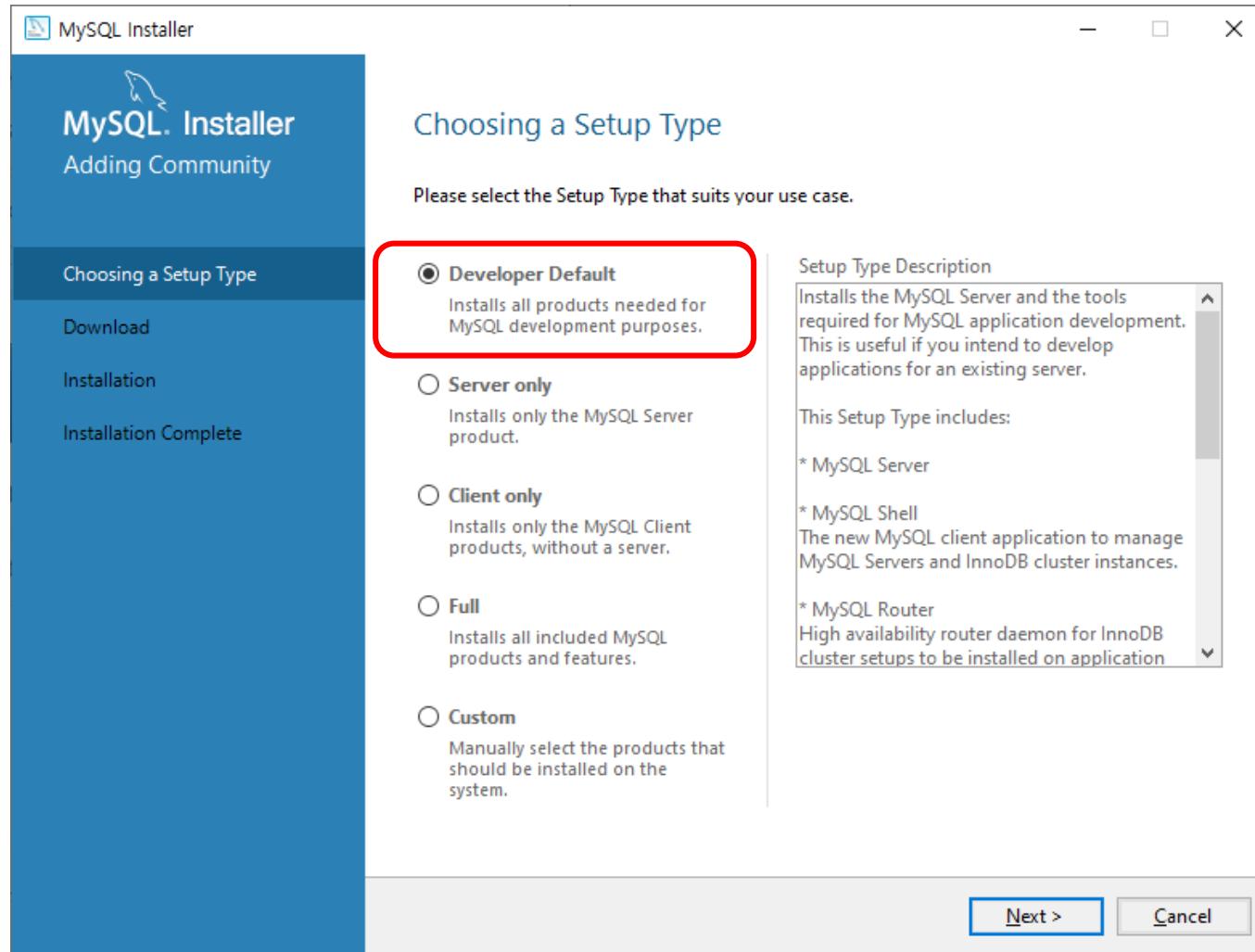
**Sign Up »**

for an Oracle Web account

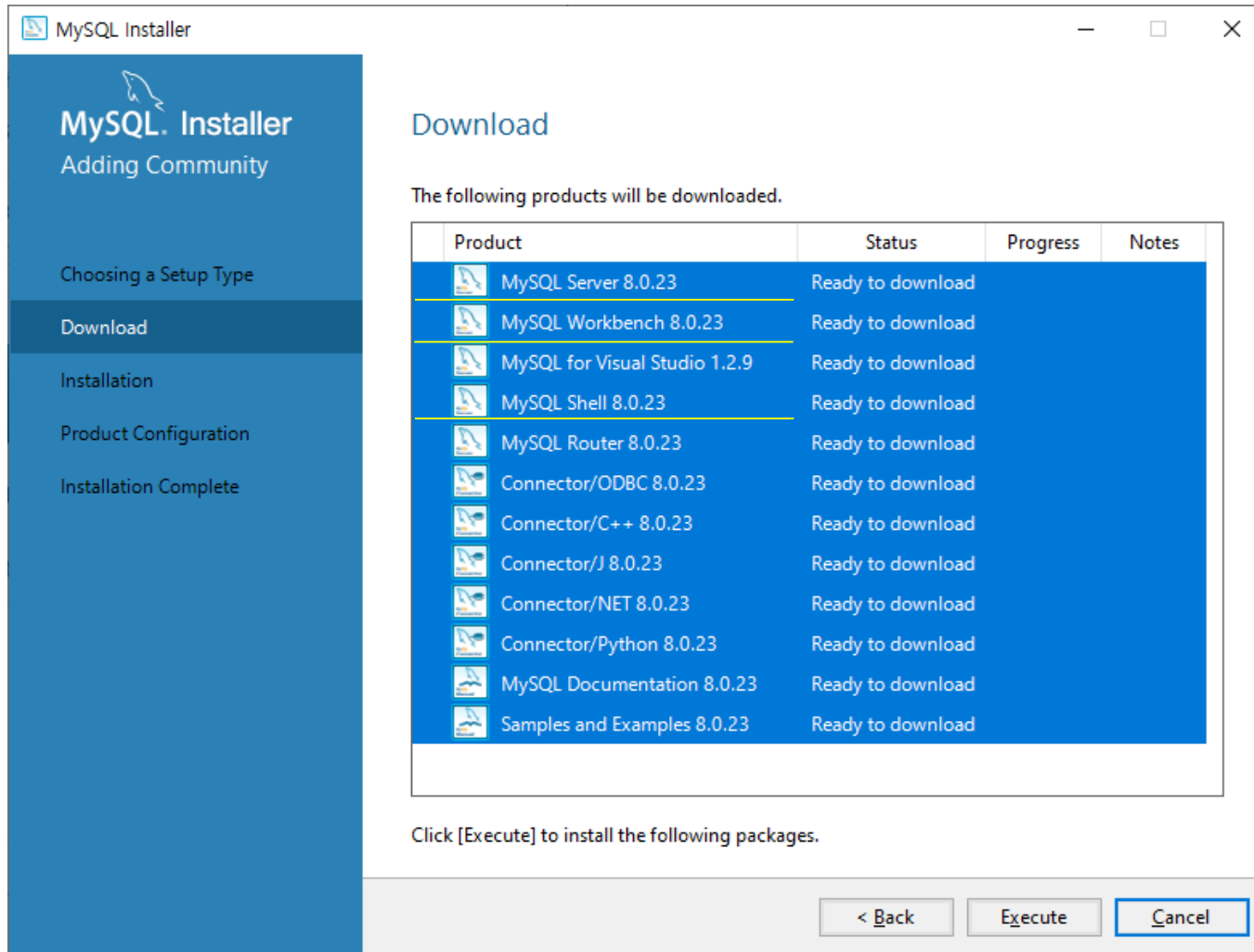
MySQL.com is using Oracle SSO for authentication. If you already have an Oracle Web account, click the Login link. Otherwise, you can signup for a free account by clicking the Sign Up link and following the instructions.

**No thanks, just start my download.**

# 4. Setup practice environment

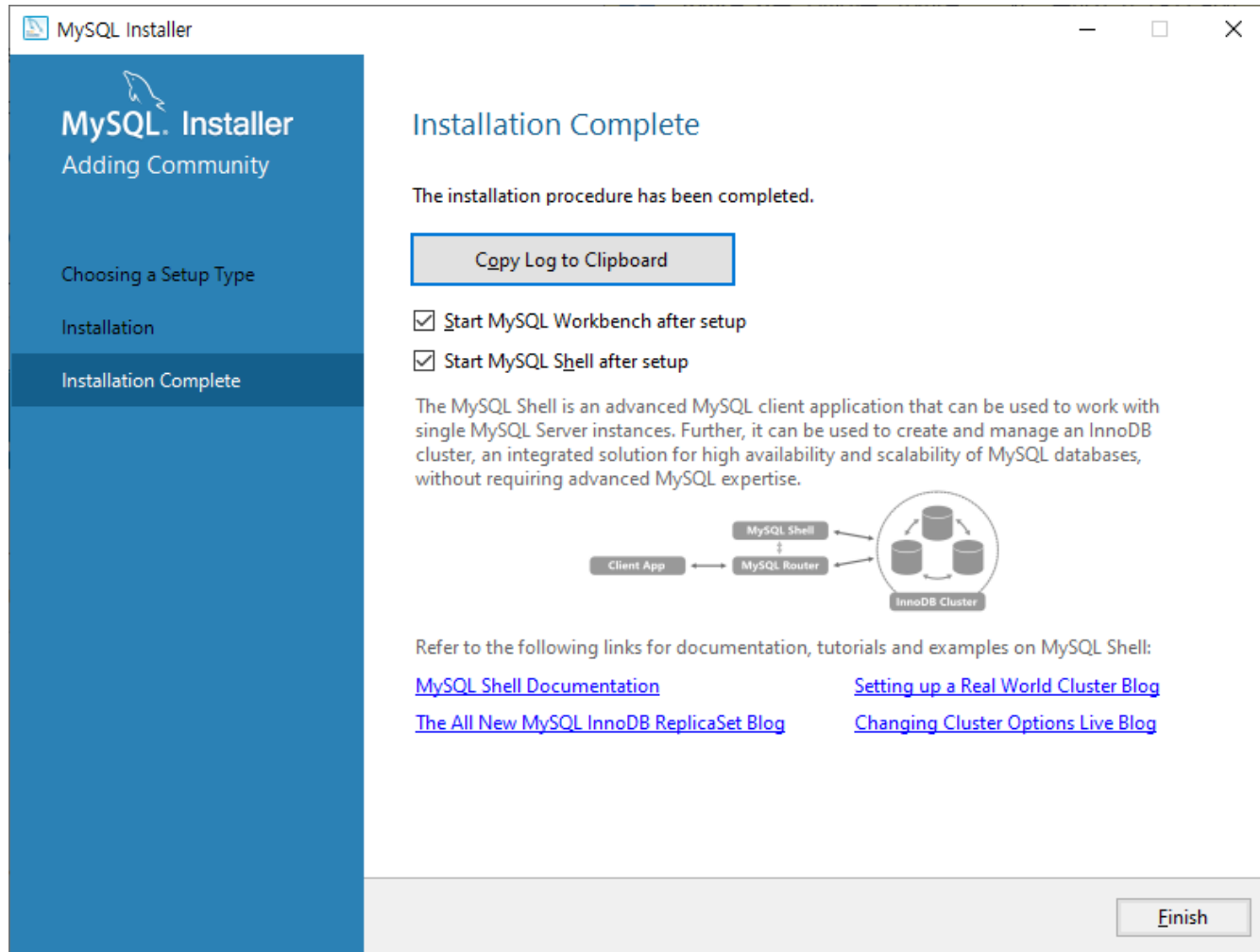


# 4. Setup practice environment

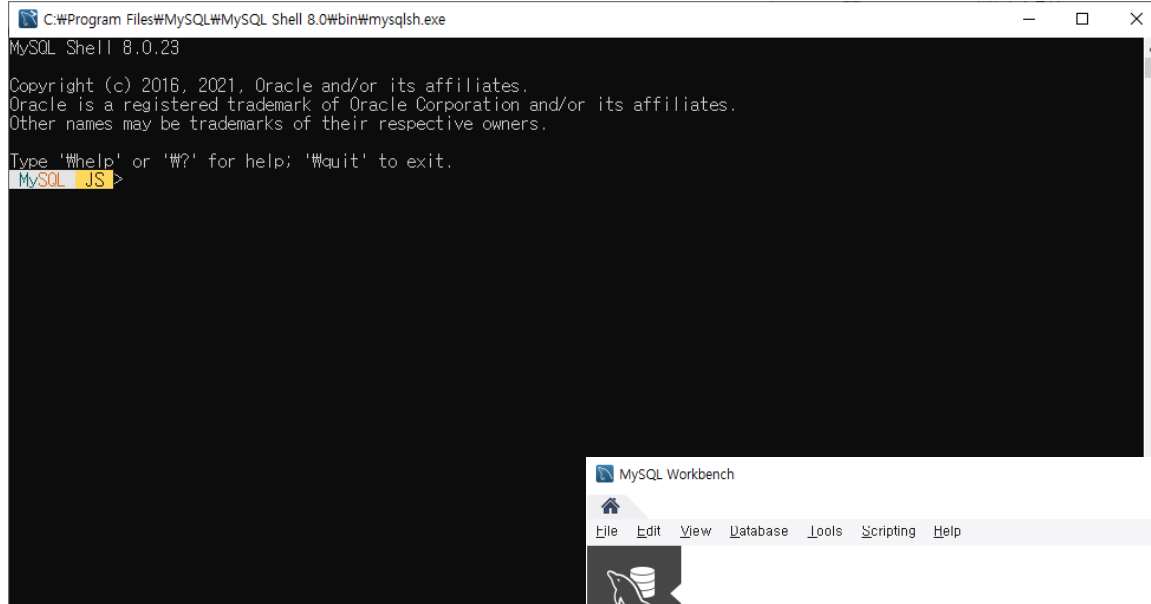


# 4. Setup practice environment

root @w: km@1032203



# 4. Setup practice environment

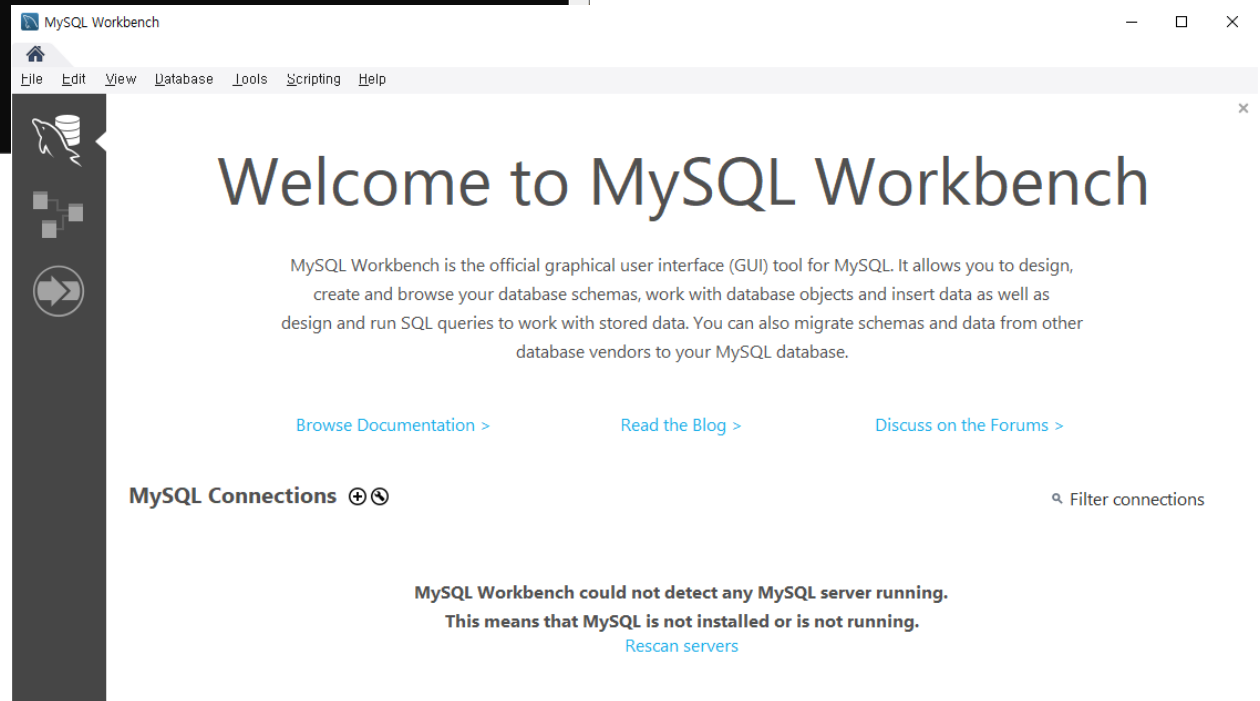


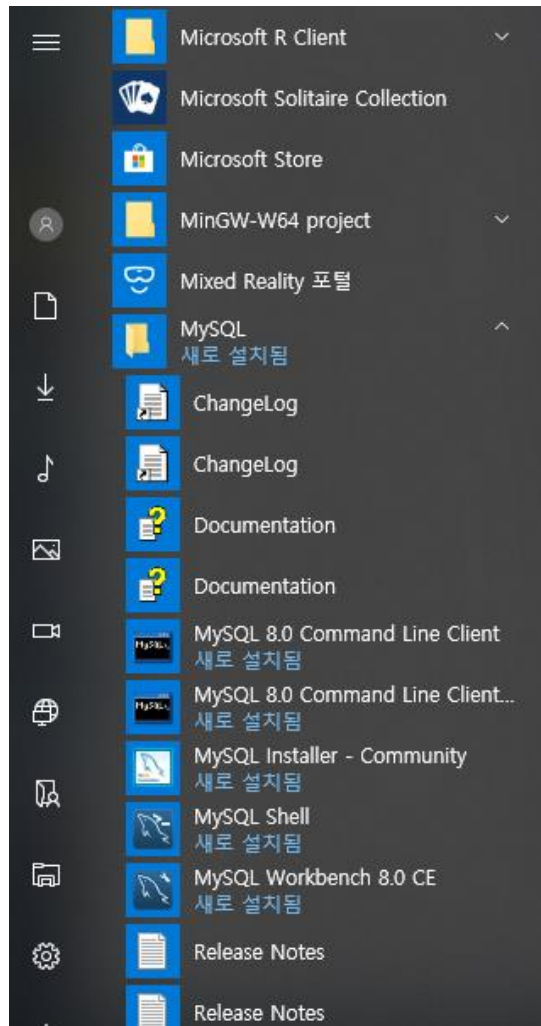
A screenshot of a MySQL Shell terminal window. The title bar shows the path 'C:\Program Files\MySQL\MySQL Shell 8.0\bin\mysqlsh.exe'. The terminal text includes the version 'MySQL Shell 8.0.23', copyright information for Oracle (2018, 2021), and a prompt 'MySQL JS>'.

```
C:\Program Files\MySQL\MySQL Shell 8.0\bin\mysqlsh.exe
MySQL Shell 8.0.23

Copyright (c) 2018, 2021, Oracle and/or its affiliates.
Oracle is a registered trademark of Oracle Corporation and/or its affiliates.
Other names may be trademarks of their respective owners.

Type 'help' or '?' for help; 'quit' to exit.
MySQL JS>
```





공부하느라  
수고했습니다