

Data Understanding (DataBase)

2021. 6. 14 ~ 6. 18

정 준 수 Ph.D

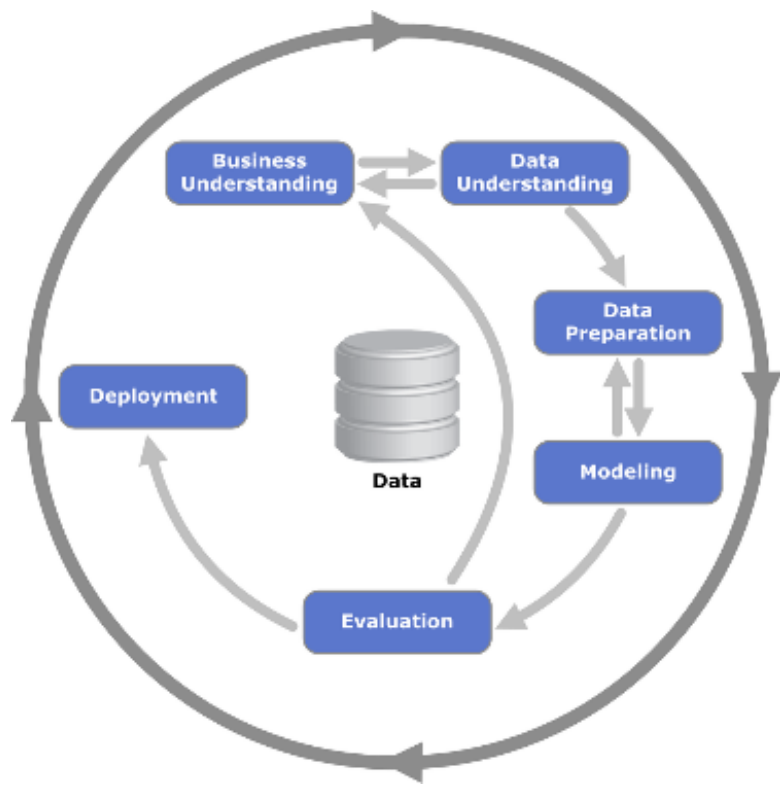
과정 목표

빅 데이터 분석 프로젝트 수행을 위한 데이터 분석 기획, 수집, 분석 및 보고서 작성에 필요한 내용 학습

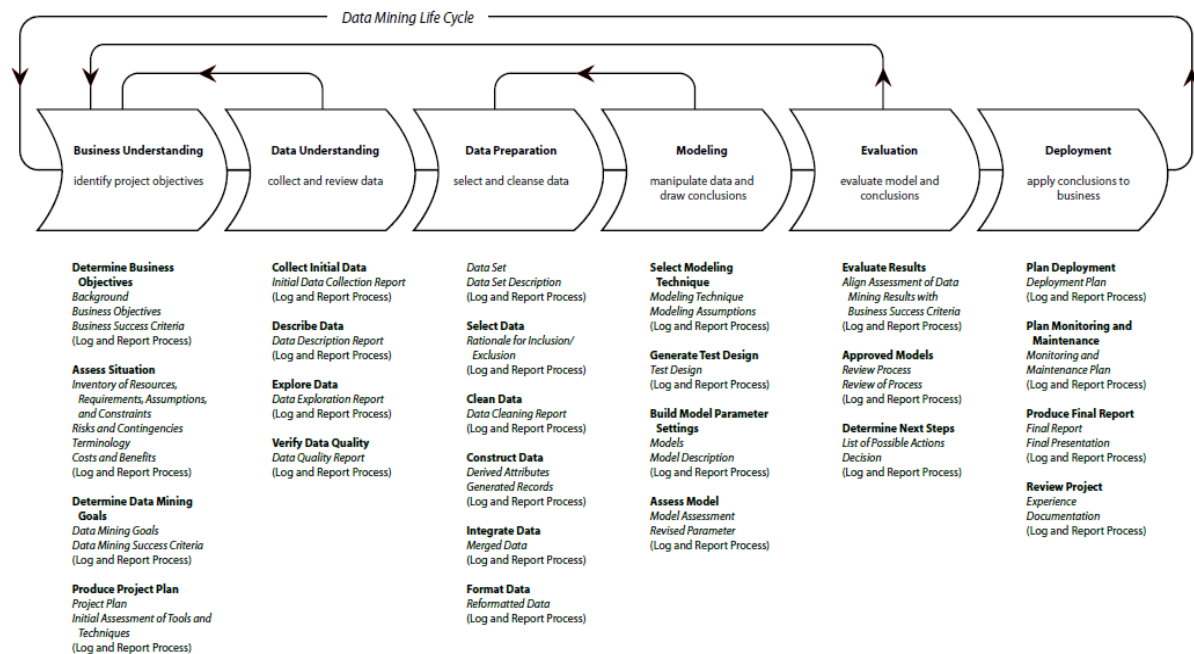
1. Python 활용(객체지향 파이썬 프로그래밍)
2. 데이터 처리 패키지 활용한 통계 분석
3. 풀스택 인터페이스 개발 프로그래밍
4. 웹 데이터 수집 및 다이나믹 시각화
5. 공공 데이터 분석, 시각화 및 발표(Project 수행을 통한 Portfolio 작성)

CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM(Cross Industry Standard Process for Data Mining)은 데이터 마이닝 전문가가 사용하는 일반적인 접근 방식을 설명한 가장 널리 사용되는 공개 표준 분석 모델입니다.



Phases

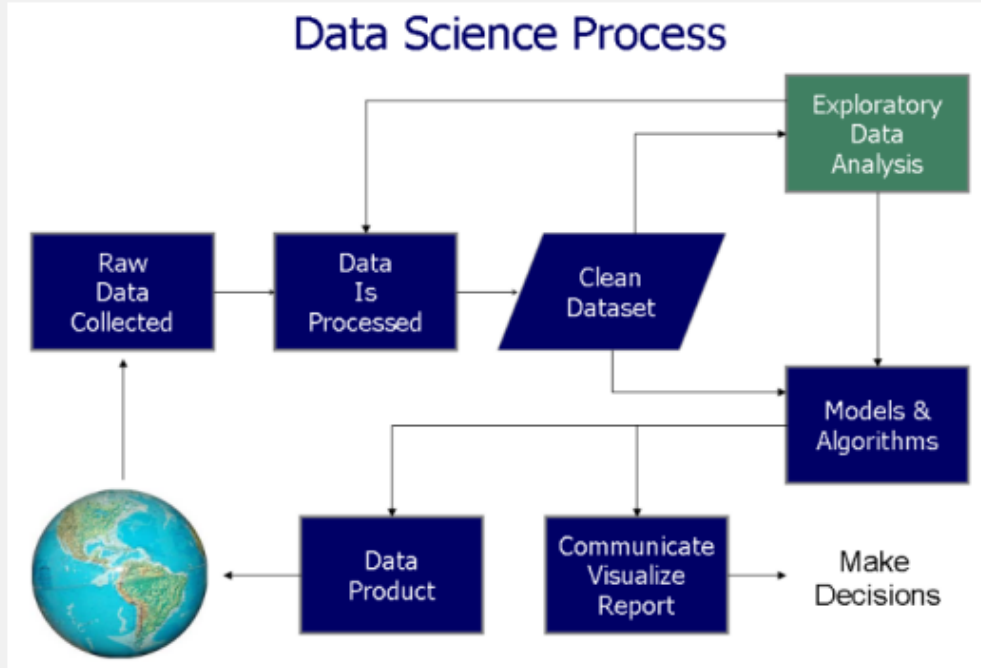


a visual guide to CRISP-DM methodology

SOURCE CRISP-DM 1.0
<http://www.crisp-dm.org/download.htm>
 DESIGN Nicole Leaper
<http://www.nicoleleaper.com>



분석 입문



“The goal of a model is to provide a simple low-dimensional summary of a dataset”

Hadley Wickham

Data Driven 서비스는 시나리오 기반으로 비즈니스 모델의 개입하여 최적화하고 플랫폼 기능을 데이터(추천, 예측, 시뮬레이션 등)로 제공하는 것

인사이트는 가설기반 접근으로 고객의 경험을 숫자로 표현하는 것 (How to : 탐색, 모델링, 시각화)

AARRR

Funnel 중에서 마케팅,
Growth 관련해서 가장 유명한
개념이 AARRR

실리콘 밸리의 스타트업 투자
및 엑셀러레이터인 500
Startups의
Dave McClure라는 사람이
이야기해서 유명해졌음



Funnel Analysis



Ex) MOOC (Massive Open Online Course)

Business Objective : Helping students get jobs.
Financial sustainability

홈페이지를 접속한 **유니크** 숫자가 100,000명
강의 리스트를 접속한 **유니크** 숫자가 80,000명
계정을 만든 **유니크** 숫자가 60,000명
강의를 등록한 **유니크** 숫자가 40,000명
완료한 **유니크** 숫자가 5,000명
직업을 구한 **유니크** 숫자가 1,000명

Unit of diversion

- User Id, Anonymous ID(cookie), Event
- Device Id, IP address

KPI 지표분석

KPI = Key Performance Indicator, 핵심 지표

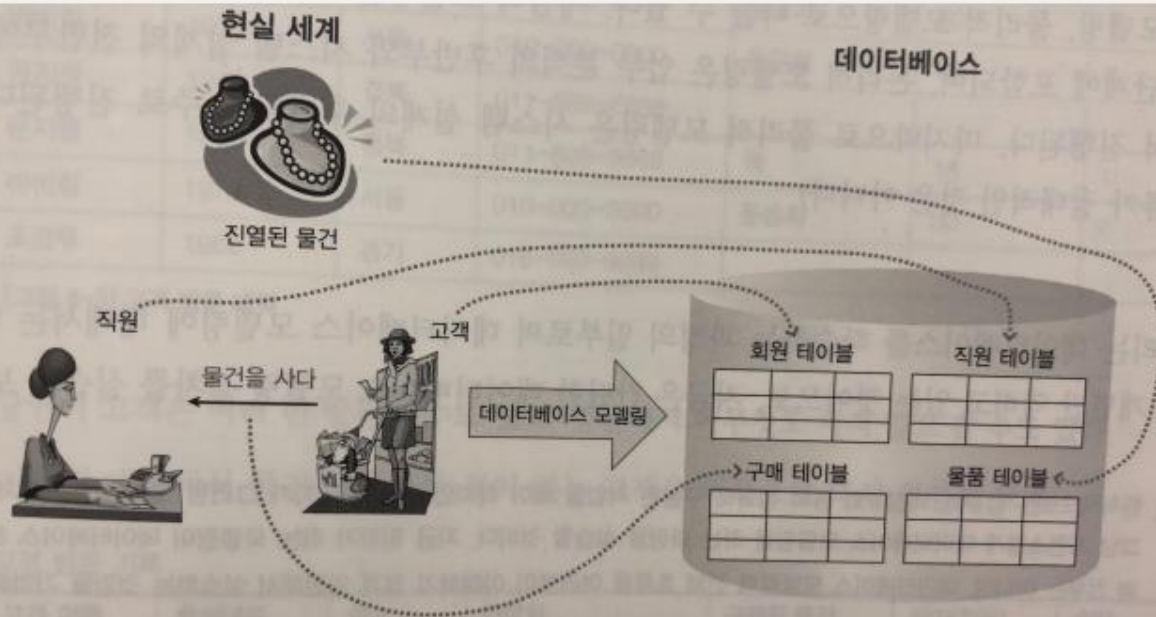
| 분류 | 용어명 | 설명 |
|----|------------|---|
| 유저 | DAU | Daily Active User. 일 방문자수(로그인 기준) |
| | 액티브 유저 | 최근 n일 이내 방문 유저수 |
| | 휴면 유저 | 최근 n일 이내 방문 기록이 없는 유저 |
| | 잔존율 | 특정일의 유저를 100으로 가정했을 때, 특정일 이후 일별로 유입된 유저의 비율 |
| 매출 | 매출액 | 유저가 결제한 금액 |
| | BU | Buying User, 결제 유저 = Paying User (PU)라고도 부름 |
| | BUR(구매전환율) | Buying User Ratio, 전체 유저 중 구매 유저로 전환된 비중 |
| | ARPU | Average Revenue Per User, 전체 유저당 결제 금액 |
| | ARPPU | Average Revenue Per Paying User, 결제 유저당 결제 금액 |

데이터베이스

데이터베이스는 여러 사람에 의해 공유되어 사용할 목적으로 통합하여 관리되는 데이터의 집합, 줄여서 **DB**라고 말함

관계형 데이터베이스 (RDBMS)는 데이터를 로우(Row : 행), 컬럼(Column : 열)과 같이 일종의 표 형태로 저장하고 종속성은 관계(Relation)으로 표현함

컬럼의 구조와 데이터의 관계가 테이블 스키마(Schema)로 사전 정의되고, SQL(구조화 질의 언어)에 의해서 정교한 검색 쿼리를 제공하며 상상하는 거의 모든 방식으로 데이터를 다룰 수 있게 됨



[그림 4-2] 데이터베이스 모델링의 개념

앞의 그림을 보면 데이터베이스 모델링이 어떤 것인지 쉽게 알 수 있다. 그림에 나온 현실

SQL

- **2차원의 정형데이터를 다루는 가장 효율적인 언어.**
(Data를 사용하는 거의 모든 프로그램에서 SQL을 사용하지 않고 효율적인 개발 불가)[DBMS, Hive, Spark, R(sqldf), ...]
- **Data의 처리 과정을 기술하는 방법이 아닌 원하는 Data의 요건을 기술하는 방식**[SQL = Data 요건정의서]
(단순성, 높은 생산성)
- **Data의 요건이 기술된 요건 정의서 형태이므로 요건의 변경시 적은 수정으로 요건을 반영할 수 있음**
(효율적인 유지보수)



SQL

- **DDL(Data Definition Language : 데이터 정의 언어)**
 - CREATE
 - DROP
 - ALTER
- **DML(Data Manipulation Language : 데이터 조작 언어)**
 - INSERT
 - UPDATE
 - DELETE
 - SELECT
- **DCL(Data Control Language : 데이터 제어 언어)**
 - GRANT
 - REVOKE

기본 문법(SELECT)

SELECT ... 선택하고자 하는 속성목록을 기술.

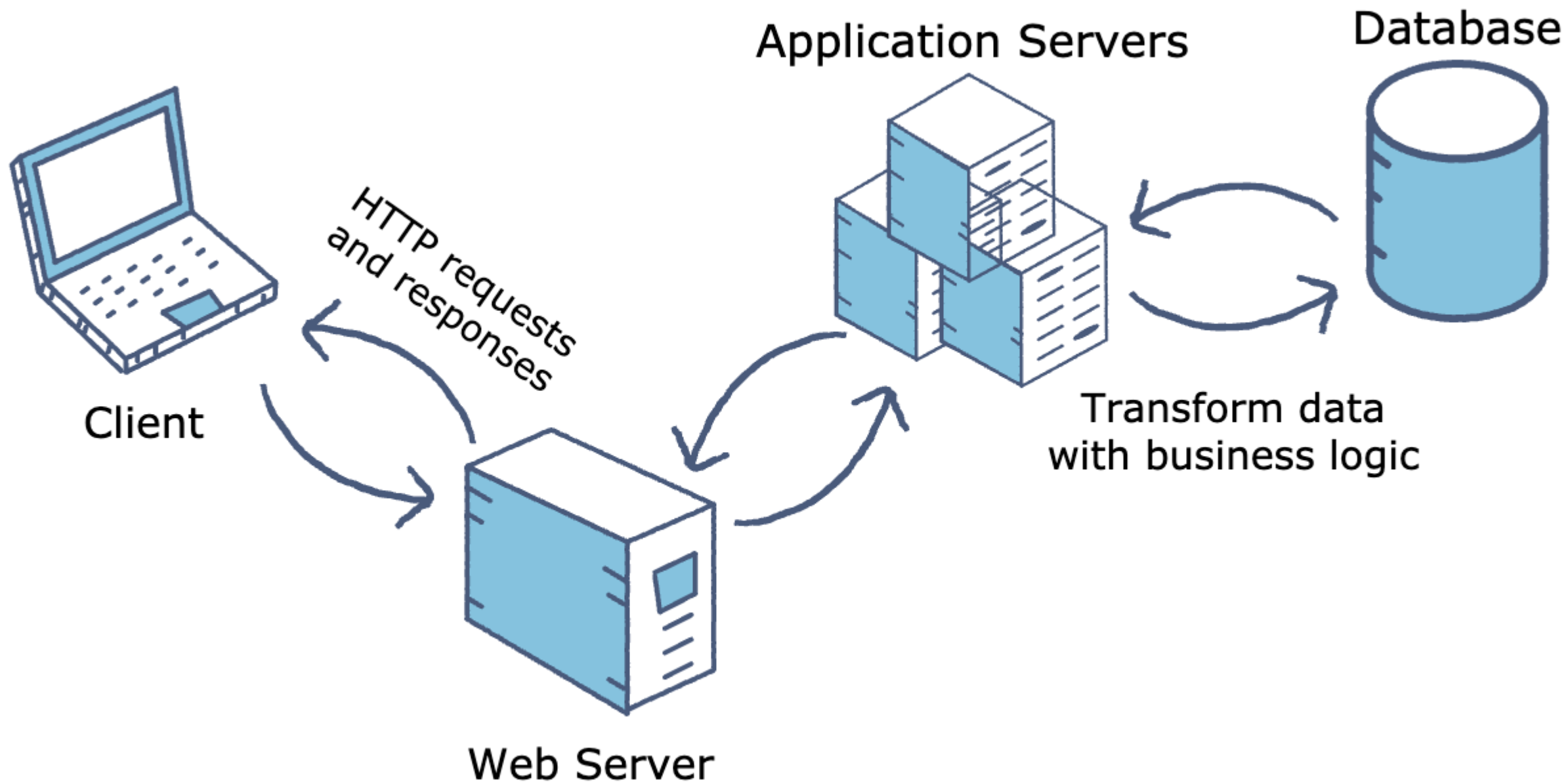
FROM ... Source Data Set을 기술.

[WHERE] ... 선택하고자 하는 행의 조건의 목록을 기술.

[GROUP BY] ... 집계하고자 하는 기준속성목록을 기술.

[HAVING] ... 집계된 행을 기준으로 선택하고자 하는 행의
조건목록을 기술함.(GROUP BY가 선행조건)

[ORDER BY] ... 정렬의 기준이 되는 속성목록과 각 속성 목록의
정렬방식(오림차순, 내림차순)를 기술.



Web Server

- Deliver static content.
- Content is delivered using the HTTP protocol only.
- Serves only web-based applications.
- No support for multi-threading.
- Facilitates web traffic that is not very resource intensive.

Application Server

- Delivers dynamic content.
- Provides business logic to application programs using several protocols (including HTTP).
- Can serve web and enterprise-based applications.
- Uses multi-threading to support multiple requests in parallel.
- Facilitates longer running processes that are very resource-intensive.

SQL vs NoSQL

관계형 데이터 베이스

- 관계형 데이터베이스 관리 시스템(relational database management system, **RDBMS**)
- [IBM](#) 산호세 연구소의 [에드거 F. 커드](#)가 도입한 [관계형 모델](#)을 기반으로 하는 [데이터베이스 관리 시스템](#)임
- 현재 사용되는 대부분의 데이터베이스는 [관계형 데이터베이스](#) 모델을 기반함
- 기업형 DBMS
 - [오라클 데이터베이스](#)
 - [마이크로소프트 SQL 서버](#)
 - [MySQL](#) (오라클)
 - IBM [DB2](#)
 - IBM [Informix](#)
 - [SAP Sybase Adaptive Server Enterprise](#)
 - [SAP Sybase IQ](#)
 - [테라데이터](#)
- TOP 3 [오픈 소스](#) DBMS
 - [MySQL](#)
 - [PostgreSQL](#)
 - [SQLite](#)
- MySQL이 오라클에 인수된 이후 가장 많이 사용되는 MySQL의 [포크](#)는 [MariaDB](#)

NoSQL

- **NoSQL** 데이터베이스는 전통적인 [관계형 데이터베이스](#) 보다 덜 제한적인 [일관성 모델](#)을 이용하는 데이터의 저장 및 검색을 위한 매커니즘을 제공
- NoSQL 데이터베이스는 단순 검색 및 추가 작업을 위한 매우 최적화된 키 값 저장 공간
- NoSQL 데이터베이스는 [빅데이터](#)와 [실시간 웹](#) 애플리케이션의 상업적 이용에 널리 쓰임
- [SQL](#) 계열 쿼리 언어를 사용할 수 있다는 사실을 강조한다는 면에서 "Not only SQL"로 불림
 - 컬럼: [H베이스](#), [아큐물로](#)
 - 문서: [몽고DB](#), [카우치베이스](#)
 - 키 값: [다이나모](#), [리악](#), [레디스](#), [캐시](#), [프로젝트 볼드모트](#)
 - 그래프: [Neo4J](#), [AgensGraph](#), [알레그로그래프](#), [버투오소](#)

General Availability (GA) Releases

Archives



MySQL Installer 8.0.25

Select Operating System:

Microsoft Windows



[Looking for previous GA versions?](#)

Windows (x86, 32-bit), MSI Installer

8.0.25

2.4M

[Download](#)

(mysql-installer-web-community-8.0.25.0.msi)

MD5: 7f64b7ca2c526c359919d9db1bc8d669 | [Signature](#)

Windows (x86, 32-bit), MSI Installer

8.0.25

435.7M

[Download](#)

(mysql-installer-community-8.0.25.0.msi)

MD5: 0c64075a9cc4ec00cce35806761aafc9 | [Signature](#)



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

MySQL Installer

Adding Community

Choosing a Setup Type













Installation

Product Configuration

Installation Complete

Installation

The following products will be installed.

| | Product | Status | Progress | Notes |
|---|---|----------|----------|-------|
| ✓ |  MySQL Server 8.0.25 | Complete | | |
| ✓ |  MySQL Workbench 8.0.25 | Complete | | |
| ✓ |  MySQL for Visual Studio 1.2.9 | Complete | | |
| ✓ |  MySQL Shell 8.0.25 | Complete | | |
| ✓ |  MySQL Router 8.0.25 | Complete | | |
| ✓ |  Connector/ODBC 8.0.25 | Complete | | |
| ✓ |  Connector/C++ 8.0.25 | Complete | | |
| ✓ |  Connector/J 8.0.25 | Complete | | |
| ✓ |  Connector/NET 8.0.25 | Complete | | |
| ✓ |  Connector/Python 8.0.25 | Complete | | |
| ✓ |  MySQL Documentation 8.0.25 | Complete | | |
| ✓ |  Samples and Examples 8.0.25 | Complete | | |

Show Details >

< Back

Next >

Cancel

MySQL Installer

MySQL Server 8.0.25

Type and Networking

Authentication Method

Accounts and Roles

Windows Service

Apply Configuration

Type and Networking

Server Configuration Type

Choose the correct server configuration type for this MySQL Server installation. This setting will define how much system resources are assigned to the MySQL Server instance.

Config Type:

Connectivity

Use the following controls to select how you would like to connect to this server.

☒ TCP/IP Port: X Protocol Port:

☒ Open Windows Firewall ports for network access

☐ Named Pipe Pipe Name:

☐ Shared Memory Memory Name:

Advanced Configuration

Select the check box below to get additional configuration pages where you can set advanced and logging options for this server instance.

☐ Show Advanced and Logging Options

[Next >](#)

[Cancel](#)

MySQL Installer

MySQL Server 8.0.25

Type and Networking

Authentication Method

Accounts and Roles

Windows Service

Apply Configuration

Authentication Method

☒ **Use Strong Password Encryption for Authentication (RECOMMENDED)**

MySQL 8 supports a new authentication based on improved stronger SHA256-based password methods. It is recommended that all new MySQL Server installations use this method going forward.



Attention: This new authentication plugin on the server side requires new versions of connectors and clients which add support for this new 8.0 default authentication (caching_sha2_password authentication).

Currently MySQL 8.0 Connectors and community drivers which use libmysqlclient 8.0 support this new method. If clients and applications cannot be updated to support this new authentication method, the MySQL 8.0 Server can be configured to use the legacy MySQL Authentication Method below.

☐ **Use Legacy Authentication Method (Retain MySQL 5.x Compatibility)**

Using the old MySQL 5.x legacy authentication method should only be considered in the following cases:

- If applications cannot be updated to use MySQL 8 enabled Connectors and drivers.
- For cases where re-compilation of an existing application is not feasible.
- An updated, language specific connector or driver is not yet available.

Security Guidance: When possible, we highly recommend taking needed steps towards upgrading your applications, libraries, and database servers to the new stronger authentication. This new method will significantly improve your security.

< Back

Next >

Cancel

Pwd: 8475

Username :

MySQL[®] Installer

MySQL Server 8.0.25

Type and Networking

Authentication Method

Accounts and Roles

Windows Service

Apply Configuration

Authentication Method

☒ **Use Strong Password Encryption for Authentication (RECOMMENDED)**

MySQL 8 supports a new authentication based on improved stronger SHA256-based password methods. It is recommended that all new MySQL Server installations use this method going forward.



Attention: This new authentication plugin on the server side requires new versions of connectors and clients which add support for this new 8.0 default authentication (caching_sha2_password authentication).

Currently MySQL 8.0 Connectors and community drivers which use libmysqlclient 8.0 support this new method. If clients and applications cannot be updated to support this new authentication method, the MySQL 8.0 Server can be configured to use the legacy MySQL Authentication Method below.

☐ **Use Legacy Authentication Method (Retain MySQL 5.x Compatibility)**

Using the old MySQL 5.x legacy authentication method should only be considered in the following cases:

- If applications cannot be updated to use MySQL 8 enabled Connectors and drivers.
- For cases where re-compilation of an existing application is not feasible.
- An updated, language specific connector or driver is not yet available.

Security Guidance: When possible, we highly recommend taking needed steps towards upgrading your applications, libraries, and database servers to the new stronger authentication. This new method will significantly improve your security.

< Back

Next >

Cancel

Me



All hosts

DB admin

Pwd: 8475

MySQL User Account

Please specify the user name, password, and database role.


 User Name: 

Host:

Role:

Authentication: ☒ MySQL

MySQL user credentials

Password: 

Confirm Password:

Password strength: Max

OK Cancel

MySQL Installer

MySQL Server 8.0.25

Type and Networking

Authentication Method

Accounts and Roles

Windows Service

Apply Configuration

Windows Service

☒ Configure MySQL Server as a Windows Service

Windows Service Details

Please specify a Windows Service name to be used for this MySQL Server instance. A unique name is required for each instance.

Windows Service Name:

☒ Start the MySQL Server at System Startup

Run Windows Service as ...

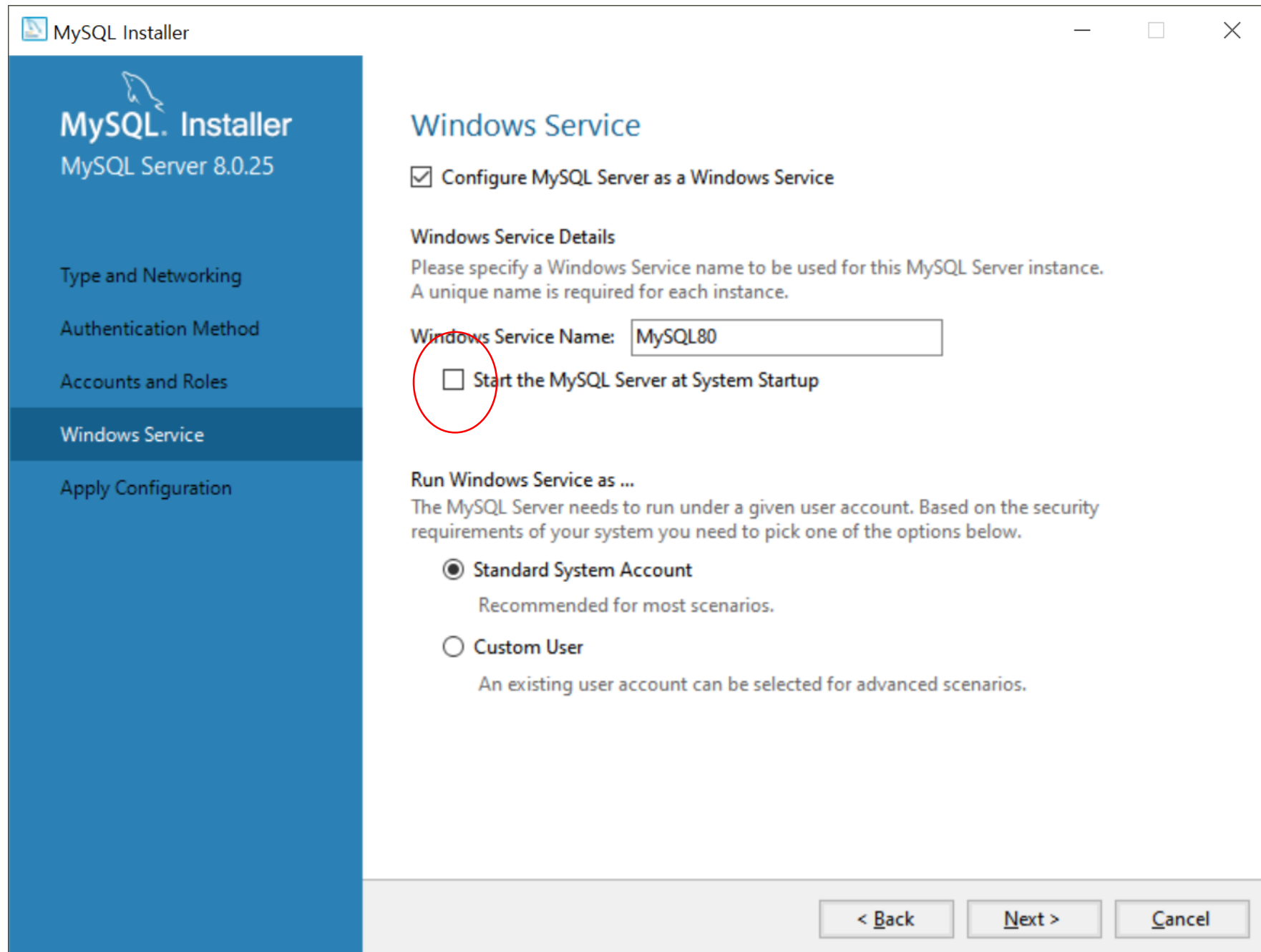
The MySQL Server needs to run under a given user account. Based on the security requirements of your system you need to pick one of the options below.

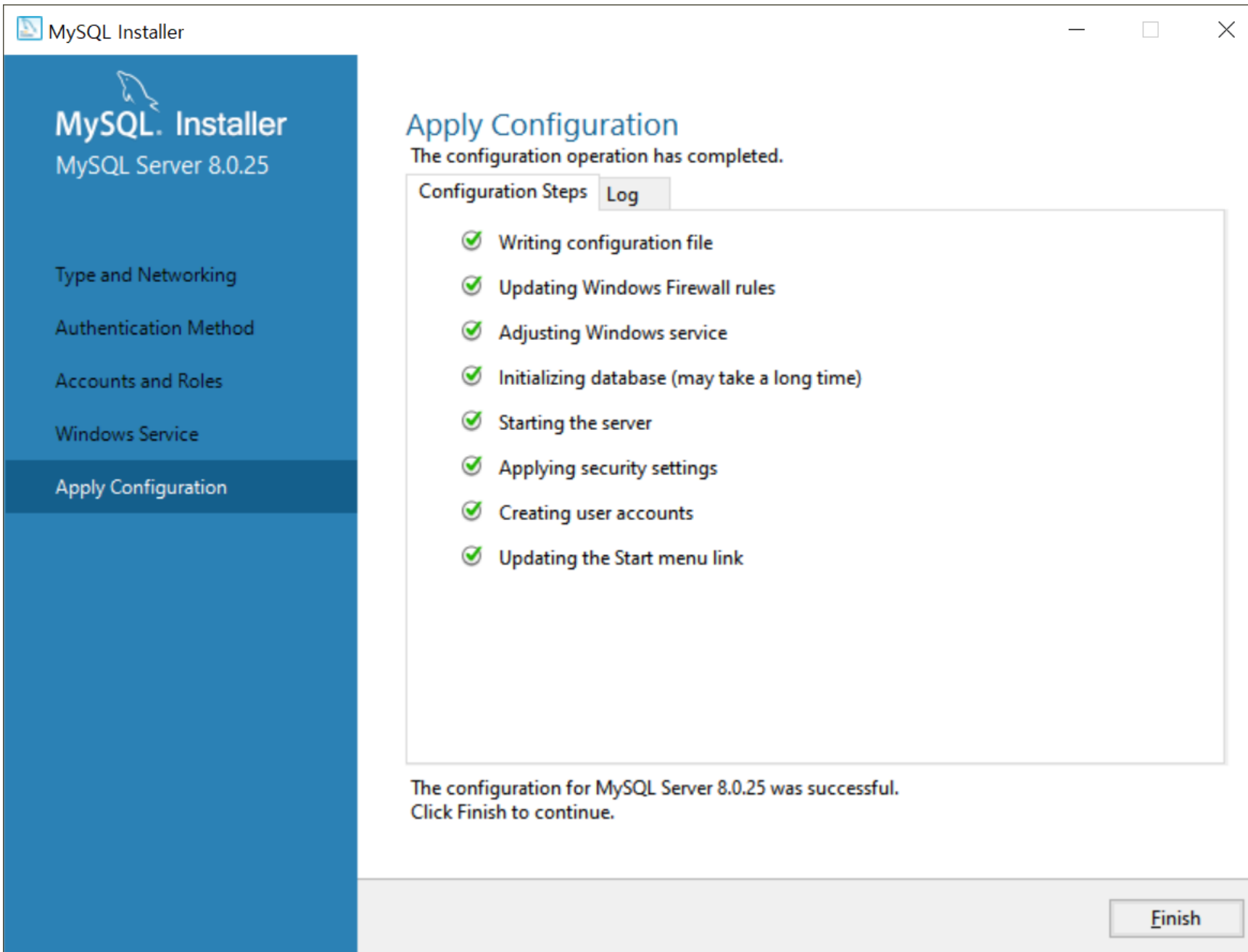
☒ Standard System Account

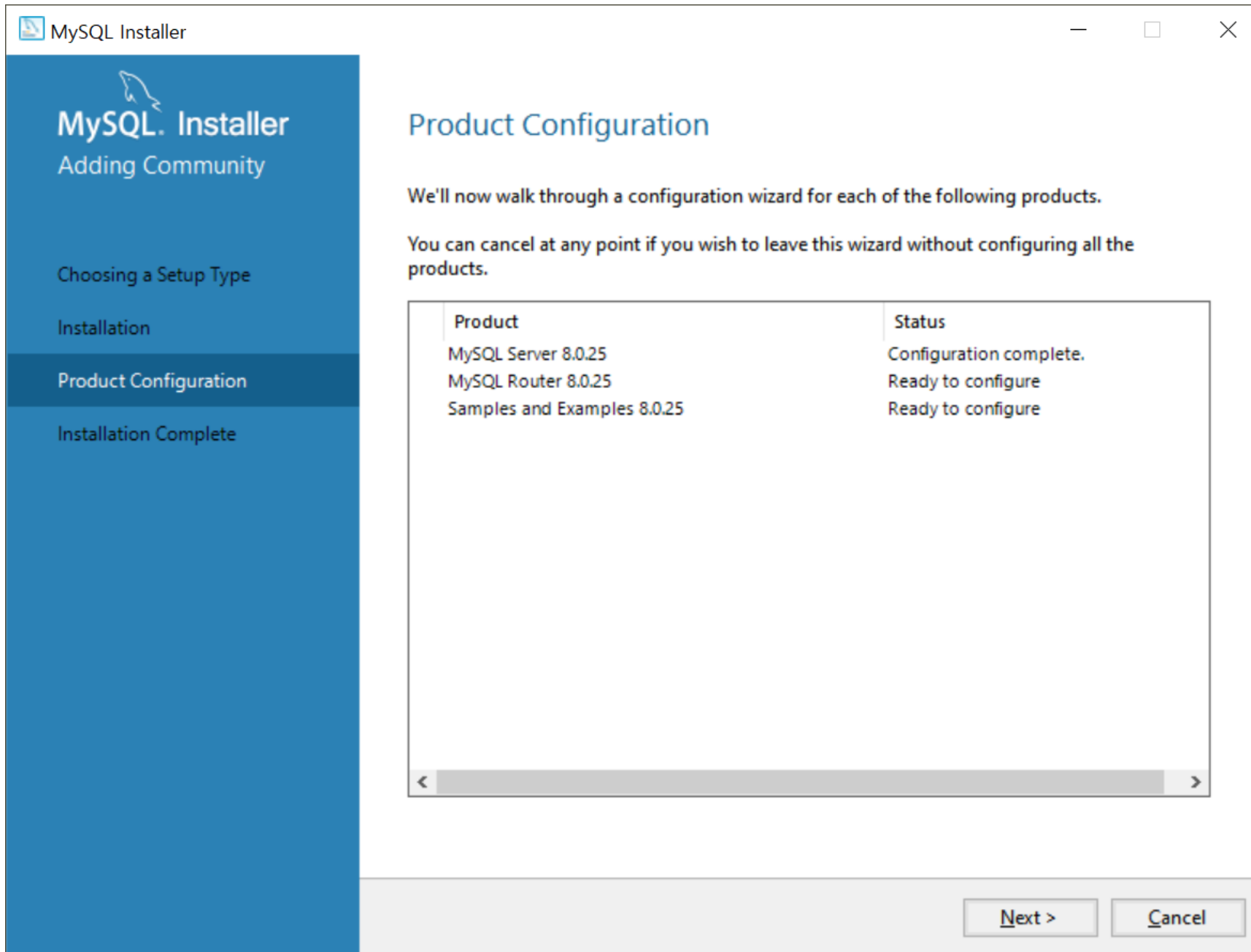
Recommended for most scenarios.

☐ Custom User

An existing user account can be selected for advanced scenarios.







MySQL Installer

MySQL Router 8.0.25

MySQL Router Configuration

MySQL Router Configuration

☐ Bootstrap MySQL Router for use with InnoDB cluster

This wizard can bootstrap MySQL Router to direct traffic between MySQL applications and a MySQL InnoDB cluster. Applications that connect to the router will be automatically directed to an available read/write or read-only member of the cluster.

The bootstrapping process requires a connection to the InnoDB cluster. In order to register the MySQL Router for monitoring, use the current Read/Write instance of the cluster.

Hostname:

Port:

Management User:

Password:

MySQL Router requires specification of a base port (between 80 and 65532). The first port is used for classic read/write connections. The other ports are computed sequentially after the first port. If any port is indicated to be in use, please change the base port.

Classic MySQL protocol connections to InnoDB cluster:

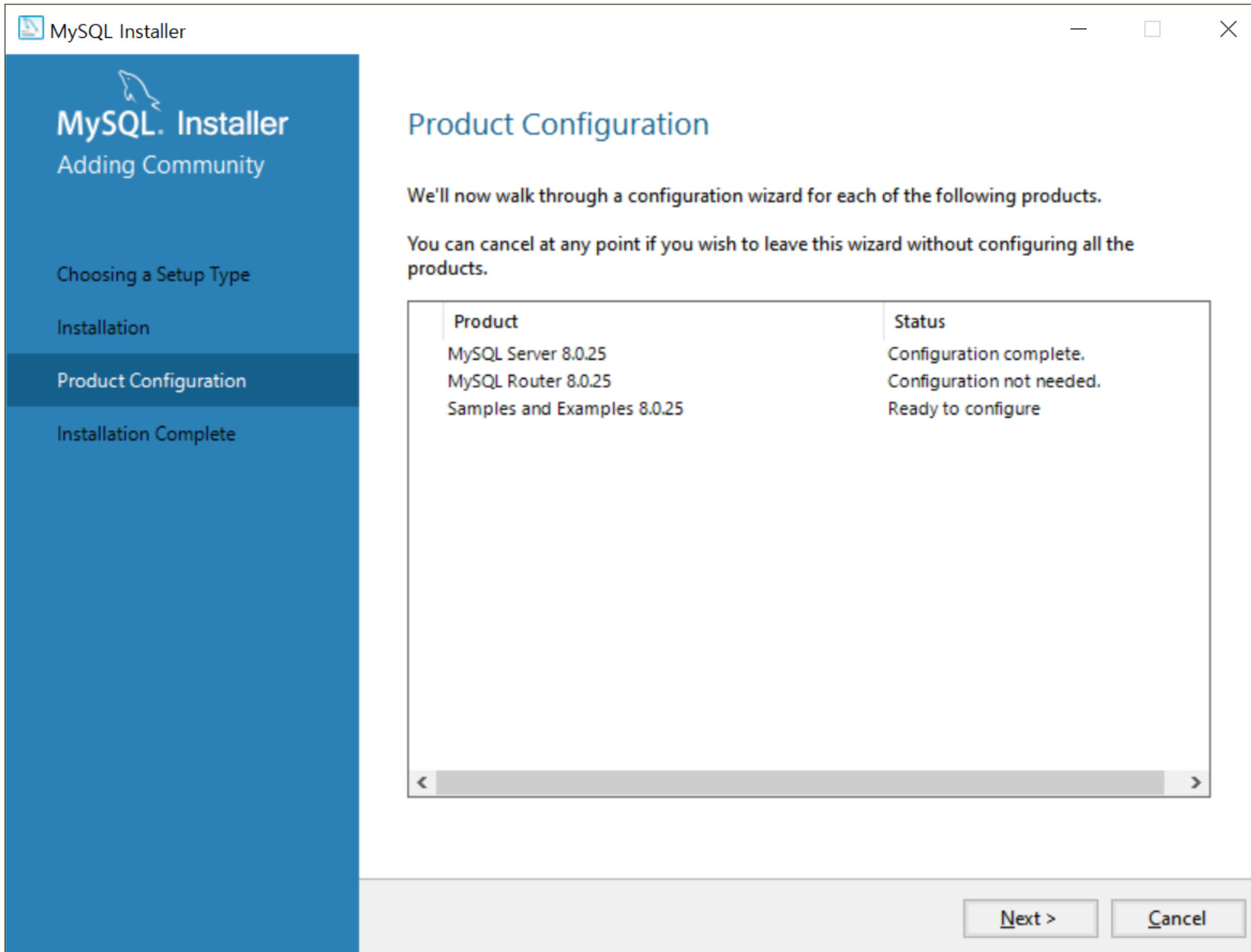
Read/Write:

Read Only:

MySQL X protocol connections to InnoDB cluster:

Read/Write:

Read Only:



MySQL Installer

MySQL. Installer
Samples and Examples

Connect To Server

Apply Configuration

Connect To Server


Select the MySQL server instances from the list to receive sample schemas and data.

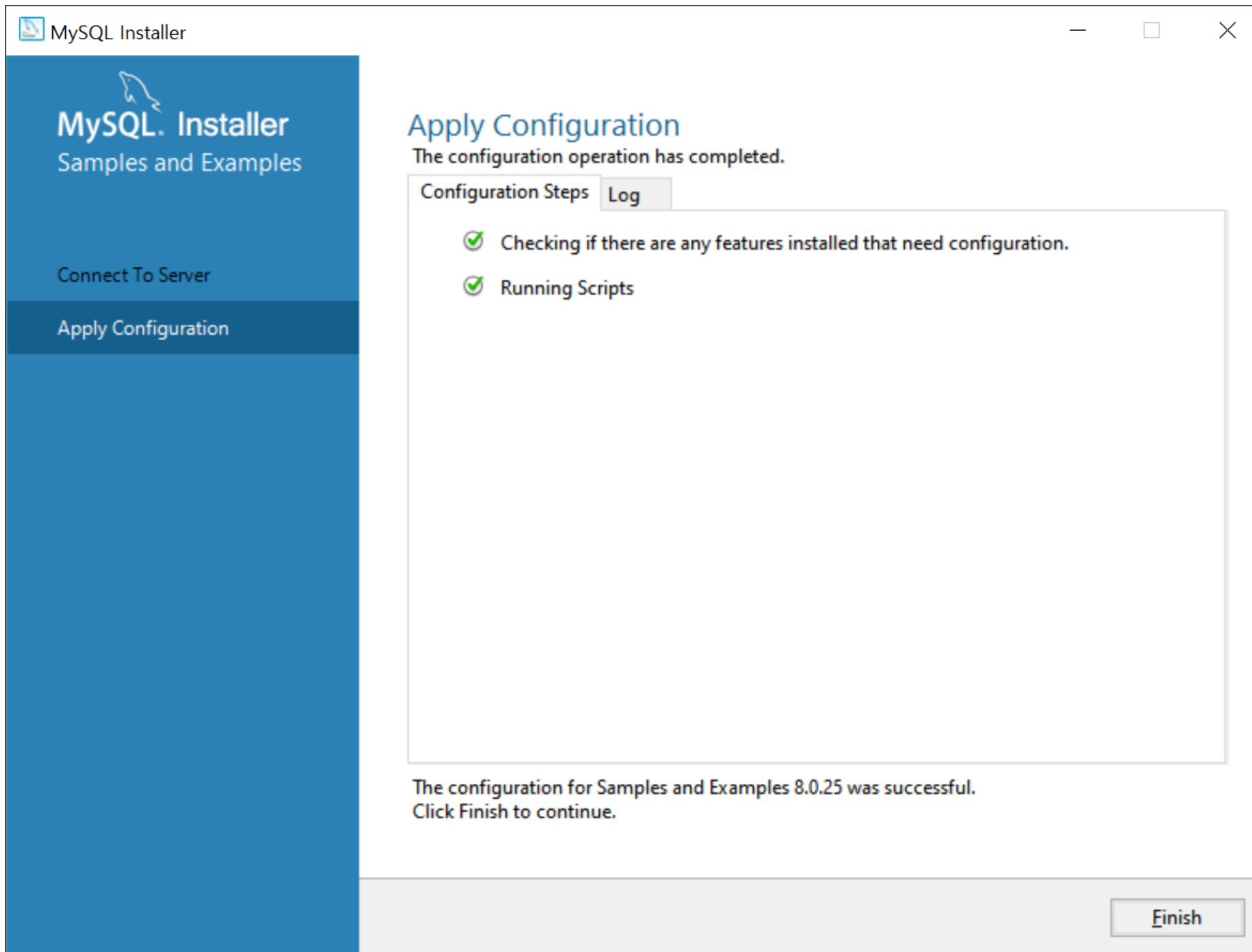
| | Server | Port | Arch... | Type | Status |
|-------------------------------------|---------------------|------|---------|--------------------|-----------------------|
| <input checked="" type="checkbox"/> | MySQL Server 8.0.25 | 3306 | X64 | Stand-alone Server | Connection succeeded. |

Provide the credentials that should be used (requires root privileges).
Click "Check" to ensure they work.

User name: Credentials provided in Server configuration

Password:





MySQL. Installer

Adding Community

Choosing a Setup Type

Installation

Product Configuration

Installation Complete

Installation Complete

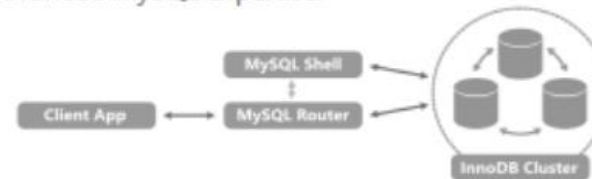
The installation procedure has been completed.

[Copy Log to Clipboard](#)

☒ [Start MySQL Workbench after setup](#)

☒ [Start MySQL Shell after setup](#)

The MySQL Shell is an advanced MySQL client application that can be used to work with single MySQL Server instances. Further, it can be used to create and manage an InnoDB cluster, an integrated solution for high availability and scalability of MySQL databases, without requiring advanced MySQL expertise.



Refer to the following links for documentation, tutorials and examples on MySQL Shell:

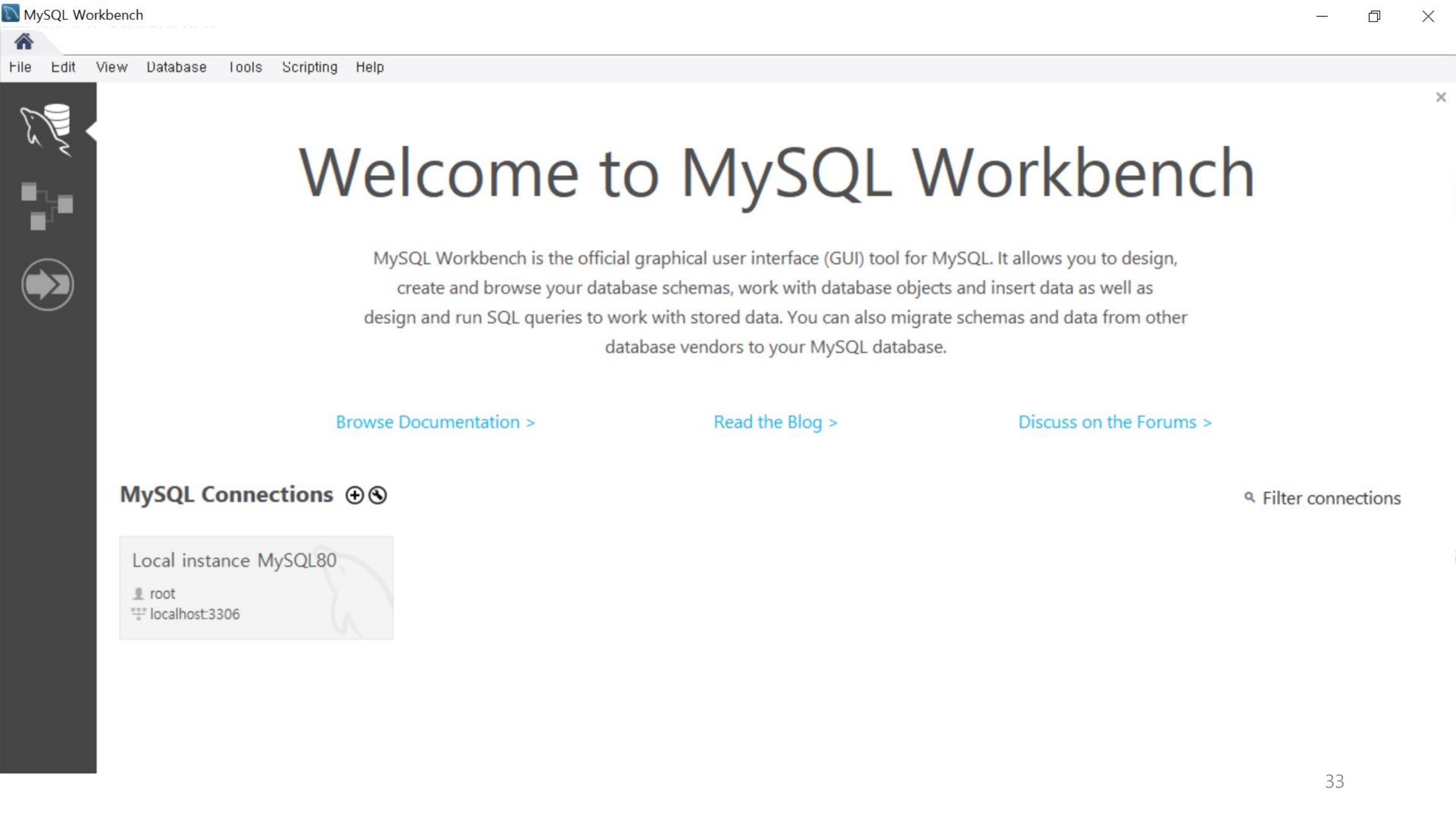
[MySQL Shell Documentation](#)

[Setting up a Real World Cluster Blog](#)

[The All New MySQL InnoDB ReplicaSet Blog](#)

[Changing Cluster Options Live Blog](#)

[Finish](#)



Welcome to MySQL Workbench

MySQL Workbench is the official graphical user interface (GUI) tool for MySQL. It allows you to design, create and browse your database schemas, work with database objects and insert data as well as design and run SQL queries to work with stored data. You can also migrate schemas and data from other database vendors to your MySQL database.

[Browse Documentation >](#)


[Read the Blog >](#)


[Discuss on the Forums >](#)

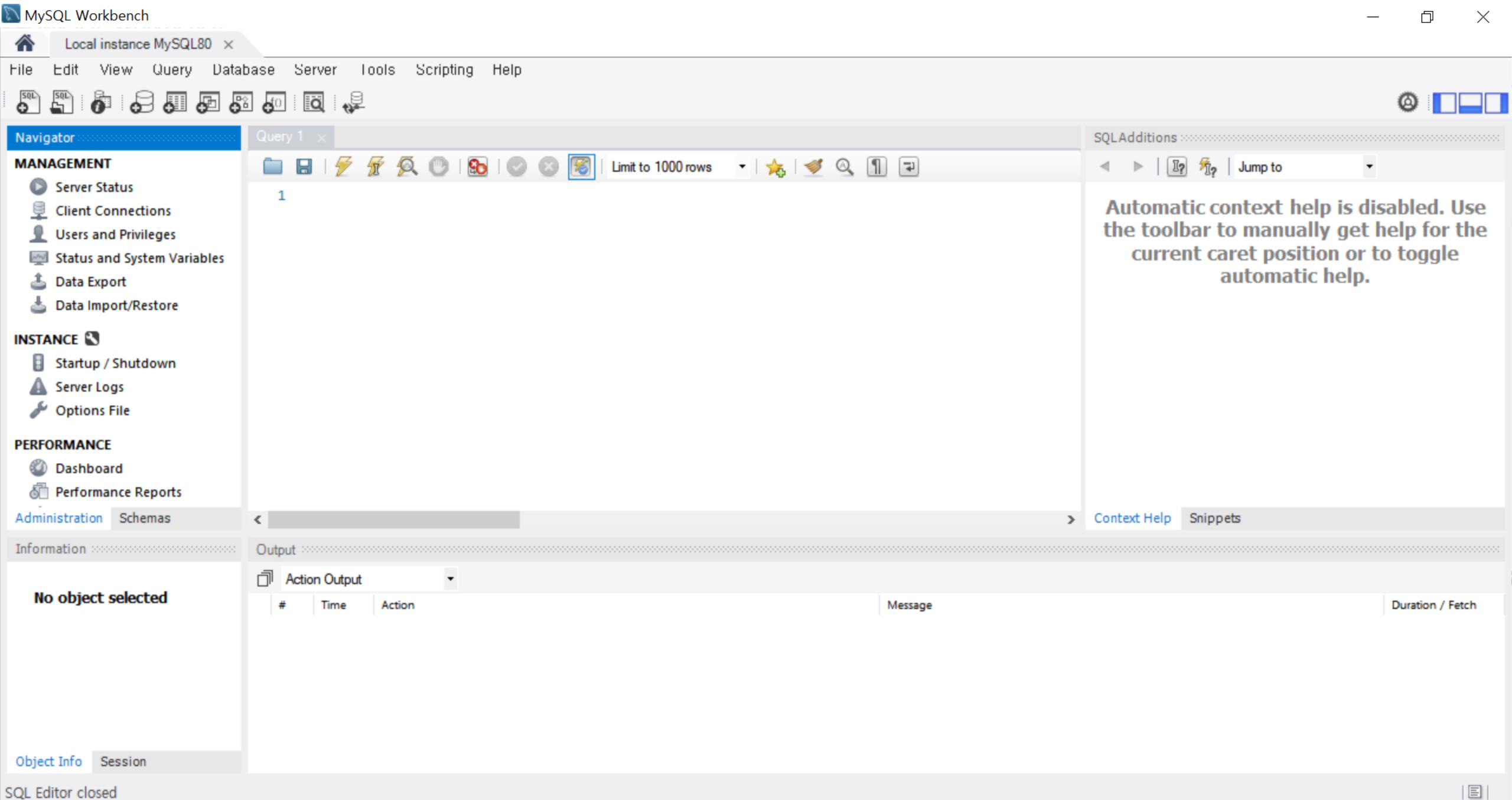
MySQL Connections

 Filter connections

Local instance MySQL80

 root

 localhost:3306



2. 기본 DBMS 사용법

● MySQL 기본 사용법

로그인 후 `sql>` 프롬프트 상에서

| | |
|----------------------------------|------------------------|
| <code>show databases;</code> | DB들의 리스트를 표시하라 |
| <code>use world;</code> | WORLD DB를 사용한다 |
| <code>show tables;</code> | WORLD DB의 테이블리스트를 표시하라 |
| <code>desc city;</code> | city 테이블의 구조를 표시달라 |
| <code>select * from city;</code> | city테이블의 내용을 표시하라 |

학습정리

- 지금까지 'MySQL 설치 및 기본 사용법'에 대해 살펴보았습니다.

MySQL 다운로드

My원도우용 SQL Community Edition 다운로드

MySQL 설치

관리자 패스워드 등록

MySQL 콘솔모드 로그인

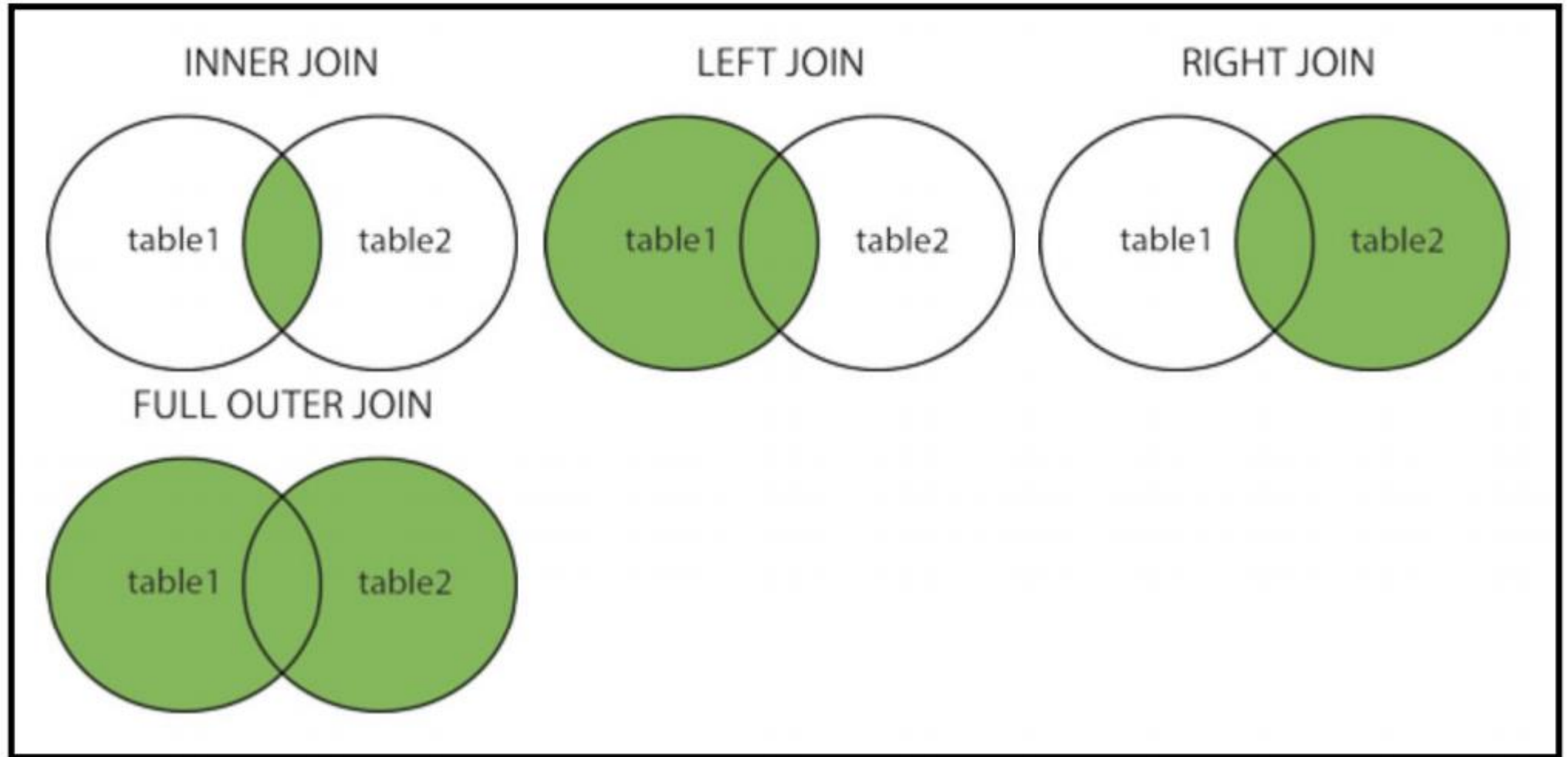
CLI(Command Line Interface) 방식

MySQL Workbench 로그인

GUI(Graphic User Interface) 방식

WORLD DB

여러개의 테이블 이용 - 조인



여러개의 테이블 이용 - 내부조인

```
SELECT column_name(s)  
FROM table1  
INNER JOIN table2  
ON table1.column_name = table2.column_name;
```

여러개의 테이블 이용 - 왼쪽조인

```
SELECT column_name(s)
FROM table1
LEFT JOIN table2
ON table1.column_name = table2.column_name;
```

여러개의 테이블 이용 - 오른쪽조인

```
SELECT column_name(s)
FROM table1
RIGHT JOIN table2
ON table1.column_name = table2.column_name;
```

여러개의 테이블 이용 - 외부조인

```
SELECT column_name(s)
FROM table1
FULL OUTER JOIN table2
ON table1.column_name = table2.column_name
WHERE condition;
```

여러개의 테이블 이용 - 셀프조인

```
SELECT column_name(s)  
FROM table1 T1, table1 T2  
WHERE condition;
```


여러개의 테이블 이용 - 하위질의(서브쿼리)

```
SELECT column1, column2, ...  
FROM (SELECT ...)
```

데이터 분석을 위한 파이썬 연동 SQL 사용 실습

파이썬에서 DB 연동

```
import pymysql.cursors

# Connect to the database
connection = pymysql.connect(host='localhost',
                             user='user',
                             password='passwd',
                             db='db',
                             charset='utf8mb4',
                             cursorclass=pymysql.cursors.DictCursor)

try:
    with connection.cursor() as cursor:
        # Read a single record
        sql = 'SELECT `id`, `password` FROM `users` WHERE `email`=%s'
        cursor.execute(sql, ('webmaster@python.org',))
        result = cursor.fetchone()
        print(result)
finally:
    connection.close()
```

pandas에서 DB 디비연동

```
import pymysql

from sqlalchemy import create_engine

import pandas.io.sql as pSql

# 연결

db_url = 'mysql+pymysql://root:12341234@localhost/python_db'

db_url

# 엔진생성(절차)

engine = create_engine( db_url, encoding='utf8' )

# 실연결




conn = engine.connect()

# 삽입

df_dict.to_sql( name='tbl_trades', con=conn, if_exists='append', index=False )

# 해제

conn.close()
```

| 성명 | 후기 |
|---|---|
|  | <p>이번에 진행을 하며 전체 프로젝트의 주제와 일정 그리고 세부 내용들이 바뀌는 과정에 있어서 힘든 점이 많았지만 그에 맞춰 팀 주제를 바꿔 나감에 따라 배우는 내용도 많았습니다. 시각화를 하는 과정이 매우 재미있었고 그 과정에서 R 을 더 잘 쓸 수 있게 되어 큰 수확이라고 생각합니다. 다음 프로젝트는 더 잘 할 수 있을 거라는 자신감이 생겼으며 다음 프로젝트의 주제는 조금 더 신중하게 선택해야겠다는 다짐을 하게 되었습니다.</p> |
|  | <p>직접 데이터를 가공,처리 하여 눈에 보이는 지표로 시각화 해보니 신기하였고, 좀 더 분석과정을 진행해보면, 실제 행정수요, 민간회사 마케팅 등에 활용 될 수 있는 유의미한 결과들이 나올 수 있을 것 같다. 그리고 계절에 따른 인구 유동이 달라진다는 새로운 인사이트를 도출 해낼 수 있어서 의미 있는 작업이었던 같다.</p> |
|  | <p>프로젝트 일정이 자꾸 변경되서 프로젝트 하기 힘들었고 시각화라는 주제에 얽매이기 보다는 각자 원하는 주제를 선정해서 프로젝트를 진행하는게 더 나을 것 같다는 생각이 들었다. 또 프로젝트 일정이 너무 일러서 배운 것을 적용하기 보다는 기존에 알던 것을 답습할 수 밖에 없어 아쉬웠다.</p> |

강사 소개

정 준 수 / Ph.D (heinem@naver.com)

- 前) 삼성전자 연구원
- 前) 삼성의료원 (삼성생명과학연구소)
- 前) 삼성SDS (정보기술연구소)
- 現) (사)한국인공지능협회, AI, 머신러닝 강의
- 現) 한국소프트웨어산업협회, AI, 머신러닝 강의
- 現) 서울디지털재단, AI 자문위원
- 現) 한성대학교 교수(겸)
- 전문분야: 시각 모델링, 머신러닝(ML), RPA
- <https://github.com/JSJeong-me/>

