

# 9기 자기주도 학습

## Mysql/Maria DB data processing



# Mysql/Maria DB data processing

## 최 호 근 Project consultant

---

- 와이즈넷 분석설계 수석 컨설턴트
- 서울시 민원 데이터분석 자문위원
- SK플래닛 데이터분석 파트장
- 現 국가대표 인도어사이클 체조선수



# 1. 과제 개요



## ☑ 실무 프로젝트시 데이터 이행 ?

- ✓ 프로젝트 오픈시 실제 데이터를 운영DB에 반영하는 작업
- ✓ 대량데이터의 경우 오픈 후에는 증분(신규/수정/삭제)데이터만 반영(스케줄 또는 API 연계등)

## ☑ 데이터 적재 시점

- ✓ 서버 Down 타임에 적재(운영서버가 다른DB서버로 교체 불가인 경우)
- ✓ 사전 적재(신규서버 도입인 경우 미리 적재가 가능)
- ✓ 서비스 오픈 후 적재 (내부오픈 등으로 외부 오픈일이 따로 있는 경우 운영자가 따로 정해진 시간에 적재)

# 1. 과제 개요



## ☑ Mysql 과 MariaDB 사용 이유

- ✓ DB시장에서 사용할 수 있는 최초의 오픈 DB중 하나이며 가장 널리 사용되고 있는 관계형 데이터베이스 관리 시스템
- ✓ 오픈소스 라이선스를 따르기는 하지만 상업적 이용시 상업용 라이선스를 구입해야 함
- ✓ MySQL의 창시자인 몬티 와이드니어가 만든 프로젝트가 MariaDB 두 DB의 성능 차이는 크게 없다
- ✓ MariaDB와 MySQL의 호환성은 매우 높다
- ✓ AWS의 Amazon Aurora와 교체사용하기 용이하다
- ✓ 상용 버전이 별도로 없다는 것이 mariaDB의 장점
- ✓ 상대적으로 많은 레퍼런스들이 있다는 것이 MySQL의 장점

# 1. 과제 개요



## ☑ 정제된 텍스트 데이터 입력시 사용하는 방법을 이용한 실습

- ✓ 텍스트 데이터 이관(다양한 형식의 텍스트파일 CSV,TXT,JSON ...) 및 가공



## 2. 전국 관광안내소 데이터 다운로드



### 표준데이터 상세



#### CSV 전국관광안내소표준데이터

공공데이터 제공 표준데이터 속성정보(허용값, 표현형식/단위 등)는 [공공데이터 제공 표준] 전문을 참고하시기 바랍니다.  
(공공데이터포털>정보공유>자료실) 각 기관에서 등록한 표준데이터를 취합하여 제공하기 때문에 갱신주기는 개별 파일마다  
다릅니다.(기관에서 등록한 데이터를 취합한 것으로 개별 파일별 갱신시점이 다름)



1



0



관심

### 표준데이터 기본정보

[메타데이터 다운로드](#)[오류신고 및 문의](#)

제공범위(대상)	「관광진흥법」 등에 따라 지방자치단체에서 관리하는 관광객에게 관광안내와 홍보 등의 편익을 제공하기 위한 관광안내소에 대한 정보(비상설안내소는 제공범위(대상)에서 제외됨)		
관련법령	관광진흥법, 지방자치단체 관광안내소 설치 및 운영조례 및 관광 진흥 등에 관한 조례 등		
소관기관	문화체육관광부	제공기관	지방자치단체
표준데이터셋제공시스템	해당없음		
갱신주기	연간	수정일	2022-12-11
분류체계	문화체육관광 - 관광	키워드	관광,관광지,안내

<https://www.data.go.kr/data/15013112/standard.do>

## 2. 전국관광안내소 데이터 다운로드

그리드

Open API

추천데이터

※ XLS 이외의 파일은 다운로드시 다소 시간이 걸릴 수 있습니다.

조희조건 선택 ▼

검색어 입력

검색

다운로드

XLS

XML

JSON

RDF

CSV

관광안내소명	안내소위치명
김해수로왕비릉관광안내소	김해시 수로왕비릉 입구 앞
김해관광통역안내소	김해시 수로왕릉 입구 앞
동래구 관광안내소	수안치안센터 1층
완주 송광사 관광안내소	완주 송광사
진해해양공원	진해해양공원 내
충주종합관광안내소	중앙탑
터미널관광안내소	터미널
수안보관광안내소	수안보
낙산관광안내소	강원도 양양군 강현면 동해대로 3094
공항관광안내소	강원도 양양군 손양면 공항로 201 양양국제공항 1층
터미널관광안내소	강원도 양양군 양양읍 양양로 23
평택호관광안내소	평택호관리사무소1층
평택항관광안내소	평택항여객터미널입구
신장소핑물관광안내소	k-55부대앞 신장소핑물입구
예술인광장관광안내소	안정리 예술인광장
김해봉하마을관광안내소	김해 진영 봉하마을 입구(주차장 건너편)
김해가야테마파크안내소	김해가야테마파크 내 (야외무대 건너편)
김해종합관광안내소	김해종합관광안내소 위치: 김해시 진영읍 진영로 150 (김해시민회관 앞)



### 3. 데이터 가공 및 변환

- ✓ 텍스트 데이터 중 입력 대상 데이터 가공  
구분자 확인( csv변환시 구분기호 )



관광안내소명,안내소위치명,시도명,시군구명,안내소소개,부가서비스정보,휴무일,운영시작시각(하절기),운영  
김해수로왕비릉관광안내소,김해시 수로왕비릉 입구 앞,경상남도,김해시,"금관가야의 시조인 수로왕의 비이  
김해관광통역안내소,김해시 수로왕릉 입구 앞,경상남도,김해시,"금관가야의 시조 김수로왕의 능으로 김해  
동래구 관광안내소,수안치안센터 1층,부산광역시,동래구,동래구 관광지 안내,,토+일+동절기(12월~3월)휴무  
완주 송광사 관광안내소,완주 송광사,전라북도,완주군,전라북도 완주군 소양면 대흥리종남산(終南山)에 있  
진해해양공원,진해해양공원 내,경상남도,창원시,"다양한 체험관과 전시관이 있어 해양역사, 해양 생물, 바다  
충주종합관광안내소,중앙탑,충청북도,충주시,"머물수록 매력있는 도시, 충주시 관광안내소입니다.",쉼터+물  
터미널관광안내소,터미널,충청북도,충주시,"머물수록 매력있는 도시, 충주시 관광안내소입니다.",핸드폰 충  
수안보관광안내소,수안보,충청북도,충주시,"머물수록 매력있는 도시, 충주시 관광안내소입니다.",없음,설날-



### 3. 데이터 가공 및 변환

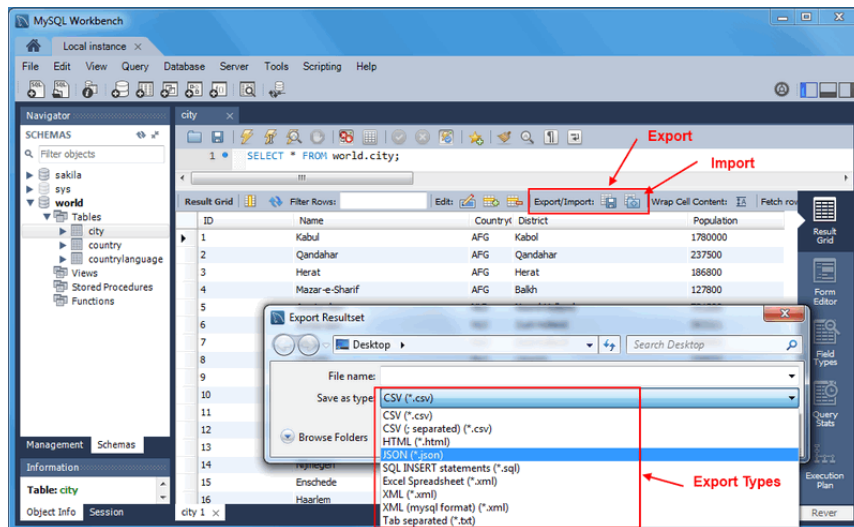
- ✓ 텍스트 속성값 확인  
필드 확인 후 추후에 필요 없는 컬럼 값 제거  
입력을 위한 파일 형식 변환(UTF-8)



관광안내소명	안내소위치명	시도명	시군구명	안내소소기	부가서비스정	휴무일
김해수로왕비릉관광안내소	김해시 수로왕	경상남도	김해시	금관가야의	관광통역(예약	1/1, 명절 전날
김해관광통역안내소	김해시 수로왕	경상남도	김해시	금관가야의	관광통역(영어	1/1, 명절 및 마
동래구 관광안내소	수안치안센터	부산광역시	동래구	동래구 관광지 안내		토+일+동절기
완주 송광사 관광안내소	완주 송광사	전라북도	완주군	전라북도 완주군 소양면		주말,공휴일
진해해양공원	진해해양공원	경상남도	창원시	다양한 체험관과 전시관		매년 1월 1일+
충주종합관광안내소	중앙탑	충청북도	충주시	머물수록 더	심터+물품보관	설날+추석

## 4. 데이터 import

- ✓ 다양한 방법을 통해 데이터 적재
- ✓ Mysql workbench, mysql load data 쿼리등 사용
- ✓ 전국관광안내소 → ssafy\_session테이블 테이블로 import

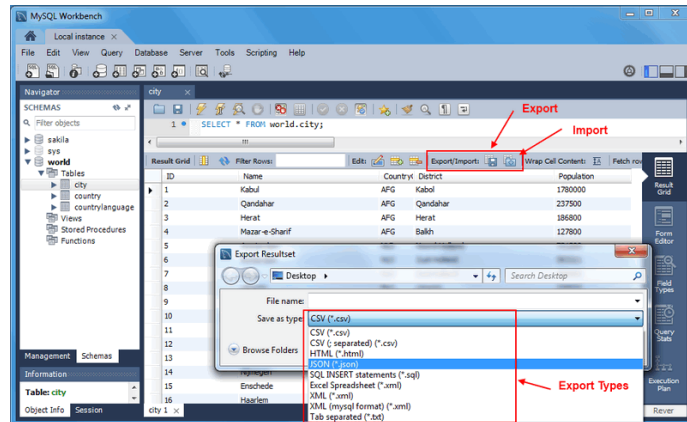


```
mysql> use bharathi;
Database changed
mysql> show tables;
+-----+
| Tables_in_bharathi |
+-----+
| cash_order          |
+-----+
1 row in set (0.00 sec)

mysql> load data local infile 'C:\Python27\cash_order.csv'
-> into table cash_order
-> fields terminated by ','
-> lines terminated by '\n'
-> ;
ERROR 2 (HY000): File 'C:\Python27\cash_order.csv' not found (Errcode: 2 - No
file or directory)
mysql> load data local infile 'C:/Python27/cash_order.csv'
-> into table cash_order
-> fields terminated by ','
-> lines terminated by '\n';
Query OK, 9868890 rows affected, 65535 warnings (23 min 19.78 sec)
Records: 142316285 Deleted: 0 Skipped: 132447395 Warnings: 142316285
```

## 5. 데이터 export

- ✓ 최종데이터 가공 및 내보내기
- ✓ 본인 소속된 지역데이터만 남김(싸피대전의 경우 대전의 관광안내소주소 정보만 테이블에 존재하도록 함)
- ✓ 해당 테이블 데이터 내보내기 (CSV파일, 파일명: ssafy\_session\_ssafy아이디.csv)
- ✓ 테이블 select 결과 화면 스크린샷 저장



```
mysql> use bharathi;
Database changed
mysql> show tables;

+-----+
| Tables_in_bharathi |
+-----+
| cash_order          |
+-----+
1 row in set (0.00 sec)

mysql> load data local infile 'C:\Python27\cash_order.csv'
-> into table cash_order
-> fields terminated by ','
-> lines terminated by '\n'
ERROR 2 (HY000): File 'C:\Python27\cash_order.csv' not found (Errcode: 2 - No
file or directory)
mysql> load data local infile 'C:/Python27/cash_order.csv'
-> into table cash_order
-> fields terminated by ','
-> lines terminated by '\n';
Query OK, 9868890 rows affected, 65535 warnings (23 min 19.78 sec)
Records: 142316285 Deleted: 0 Skipped: 132447395 Warnings: 142316285
```

## 6. error 발생시 원인 확인 및 해결

**Configure Import Settings**

Detected file format: csv

Encoding: utf-8

Columns:

**MySQL Workbench**

**Table Data Import**

Can't analyze file. Please try to change  
If that doesn't help, maybe the file is  
file is empty.

**word: 오류**

**SQL 오류 (1064):** You have an error in your SQL syntax; check the manual that corresponds to your MariaDB server version for the right syntax to use near '...' at line 1

**확인**

**원인을 알아야 같은 실수를 방지**

**오류**

**SQL 오류 (22):** File 'D:\oradmp\영구보관\_oradmp201805041.txt' not found (Errcode: 22 - Invalid argument)

**확인**

**해당 오류에 대한 몇 가지 도움말 찾기**

**Unhandled exception:** quotechar must be set if quoting enabled

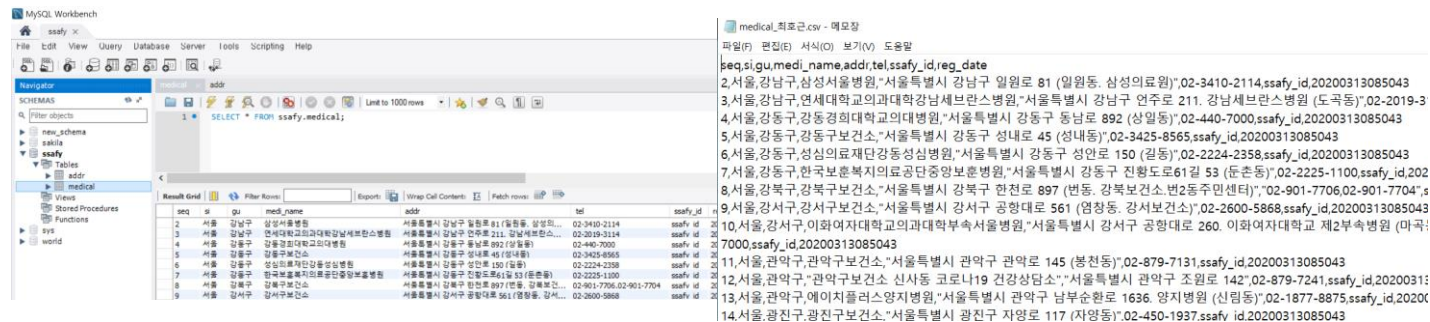
Check the log for more details.

**OK**

< Back Next > Cancel

## 7. 최종산출물

- ✓ 해당데이터 내보내기 (CSV파일, 파일명: ssafy\_session\_ssafy아이디.csv)
- ✓ 데이터 셀렉트 화면 샘플 스크린샷 저장 (select \* from 테이블명 쿼리결과 스크린샷, 본인의 ssafy아이디, 입력날짜 가 나오도록 함)
- ✓ Readme(MD파일)



MySQL Workbench interface showing a query result for medical data. The query is `SELECT * FROM ssafy.medical;`. The result grid displays columns: seq, si, su, med\_name, addr, tel, and ssafy\_id. The data includes various medical facilities and their addresses, such as '삼성서울병원' and '강남세브란스병원'.

seq	si	su	med_name	addr	tel	ssafy_id
2	서울	강남구	삼성서울병원	서울특별시 강남구 삼성동 81 (일원동, 삼성의료원)	02-3410-2114	ssafy_id,20200313085043
3	서울	강남구	연세대학교의과대학강남세브란스병원	서울특별시 강남구 연주로 211, 강남세브란스병원 (도곡동)	02-2019-3114	ssafy_id,20200313085043
4	서울	강남구	강남세브란스병원	서울특별시 강남구 삼성동 81 (일원동)	02-3410-2114	ssafy_id,20200313085043
5	서울	강남구	강남세브란스병원	서울특별시 강남구 삼성동 81 (일원동)	02-3410-2114	ssafy_id,20200313085043
6	서울	강남구	강남세브란스병원	서울특별시 강남구 삼성동 81 (일원동)	02-3410-2114	ssafy_id,20200313085043
7	서울	강남구	강남세브란스병원	서울특별시 강남구 삼성동 81 (일원동)	02-3410-2114	ssafy_id,20200313085043
8	서울	강남구	강남세브란스병원	서울특별시 강남구 삼성동 81 (일원동)	02-3410-2114	ssafy_id,20200313085043
9	서울	강남구	강남세브란스병원	서울특별시 강남구 삼성동 81 (일원동)	02-3410-2114	ssafy_id,20200313085043

README.md

Today I Learned(TIL) 일일과제 에세이

