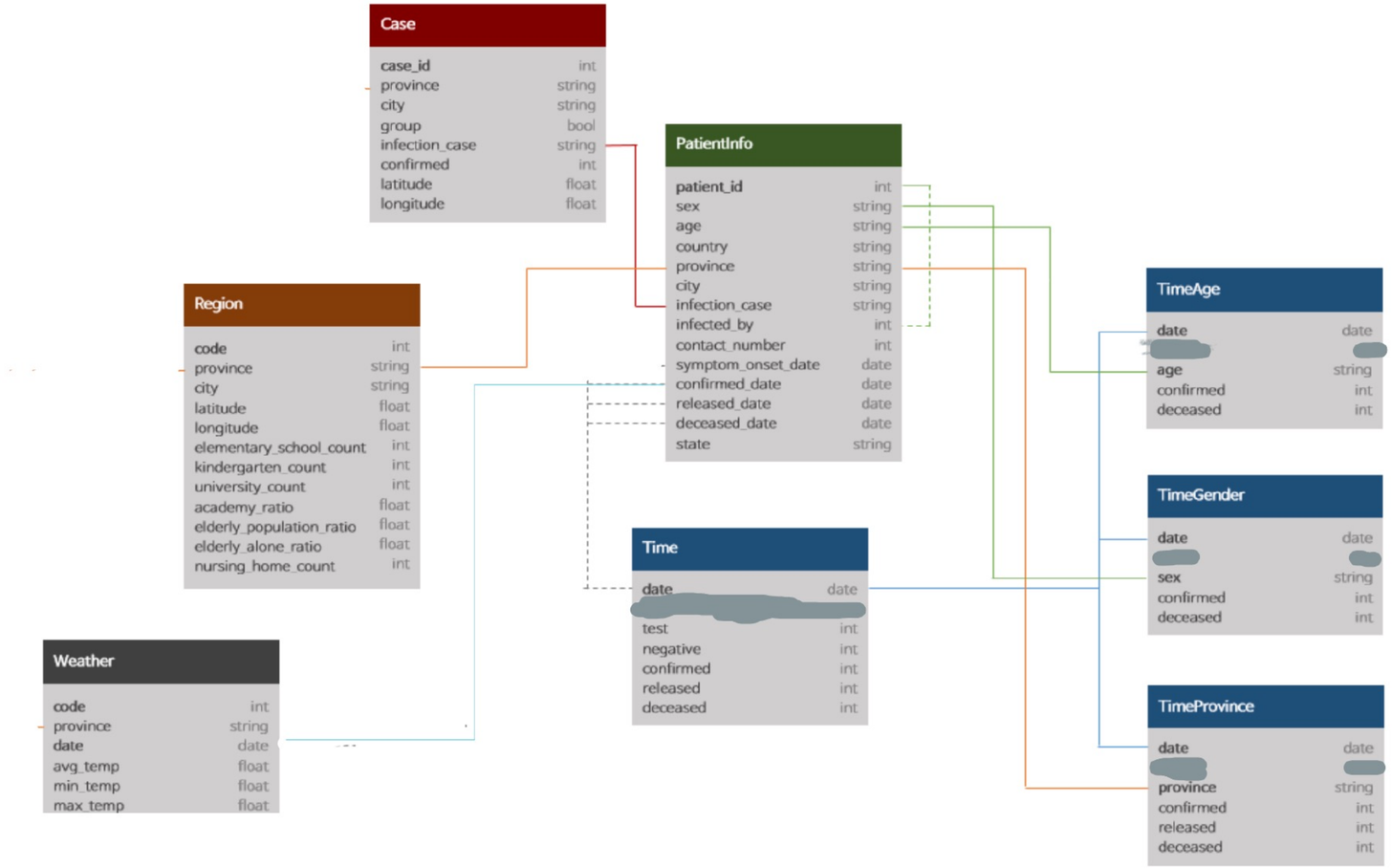
Team Project
KOR – COVID19 Database

Inuk Jung
College of IT Engineering, School of Computer Science and Engineering

Scoring metric

- Total 100 point
  1. 15 point – ER model 완료!
  2. 15 point – Relational model 완료!
  3. 15 point – DB 구축(MySQL, Python)
  4. 15 point – APACHE/PHP연동
  5. 15 point – Search function (필수 기능 2개)
  6. 15 point – SQL tasks
  7. 10 point – Map visualization

팀 프로젝트 2차시


Case
case_id int
province string
city string
group bool
infection_case string
confirmed int
latitude float
longitude float
PatientInfo
patient_id int
sex string
age string
country string
province string
city string
infection_case string
infected_by int
contact_number int
symptom_onset_date date
confirmed_date date
released_date date
deceased_date date
state string
Region
code int
province string
city string
latitude float
longitude float
elementary_school_count int
kindergarten_count int
university_count int
academy_ratio float
elderly_population_ratio float
elderly_alone_ratio float
nursing_home_count int
Time
date date
test int
negative int
confirmed int
released int
deceased int
Weather
code int
province string
date date
avg_temp float
min_temp float
max_temp float
TimeAge
date date
age string
confirmed int
deceased int
TimeGender
date date
sex string
confirmed int
deceased int
TimeProvince
date date
province string
confirmed int
released int
deceased int

팀프로젝트 2차시

- 1차시에서 팀별로 설계 하였던 ER 모델, Relation Model 및 제공된 relation을 바탕으로 팀별로 테이블 생성 후 파이썬으로 insert .
- 팀프로젝트는 데이터베이스는 서버가 아닌 팀별 개인노트북 mysql 사용.
- 총 8개의 테이블 생성 및 insert, 2개의 csv 파일 제공
- 데이터 slicing, parsing시 발생하는 데이터로 인한 에러는 팀별로 자율적으로 처리.

  cf) 어떤 식으로 처리하였는지 README file생성 및 작성 후 같이 제출

제출 목록 : 테이블 생성 sql 파일 하나, 8개의 파이썬 파일(insert), README file

마감 : 다음주 수업 전 자정 까지.

제공 되는 데이터 : K_COVID19.csv → Patinetinfo, Case, Region, Weather

```
patient_id,sex,age,country,province,city,infection_case,infected_by,contact_number,symptom_onset_date,confirmed_date,released_da
1000000001,male,50s,Korea,Seoul,Gangseo-gu,"overseas inflow",NULL,75,2020-01-22,2020-01-23,2020-02-05,NULL,released,4.6,0,9.9,10
1000000002,male,30s,Korea,Seoul,Jungnang-gu,"overseas inflow",NULL,31,NULL,2020-01-30,2020-03-02,NULL,released,5.2,1.4,10.4,1000
1000000003,male,50s,Korea,Seoul,Jongno-gu,"contact with patient",2002000001,17,NULL,2020-01-30,2020-02-19,NULL,released,5.2,1.4,
1000000004,male,20s,Korea,Seoul,Mapo-gu,"overseas inflow",NULL,9,2020-01-26,2020-01-30,2020-02-15,NULL,released,5.2,1.4,10.4,100
1000000005,female,20s,Korea,Seoul,Seongbuk-gu,"contact with patient",1000000002,2,NULL,2020-01-31,2020-02-24,NULL,released,3.9,1.
1000000006,female,50s,Korea,Seoul,Jongno-gu,"contact with patient",1000000003,43,NULL,2020-01-31,2020-02-19,NULL,released,3.9,1.
1000000007,male,20s,Korea,Seoul,Jongno-gu,"contact with patient",1000000003,0,NULL,2020-01-31,2020-02-10,NULL,released,3.9,1.4,8.
1000000008,male,20s,Korea,Seoul,etc,"overseas inflow",NULL,0,NULL,2020-02-02,2020-02-24,NULL,released,1.5,-2.1,5.3,1000036,NULL,
1000000009,male,30s,Korea,Seoul,Songpa-gu,"overseas inflow",NULL,68,NULL,2020-02-05,2020-02-21,NULL,released,-8.3,-11,-4.9,10000
1000000010,female,60s,Korea,Seoul,Seongbuk-gu,"contact with patient",1000000003,6,NULL,2020-02-05,2020-02-29,NULL,released,-8.3,
1000000011,female,50s,China,Seoul,Seodaemun-gu,"overseas inflow",NULL,23,NULL,2020-02-06,2020-02-29,NULL,released,-6.4,-11.8,0.4,
1000000012,male,20s,Korea,Seoul,etc,"overseas inflow",NULL,0,NULL,2020-02-07,2020-02-27,NULL,released,-1.7,-7.2,2.2,1000036,NULL
1000000013,male,80s,Korea,Seoul,Jongno-gu,"contact with patient",1000000017,117,NULL,2020-02-16,NULL,NULL,deceased,-1.4,-4.3,7.7,
1000000014,female,60s,Korea,Seoul,Jongno-gu,"contact with patient",1000000013,27,2020-02-06,2020-02-16,2020-03-12,NULL,released,
1000000015,male,70s,Korea,Seoul,Seongdong-gu,"Seongdong-gu APT",NULL,8,2020-02-11,2020-02-19,NULL,NULL,released,1,-4.4,6.4,10000
1000000016,male,70s,Korea,Seoul,Jongno-gu,"contact with patient",1000000017,NULL,NULL,2020-02-19,2020-03-11,NULL,released,1,-4.4,
1000000017,male,70s,Korea,Seoul,Jongno-gu,"contact with patient",1000000003,NULL,NULL,2020-02-20,2020-03-01,NULL,released,4.6,-0.
1000000018,male,20s,Korea,Seoul,etc,etc,NULL,NULL,NULL,2020-02-20,NULL,NULL,released,4.6,-0.6,10.8,1000038,NULL,0,100,NULL,NULL,
1000000019,female,70s,Korea,Seoul,Jongno-gu,"contact with patient",1000000021,NULL,NULL,2020-02-20,2020-03-08,NULL,released,4.6,
1000000020,female,70s,Korea,Seoul,Seongdong-gu,"Seongdong-gu APT",1000000015,NULL,NULL,2020-02-20,NULL,NULL,released,4.6,-0.6,10.
1000000021,male,80s,Korea,Seoul,Jongno-gu,"contact with patient",1000000016,NULL,NULL,2020-02-20,2020-03-08,NULL,released,4.6,-0.
1000000022,male,30s,Korea,Seoul,Seodaemun-gu,"Eunpyeong St. Mary's Hospital",NULL,NULL,NULL,2020-02-21,NULL,NULL,released,6.7,2.
1000000023,male,50s,Korea,Seoul,Seocho-gu,"Shincheonji Church",NULL,NULL,NULL,2020-02-21,NULL,NULL,released,6.7,2.1,10.9,1000021,
1000000024,male,40s,Korea,Seoul,Guro-gu,"contact with patient",NULL,NULL,NULL,2020-02-22,2020-03-14,NULL,released,4,0,7.9,100003
1000000025,male,60s,Korea,Seoul,Gangdong-gu,"Eunpyeong St. Mary's Hospital",1000000022,NULL,NULL,2020-02-22,NULL,NULL,released,4,
1000000026,male,30s,Korea,Seoul,Seocho-gu,etc,NULL,NULL,2020-02-21,2020-02-22,2020-03-11,NULL,released,4,0,7.9,1000038,NULL,0,100
1000000027,male,50s,Korea,Seoul,Gangseo-gu,"overseas inflow",NULL,NULL,NULL,2020-02-23,2020-03-04,NULL,released,2.5,-2.5,8,10000
1000000028,female,70s,Korea,Seoul,Jongno-gu,"Eunpyeong St. Mary's Hospital",NULL,NULL,NULL,2020-02-23,2020-03-11,NULL,released,2.
1000000029,female,20s,Korea,Seoul,Jongno-gu,"Eunpyeong St. Mary's Hospital",1000000028,NULL,2020-02-11,2020-02-26,2020-03-11,NULL
1000000030,male,60s,China,Seoul,Gangdong-gu,"Eunpyeong St. Mary's Hospital",NULL,NULL,NULL,2020-02-23,NULL,NULL,released,2.5,-2.
```

주의 !

PatientInfo, Case, Region, Weather 테이블의 row의 갯수는 다릅니다!

- 4개의 테이블을 만들기 위한 데이터들이 하나의 csv 파일에 들어 있습니다.
- 해당 csv 파일은 총 33개의 column 으로 구성되어 있습니다. 1행에 각 열이 어떤 속성값인지 명시되어 있습니다.
- 각 row는 환자 한명에 대한 row 입니다.
- 각 row로 부터 각 4개의 테이블에 알맞게 파싱을 하신후 insert를 하시면 됩니다.

Patientinfo 테이블 : Epidemiological data of COVID-19 patients in South Korea

```
mysql> desc patientinfo;
+-------------------+-------------+------+-----+---------+-------+
| Field             | Type        | Null | Key | Default | Extra |
+-------------------+-------------+------+-----+---------+-------+
| patient_id        | bigint      | NO   | PRI | NULL    |       |
| sex               | varchar(10) | YES  |     | NULL    |       |
| age               | varchar(10) | YES  |     | NULL    |       |
| country           | varchar(50) | YES  |     | NULL    |       |
| province          | varchar(50) | YES  |     | NULL    |       |
| city              | varchar(50) | YES  |     | NULL    |       |
| infection_case    | varchar(50) | YES  |     | NULL    |       |
| infected_by       | bigint      | YES  |     | NULL    |       |
| contact_number    | int         | YES  |     | NULL    |       |
| symptom_onset_date | date       | YES  |     | NULL    |       |
| confirmed_date    | date        | YES  |     | NULL    |       |
| released_date     | date        | YES  |     | NULL    |       |
| deceased_date     | date        | YES  |     | NULL    |       |
| state             | varchar(20) | YES  |     | NULL    |       |
+-------------------+-------------+------+-----+---------+-------+
```

- Patient_id : region_code(5) + patient_number(5)
- Provice : 서울, 부산 같은 특별시 및 광역시 또는 경기도 강원도 와 같은 도
- City :
  1) province가 서울 부산 같은 특별시, 광역시인 경우 City는 강남구, 서초구, 해운대구
  2) province가 경상북도 경기도 같은 경우에는 City가 구미시, 안동시
- Infection_case : 감염 원인
  ex) overseas inflow, contact with patient, Eunpyeong St. Mary's Hospital
- Infected_by : the ID of who infected this patient
  cf) this column refers to the 'patient_id' column.
- Contact_number : 접촉한 사람들 수
- Symptom_onset_date : 증상발생 날짜
- Confirmed_date : 확진(양성 판정) 일
- Released_date : 완치(퇴원)날짜
- Deceased_date : 사망일
- State : isolated / released / deceased

Case 테이블 : Data of COVID-19 infection cases in South Korea

```
mysql> desc caseINFo;
+-----------------+------------+------+-----+---------+-------+
| Field           | Type       | Null | Key | Default | Extra |
+-----------------+------------+------+-----+---------+-------+
| case_id         | int        | NO   | PRI | NULL    |       |
| province        | varchar(50)| YES  |     | NULL    |       |
| city            | varchar(50)| YES  |     | NULL    |       |
| infection_group | tinyint(1) | YES  |     | NULL    |       |
| infection_case  | varchar(50)| YES  |     | NULL    |       |
| confirmed       | int        | YES  |     | NULL    |       |
| latitude        | float      | YES  |     | NULL    |       |
| longitude       | float      | YES  |     | NULL    |       |
+-----------------+------------+------+-----+---------+-------+
```

- Case_id : The ID of the infection case
  case_id(7) = region_code(5)+case_number(2)
- Infection_group : 집단감염 여부
  TRUE = Group infection
  FALSE =not group
- infection_case : the infection case (the name of group or other cases)
  ex) Itaewon Clubs, Guro-gu Call Center
- Confirmed : 확진자 수

Region 테이블 : Location and statistical data of the regions in South Korea

Field	Type	Null	Key	Default	Extra
region_code	int	NO	PRI	NULL	
province	varchar(50)	YES		NULL	
city	varchar(50)	YES		NULL	
latitude	float	YES		NULL	
longitude	float	YES		NULL	
elementary_school_count	int	YES		NULL	
kindergarten_count	int	YES		NULL	
university_count	int	YES		NULL	
academy_ratio	float	YES		NULL	
elderly_population_ratio	float	YES		NULL	
elderly_alone_ratio	float	YES		NULL	
nursing_home_count	int	YES		NULL	

Weather 테이블 : Data of the weather in the regions of South Korea

Field	Type	Null	Key	Default	Extra
region_code	int	NO	PRI	NULL	
province	varchar(50)	YES		NULL	
wdate	date	NO	PRI	NULL	
avg_temp	float	YES		NULL	
min_temp	float	YES		NULL	
max_temp	float	YES		NULL	

- Region_code: the code of the region
- Wdate = Date

PatientInfo, Case, region, weather **테이블 생성 및** data parsing

K_COVID19.csv파일에서 테이블당 유효한 attribute만 뽑아내어 insert!

- Parsing_patient.py
- Pasring_case.py
- Parsing_region.py
- Parsing_weather.py

힌트1)
case 테이블 약 : 120 여개의 row
region 테이블 약 : 170 여개의 row
weather 테이블 약 : 2500여개의 row

힌트2)
지난 과제처럼 file을 open후 for문을 통해 한 줄씩 읽어가며 처리해도 되지만
Pandas 라이브러리의 Dataframe을 사용하는 것도 가능

K_COVID19.csv, additional_Timeinfo.csv

```
patient_id,sex,age,country,province,city,infection_case,infected_by,contact_number,symptom_onset_date,confirmed_date,released_da
1000000001,male,50s,Korea,Seoul,Gangseo-gu,"overseas inflow",NULL,75,2020-01-22,2020-01-23,2020-02-05,NULL,released,4.6,0,9.9,10
1000000002,male,30s,Korea,Seoul,Jungnang-gu,"overseas inflow",NULL,31,NULL,2020-01-30,2020-03-02,NULL,released,5.2,1.4,10.4,1000
1000000003,male,50s,Korea,Seoul,Jongno-gu,"contact with patient",2002000001,17,NULL,2020-01-30,2020-02-19,NULL,released,5.2,1.4,
1000000004,male,20s,Korea,Seoul,Mapo-gu,"overseas inflow",NULL,9,2020-01-26,2020-01-30,2020-02-15,NULL,released,5.2,1.4,10.4,100
1000000005,female,20s,Korea,Seoul,Seongbuk-gu,"contact with patient",1000000002,2,NULL,2020-01-31,2020-02-24,NULL,released,3.9,1.
1000000006,female,50s,Korea,Seoul,Jongno-gu,"contact with patient",1000000003,43,NULL,2020-01-31,2020-02-19,NULL,released,3.9,1.
1000000007,male,20s,Korea,Seoul,Jongno-gu,"contact with patient",1000000003,0,NULL,2020-01-31,2020-02-10,NULL,released,3.9,1.4,8.
1000000008,male,20s,Korea,Seoul,etc,"overseas inflow",NULL,0,NULL,2020-02-02,2020-02-24,NULL,released,1.5,-2.1,5.3,1000036,NULL,
1000000009,male,30s,Korea,Seoul,Songpa-gu,"overseas inflow",NULL,68,NULL,2020-02-05,2020-02-21,NULL,released,-8.3,-11,-4.9,10000
1000000010,female,60s,Korea,Seoul,Seongbuk-gu,"contact with patient",1000000003,6,NULL,2020-02-05,2020-02-29,NULL,released,-8.3,-
1000000011,female,50s,China,Seoul,Seodaemun-gu,"overseas inflow",NULL,23,NULL,2020-02-06,2020-02-29,NULL,released,-6.4,-11.8,0.4,
1000000012,male,20s,Korea,Seoul,etc,"overseas inflow",NULL,0,NULL,2020-02-07,2020-02-27,NULL,released,-1.7,-7.2,2.2,1000036,NULL
1000000013,male,80s,Korea,Seoul,Jongno-gu,"contact with patient",1000000017,117,NULL,2020-02-16,NULL,NULL,deceased,-1.4,-4.3,7.7,
1000000014,female,60s,Korea,Seoul,Jongno-gu,"contact with patient",1000000013,27,2020-02-06,2020-02-16,2020-03-12,NULL,released,
1000000015,male,70s,Korea,Seoul,Seongdong-gu,"Seongdong-gu APT",NULL,8,2020-02-11,2020-02-19,NULL,NULL,released,1,-4.4,6.4,10000
1000000016,male,70s,Korea,Seoul,Jongno-gu,"contact with patient",1000000017,NULL,NULL,2020-02-19,2020-03-11,NULL,released,1,-4.4,
1000000017,male,70s,Korea,Seoul,Jongno-gu,"contact with patient",1000000003,NULL,NULL,2020-02-20,2020-03-01,NULL,released,4.6,-0.
1000000018,male,20s,Korea,Seoul,etc,etc,NULL,NULL,NULL,2020-02-20,NULL,NULL,released,4.6,-0.6,10.8,1000038,NULL,0,100,NULL,NULL,
1000000019,female,70s,Korea,Seoul,Jongno-gu,"contact with patient",1000000021,NULL,NULL,2020-02-20,2020-03-08,NULL,released,4.6,-
1000000020,female,70s,Korea,Seoul,Seongdong-gu,"Seongdong-gu APT",1000000015,NULL,NULL,2020-02-20,NULL,NULL,released,4.6,-0.6,10.
1000000021,male,80s,Korea,Seoul,Jongno-gu,"contact with patient",1000000016,NULL,NULL,2020-02-20,2020-03-08,NULL,released,4.6,-0.
1000000022,male,30s,Korea,Seoul,Seodaemun-gu,"Eunpyeong St. Mary's Hospital",NULL,NULL,NULL,2020-02-21,NULL,NULL,released,6.7,2.
1000000023,male,50s,Korea,Seoul,Seocho-gu,"Shincheonji Church",NULL,NULL,NULL,2020-02-21,NULL,NULL,released,6.7,2.1,10.9,1000021,
1000000024,male,40s,Korea,Seoul,Guro-gu,"contact with patient",NULL,NULL,NULL,2020-02-22,2020-03-14,NULL,released,4,0,7.9,100003
1000000025,male,60s,Korea,Seoul,Gangdong-gu,"Eunpyeong St. Mary's Hospital",1000000022,NULL,NULL,2020-02-22,NULL,NULL,released,4,
1000000026,male,30s,Korea,Seoul,Seocho-gu,etc,NULL,NULL,2020-02-21,2020-02-22,2020-03-11,NULL,released,4,0,7.9,1000038,NULL,0,10
1000000027,male,50s,Korea,Seoul,Gangseo-gu,"overseas inflow",NULL,NULL,NULL,2020-02-23,2020-03-04,NULL,released,2.5,-2.5,8,10000
1000000028,female,70s,Korea,Seoul,Jongno-gu,"Eunpyeong St. Mary's Hospital",NULL,NULL,NULL,2020-02-23,2020-03-11,NULL,released,2.
1000000029,female,20s,Korea,Seoul,Jongno-gu,"Eunpyeong St. Mary's Hospital",1000000028,NULL,2020-02-11,2020-02-26,2020-03-11,NULL
1000000030,male,60s,China,Seoul,Gangdong-gu,"Eunpyeong St. Mary's Hospital",NULL,NULL,NULL,2020-02-23,NULL,NULL,released,2.5,-2.
```

- Additional_Timeinfo.csv 와 K_COVID19.csv의 PatientInfo를 함께 사용
- Additional_Time.csv는 총 3개의 column으로 구성되어 있습니다. 1행에 각각의 column 정보를 확인하실 수 있습니다.
- 각 row는 해당날짜의 누적 검사수, 누적 음성판정수로 이루어져 있습니다.
  (ex : 2020년 1월 20일에는 검사자 1명, 음성판정 0명)

keTime > addtional_Timeinfo.csv

```
date,test,negative
2020-01-20,1,0
2020-01-21,1,0
2020-01-22,4,3
2020-01-23,22,21
2020-01-24,27,25
2020-01-25,27,25
2020-01-26,51,47
2020-01-27,61,56
2020-01-28,116,97
2020-01-29,187,155
2020-01-30,246,199
2020-01-31,312,245
2020-02-01,371,289
2020-02-02,429,327
2020-02-03,490,414
2020-02-04,607,462
2020-02-05,714,522
2020-02-06,885,693
2020-02-07,1352,1001
2020-02-08,2097,1134
2020-02-09,2598,1683
2020-02-10,3110,2552
2020-02-11,4325,3535
2020-02-12,5624,4811
2020-02-13,6511,5921
2020-02-14,7242,6679
2020-02-15,7734,7148
2020-02-16,8161,7647
2020-02-17,8718,7980
2020-02-18,9772,8923
2020-02-19,11173,9973
2020-02-20,13202,11238
2020-02-21,16400,13016
2020-02-22,21586,15116
2020-02-23,26179,17520
2020-02-24,32756,20292
2020-02-25,40304,25447
2020-02-26,53553,31576
2020-02-27,66652,39318
2020-02-28,81167,48593
2020-02-29,94055,55723
2020-03-01,98921,61825
2020-03-02,109591,71580
2020-03-03,125851,85484
2020-03-04,136707,102965
2020-03-05,146541,118965
2020-03-06,164740,136624
2020-03-07,178189,151802
2020-03-08,188518,162008
```

TimeInfo 테이블 : COVID-19 data by date

```
mysql> desc TimeInfo;
+-----------+---------+------+-----+---------+-------+
| Field     | Type    | Null | Key | Default | Extra |
+-----------+---------+------+-----+---------+-------+
| date      | date    | NO   | PRI | NULL    |       |
| test      | int(11) | YES  |     | NULL    |       |
| negative  | int(11) | YES  |     | NULL    |       |
| confirmed | int(11) | YES  |     | NULL    |       |
| released  | int(11) | YES  |     | NULL    |       |
| deceased  | int(11) | YES  |     | NULL    |       |
+-----------+---------+------+-----+---------+-------+
```

- Date : 코로나 터진 이후 2020-06-30일까지의 날짜(primary key)
- Test : 그 날의 누적검사자 수
- Negative : 그 날의 누적 음성판정자 수
- Confirmed : 그날 누적 양성판정자 수
- Released : 확진받고 격리된 사람들 중 격리 해제된 사람들 수(누적)
- deceased : 누적사망자 수

주의1 : 모든 column은 "누적 " 됨.

Ex) 1월 20일에 확진자 1명 , 1월 21일에 확진자 1명 : 총 2명

=> 1월 21일의 confirmed은 "2" 여야함.

주의2 : Time의 date와 Patient의 Confirmed_date와 연결하셔야 합니다.

TIMEINFO **테이블 생성 및** data parsing

```
mysql> desc TimeInfo;
+-----------+---------+------+-----+---------+-------+
| Field     | Type    | Null | Key | Default | Extra |
+-----------+---------+------+-----+---------+-------+
| date      | date    | NO   | PRI | NULL    |       |
| test      | int(11) | YES  |     | NULL    |       |
| negative  | int(11) | YES  |     | NULL    |       |
| confirmed | int(11) | YES  |     | NULL    |       |
| released  | int(11) | YES  |     | NULL    |       |
| deceased  | int(11) | YES  |     | NULL    |       |
+-----------+---------+------+-----+---------+-------+
```

1)Date, test, negative 데이터는 additional_time.csv에서 가져올것

2)해당하는 date의 confirmed, released, deceased는 K_COVID19.csv의 patientinfo 에서 누적 사람수를 count해서 입력 할 것

주의! date까지의 누적 released, confirmed, deceased 환자 수를 구할 것!

TimeAge, TimeGender, TimeProvince **테이블 생성 및** data parsing

Hint1 : 각각 의 table column의 type(string,int ...)은 patientInfo(1주차실습자료) 와 TimeInfo의 column type정보를 사용하여 만들어 주십시오.

Hint2 : Time에서는 date만 primary key였습니다. 하지만 TimeAge는 (날짜,나이대)를 기준으로 복합키를 설정하셔야 할 것입니다.

Hint3 : TimeProvince의 장소는 patientinfo의 patient province를 기준으로 합니다. Case의 province(x)

Hint4 : Released_date, confirmed date, deceased date는 K_COVID19.csv의 PatientInfo에서 가져와서 사용

DO YOUR BEST!