

L. Lancia

G. Salillari

Cloud Computing

Master Degree in Data Science

Sapienza Università di Roma

# Facebook Tao

Distributed Data Store for the Social  
Graph

## Indice

Introduction

Open Data

Schema

Query

## Introduction

- Aim of this project was to build a simple relational database and executing some queries.
- We used open data about public transport network of city of Rome provided by ATAC. <http://www.agenziamobilita.roma.it/it/progetti/open-data/dataset.html>
- The DBMS used is MySQL<sup>1</sup>

---

<sup>1</sup>Ver 14.14 Distrib 5.7.11

## Open Data

ATAC provides the data in the GTFS format.

### What is GTFS?

The General Transit Feed Specification (GTFS) defines a common format for public transportation schedules and associated geographic information.<sup>2</sup>

Unfortunately not all tables required in GTFS standard are available in the ATAC dataset.

---

<sup>2</sup><https://developers.google.com/transit/gtfs/#how-do-i-start>

## File Used

<code>stops.txt</code>	Individual locations where vehicles pick up or drop off passengers.
<code>routes.txt</code>	Transit routes. A route is a group of trips that are displayed to riders as a single service.
<code>trips.txt</code>	Trips for each route. A trip is a sequence of two or more stops that occurs at specific time.
<code>stop_times.txt</code>	Times that a vehicle arrives at and departs from individual stops for each trip.
<code>calendar_dates.txt</code>	Exceptions for the service IDs defined in the <code>calendar.txt</code> file. If <code>calendar_dates.txt</code> includes ALL dates of service, this file may be specified instead of <code>calendar.txt</code> .

Table: Data Set files

## Table routes

```
CREATE TABLE `routes` (  
  `route_id` varchar(5) DEFAULT NULL,  
  `agency_id` varchar(8) DEFAULT NULL,  
  `route_short_name` varchar(5) DEFAULT NULL,  
  `route_long_name` varchar(10) DEFAULT NULL,  
  `route_type` int(1) DEFAULT NULL,  
  `route_color` varchar(6) DEFAULT NULL,  
  `route_text_color` int(6) DEFAULT NULL,  
  KEY `route_id` (`route_id`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

## Table stops

```
CREATE TABLE `stops` (  
  `stop_id` varchar(5) NOT NULL DEFAULT '',  
  `stop_name` varchar(42) NOT NULL DEFAULT '',  
  `stop_lat` decimal(26,10) DEFAULT NULL,  
  `stop_lon` decimal(26,10) DEFAULT NULL,  
  `location_type` int(1) DEFAULT NULL,  
  `parent_station` varchar(5) DEFAULT NULL,  
  PRIMARY KEY (`stop_id`,`stop_name`),  
  KEY `stop_id` (`stop_id`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

## Table trips

```
CREATE TABLE `trips` (  
  `route_id` varchar(5) DEFAULT NULL,  
  `service_id` varchar(9) DEFAULT NULL,  
  `trip_id` varchar(11) DEFAULT NULL,  
  `direction_id` int(1) DEFAULT NULL,  
  `shape_id` bigint(10) DEFAULT NULL,  
  KEY `trip_id` (`trip_id`),  
  KEY `route_id` (`route_id`),  
  CONSTRAINT `trips_ibfk_1` FOREIGN KEY (`route_id`)  
    REFERENCES `routes` (`route_id`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```



## Table times

```
CREATE TABLE `times` (  
  `trip_id` varchar(11) DEFAULT NULL,  
  `arrival_time` time DEFAULT NULL,  
  `departure_time` time DEFAULT NULL,  
  `stop_id` varchar(5) DEFAULT NULL,  
  `stop_sequence` int(11) DEFAULT NULL,  
  KEY `trip_id` (`trip_id`),  
  KEY `stop_id` (`stop_id`),  
  CONSTRAINT `times_ibfk_1` FOREIGN KEY (`trip_id`)  
    REFERENCES `trips` (`trip_id`),  
  CONSTRAINT `times_ibfk_2` FOREIGN KEY (`stop_id`)  
    REFERENCES `stops` (`stop_id`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

## Table calendar

```
CREATE TABLE `calendar` (  
  `service_id` varchar(9) DEFAULT NULL,  
  `date` date DEFAULT NULL,  
  `exception_type` int(1) DEFAULT NULL  
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

## Query

1) List all routes that stops at "Policlinico"

```
SELECT DISTINCT route_id
FROM times join stops join
      trips
ON times.`stop_id`=stops.`
   stop_id`
AND trips.`trip_id`=times.`
   trip_id`
WHERE stops.`stop_name` = "
      POLICLINICO"
```

results: N2L N13 490 495 61 N2  
649 2 19 3 N10 88 N11 MEB  
MEB1

2) List all routes that a stop at a stop containing the word "DE LOLLIS"

results:

```
SELECT DISTINCT route_id,  
                stop_name  
FROM times JOIN stops JOIN  
        trips  
ON times.`stop_id`=stops.`  
    stop_id`  
AND trips.`trip_id`=times.`  
    trip_id`  
WHERE stops.`stop_name` LIKE  
        "%DE LOLLIS%"
```

```
C3 DE LOLLIS- IRPINI  
N10 DE LOLLIS- IRPINI  
492 DE LOLLIS- IRPINI  
C3 DE LOLLIS- VERANO  
N10 DE LOLLIS- VERANO  
C2 DE LOLLIS- VERANO  
492 DE LOLLIS- VERANO  
2 VERANO- DE LOLLIS  
19 VERANO- DE LOLLIS  
3 VERANO- DE LOLLIS  
71 VERANO- DE LOLLIS
```

### 3) List the stops of the autobus 445

```
SELECT DISTINCT stop_name
FROM stops, times, trips
WHERE stops.stop_id = times.
      stop_id
AND trips.trip_id = times.
      trip_id
AND `route_id` = "445"
```

```
BOLOGNA
VENTUNO APRILE- VILLA RICOTTI
VENTUNO APRILE- NARDINI
LANCIANI- BOLDETTI
LANCIANI- DE PETRA
MONTI TIBURTINI- NOMENTANA
MONTI DI PIETRALATA
CURIONI- DE LORENZO
CURIONI- COLLINA LANCIANI
CURIONI
CURIONI- PENTA
CURIONI- REPOSSI
LARGO LANCIANI
LANCIANI- WINCKELMANN
VENTUNO APRILE- RICOTTI
CARACI- MINISTERO INFRASTRUTTURE T
```

#### 4) How many trains of MEB runs in a day?

<pre>SELECT count(trip_id),        calendar.date FROM trips join calendar ON trips.service_id =     calendar.service_id WHERE route_id = "MEB" GROUP BY calendar.date LIMIT 7</pre>	<pre>count(trip_id) date 375 2016-03-14 375 2016-03-15 375 2016-03-16 375 2016-03-17 400 2016-03-18 267 2016-03-19 241 2016-03-20</pre>
---	---

## 5) Which line departs from "VERANO" after 5pm

```
SELECT route_id,  
       departure_time  
FROM trips JOIN times JOIN  
       stops  
ON times.stop_id = stops.  
   stop_id  
AND times.trip_id = trips.  
   trip_id  
WHERE stop_name = "VERANO"  
AND departure_time = (  
SELECT MIN(`departure_time`)  
FROM times JOIN stops  
ON times.stop_id = stops.  
   stop_id  
WHERE stop_name = "VERANO"  
AND `departure_time` > "  
   17:00:00")
```

route_id	departure_time
3	17:01:00
71	17:01:00
163	17:01:00
545	17:01:00
542	17:01:00

6) Which are the three most frequented stops on sunday?

```
SELECT  stop_name,  count(  
        times.trip_id) AS cnt  
FROM    times JOIN stops JOIN  
        calendar JOIN trips  
ON      times.stop_id = stops.  
        stop_id  
AND     calendar.service_id =  
        trips.service_id  
AND     trips.trip_id = times.  
        trip_id  
WHERE   date = "2016-03-20"  
GROUP BY stop_name ORDER BY  
        cnt DESC LIMIT 3
```

```
TERMINI 4158  
PIAZZA VENEZIA 2770  
CONCA D'ORO 2493
```



## 7) List all of stops served by nocturne bus

```
SELECT distinct stop_name,  
               route_id  
FROM trips, times, stops  
WHERE trips.trip_id = times.  
       trip_id  
AND stops.stop_id = times.  
       stop_id  
AND route_id LIKE "N%"
```

```
BATTISTINI- SORIA N1  
BOCCEA- BATTISTINI N1  
BOCCEA- BRA N1  
BOCCEA- VAL CANNUTA N1  
BOCCEA- GREGORIO TREDICESIMO N1  
BOCCEA/URBANO SECONDO N1  
BOCCEA- GALEOTTI N1  
CIRCONVALLAZIONE CORNELIA- BOLOGNI  
CIRCONVALLAZIONE CORNELIA- AURELIA  
...  
CAVE ARDEATINE ^^IN9  
MARMORATA- VANVITELLI^^IN9  
ARA COELI- PIAZZA VENEZIA^^IN9  
TERME DIOCLEZIANO^^IN9
```

## 8) List all routes that connects Rebibbia to Tiburtina Station

```

SELECT DISTINCT route_id
FROM times t JOIN stops s
      JOIN trips tr
ON t.`trip_id` = tr.trip_id
AND s.stop_id = t.stop_id
WHERE s.stop_name LIKE "%
      REBIBBIA%"
AND tr.route_id = ANY (
SELECT distinct tr.route_id
FROM times t JOIN stops s
      JOIN trips tr
ON t.trip_id = tr.trip_id
      AND s.stop_id = t.
      stop_id
WHERE s.stop_name = "
      TIBURTINA"
OR s.stop_name = "STAZIONE
      TIBURTINA"
)

```

120F

163

N2

N23

MEB