

Lorenzo Lancia

Alessandro Gallo

Master Degree in Data Science  
Sapienza Università di Roma

# A sequel to AtacSql

Indexing

Project 2 for DMDS course

atac

ROMA



# Table of Contents

Introduction

A few simple queries

A more complex query ...

# Introduction

- Aim of this project was test how indexing affect the performance of database
- 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 5.7.12

- To ensure a more or less fair comparison we need to disable the dbms cache

```
SET SESSION query_cache_type=0;
```

- Then we need to enable profiling to record the execution times of each queries

```
SET profiling = 1;
```

# A few simple queries...

Let's start with a simple query:

```
select * from calendar where  
    `date`="2016-04-28"
```

```
alter TABLE calendar add key  
    (`date`)
```

without indexing it takes  
0.019732 seconds to execute  
the query

by adding the index we can  
reduce the execution time up  
to a factor 10, arriving at  
0.002789 seconds

Using a longer table the performance difference it's more accentuated. The table `times` has got  $\sim 4 \cdot 10^6$  rows.  
The query:

```
SELECT * FROM `times` WHERE  
    trip_id = "766_1426895"
```

takes 5.82667500 to be  
executed without an index

```
alter table times add  
    PRIMARY key (trip_id,  
    stop_sequence)
```

but adding the primary key  
we reduce it to 0.00509000  
seconds.

# Here comes the JOIN

Let's take the first query of the first project

```
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"
```

It's too hard to DBMS to execute it without any index

with the previous created index

```
alter table times add PRIMARY key (trip_id, stop_sequence  
    )
```

it takes 46.48726400 seconds to complete the query



adding an index also on the `stop_id` attribute

```
alter table times add key (stop_id)
```

it takes 10.43061100 seconds to complete the query

indexing an other table it's useful as well

```
alter table trips add PRIMARY KEY(trip_id)
```

it reduces the execution time to 0.03666900

indexing the third column, or adding foreign keys it's not useful to increase the speed of execution of this query.  
For example

```
alter table stops add PRIMARY KEY (stop_id)
ALTER TABLE times ADD CONSTRAINT trip_id FOREIGN KEY (
    trip_id) REFERENCES trips(trip_id)
ALTER TABLE times ADD CONSTRAINT stop_id FOREIGN KEY (
    stop_id) REFERENCES stops(stop_id)
```

doesn't alter the execution time