

L. Lancia

G. Salillari

Cloud Computing

Master Degree in Data Science

Sapienza Università di Roma

# Facebook Tao

Distributed Data Store for the Social Graph

# Table of Contents

The Data Model

Architecture

Implementation

Workload & Performance

# Introduction

## What is TAO?

### Tao

is a geographically distribute store

- deployed at Facebook
- with efficient and timely access to social graph
- using a fixed set of query
- replacing memcache
- running on thousands of machines
- provide access to many PB of data
- process a billion reads ad millions of writes each second!

## The social graph

Facebook has more than 1 billion active user

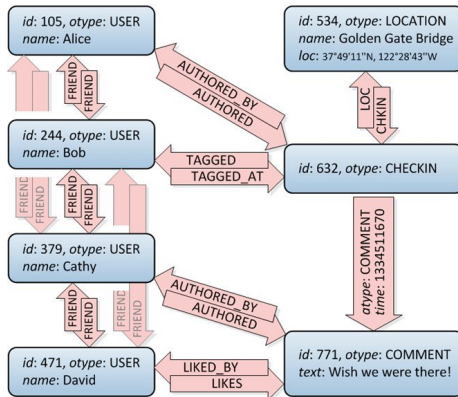
- recording relationships,
- sharing interests,
- uploading pictures and ...

The user experience of Fb comes from rapid, efficient and scalable access to the *social graph*

What's behind an entry in your Fb page?

A single Fb page aggregate and filter hundreds of items from the social graph.

producing user-tailored content and checking for privacy



## Before Tao

Facebook was storing the social graph to MySQL

Quering it from PHP

Storing result in memcache

---

Over time Fb deprecated direct access to MySQL in favor of a graph (assosiation, nodes) abstraction

## Limits

- Inefficient edge list
- Distributed Control Logic
- Expensive read-after-write consistency

Also they want to access social graph from non-PHP services

## TAO's Goals

- Provides basic access to nodes and edges
- Read optimization
- Favour availability over consistency
- Able to tolerate fast changing of the data



# Tao Data Model

# title

# title

# title