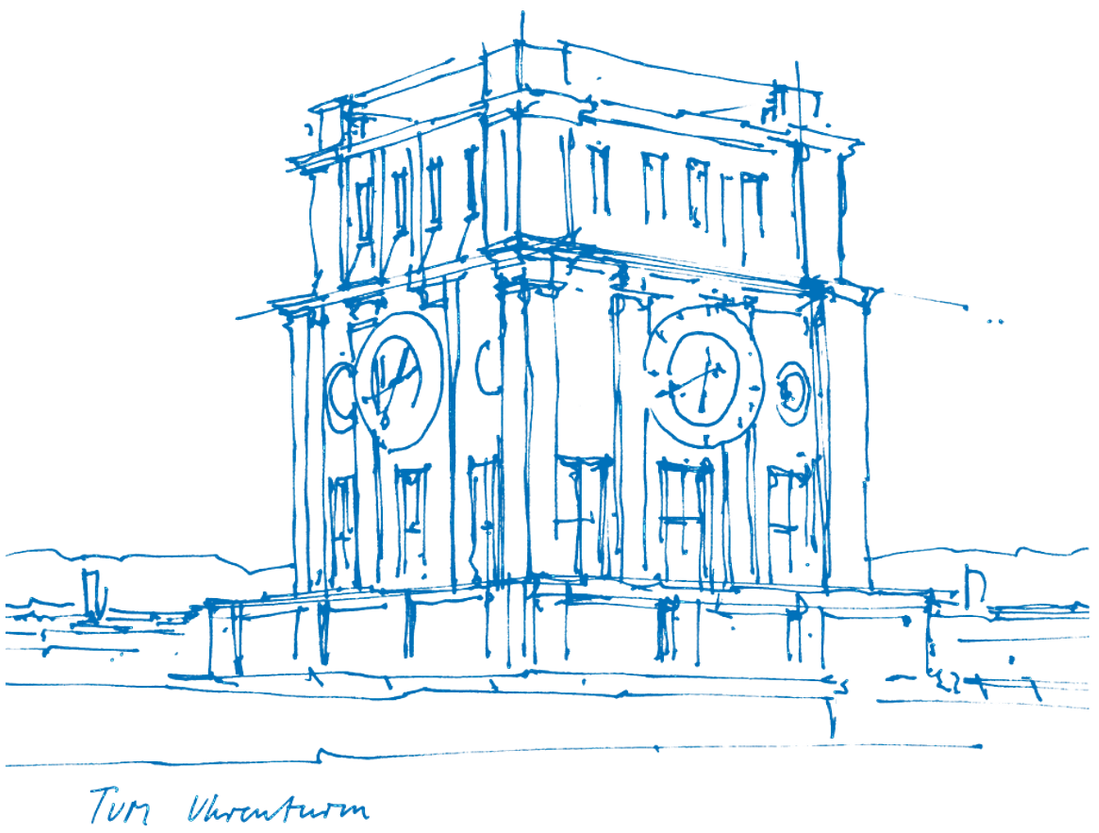
Scientific Seminar on Security in Information Technology

Relational and NoSQL Databases: A Comparison and Introduction to Hybrid and Mining Frameworks

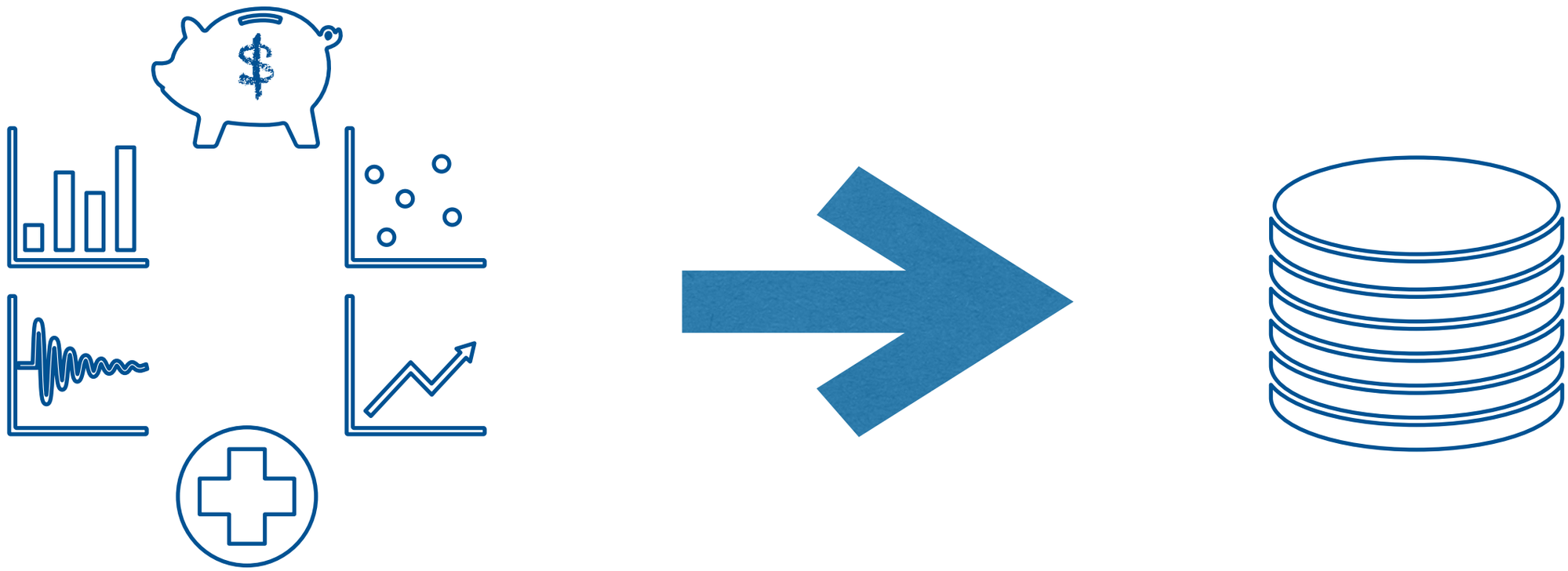
Teodor Fratiloiu

Technische Universität München
Fakultät für Elektro- und Informationstechnik
Lehrstuhl für Sicherheit in der Informationstechnik

Munich, January 22, 2021



We have a lot of data. How do we store it?



A Widely Used Solution:

In a database.[1]

1. E. Codd, A Relational Model of Data for Large Shared Data Banks (1970)

Today's Agenda

1. Crash Course on Database Families

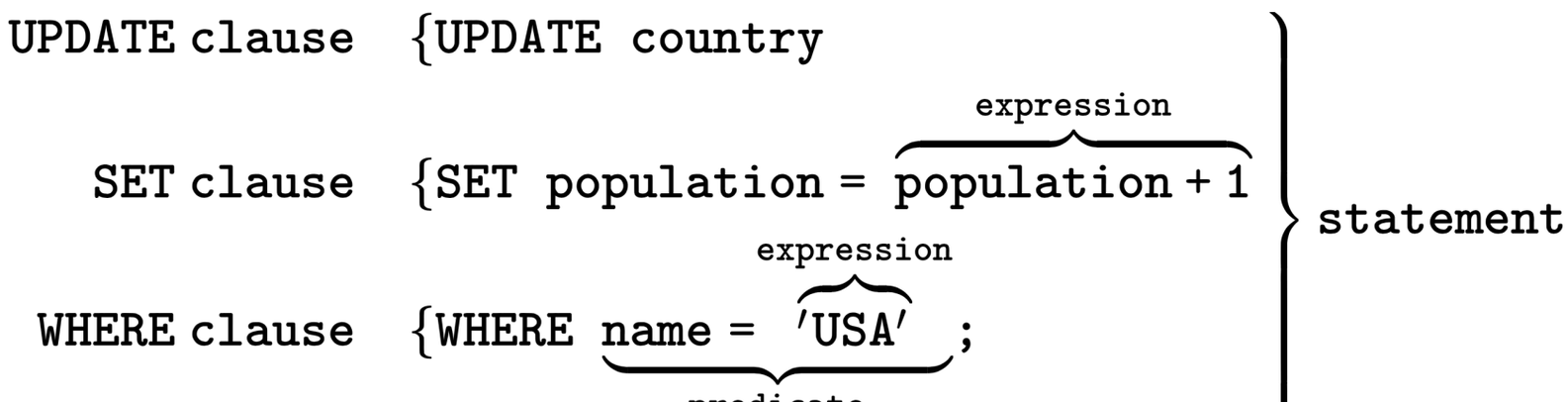
2. Comparing SQL and NoSQL

3. The Hybrid Database Approach

4. Graph Mining and Other Technologies

Part 1. Crash Course On Database Families

What *Is* a SQL Database?



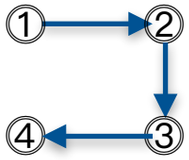
Data Tuples → queries are declarative

Scales Vertically → Large Processing & Memory Overheads

ACID (Atomicity, Consistency Isolation, Durability) property

Image Source: ANSI/ISO/IEC International Standard (IS). Database Language SQL—Part 2: Foundation (SQL/Foundation). 1999.

What *Is* a NoSQL Database?



Designed for connected and distributed systems

Data is saved as a custom structure (e.g. Graph)

BASE (Basically Available State and Eventually Consistent) property

Neo4j Case Study

- Graph Type Data (edges, nodes, attributes)
- CypherQL
- Supports ACID

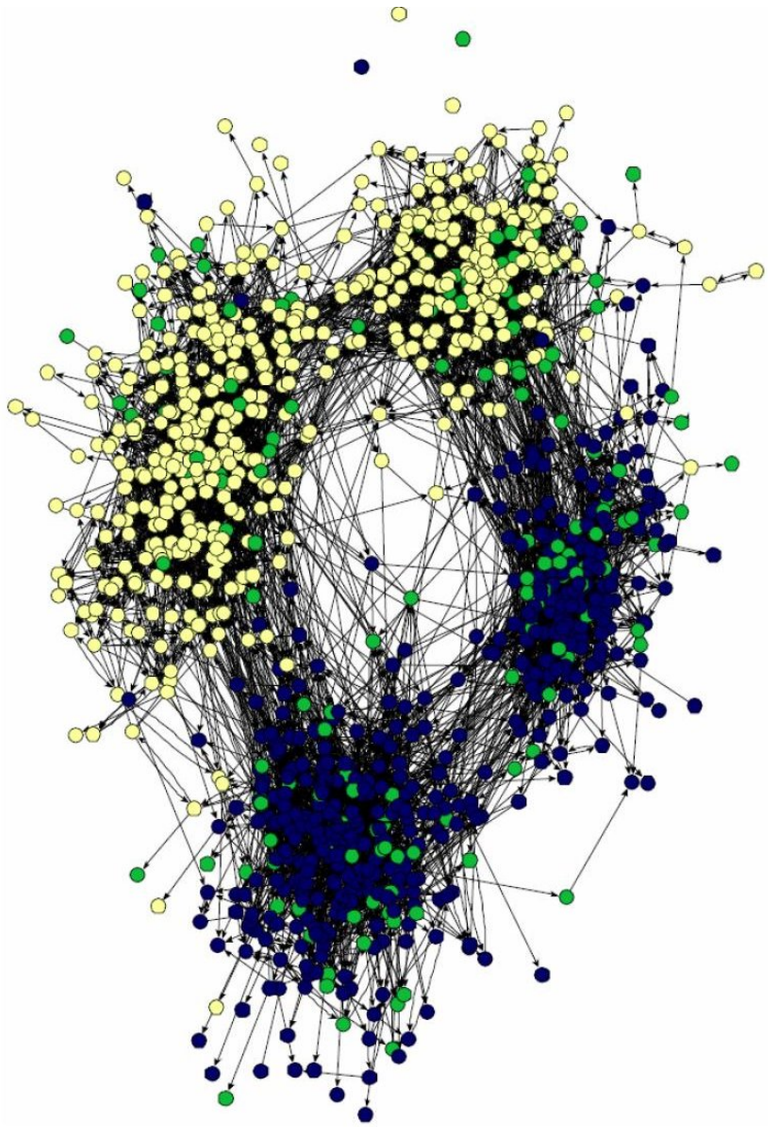


Image Source: http://people.seas.harvard.edu/~babis/kdd09.html

The CAP Theorem

Consistency: system stability after *write* operations

Availability: ready for *read* operations

Partition Tolerance: ability to function when system resources are spread out

Relational DB's support all 3 properties

NoSQL DB's usually support only Availability and Partition Tolerance

Part 2. Comparing SQL and NoSQL

Comparing the Two Sides

Team SQL

- ✅ Legacy format means wide support
- ✅ More reliable for Security Applications
- ✅ One universal query language
- ❌ Scales Vertically
- ❌ Relatively slow

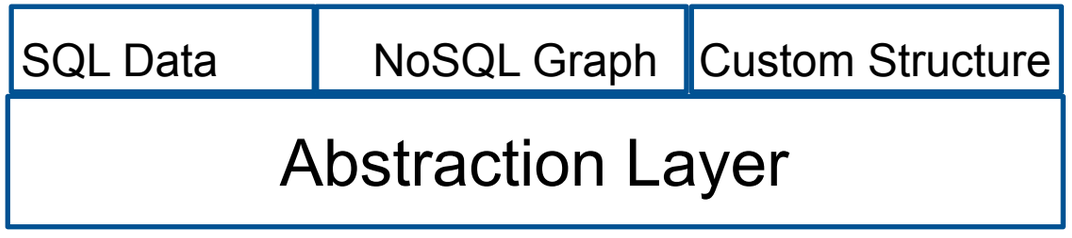
Team NoSQL

- ✅ Very Fast
- ✅ Scalable, ACID Property (some frameworks)
- ✅ Designed for distributed systems
- ❌ Custom Frameworks & Languages

Team NoSQL wins in speed comparisons.
Team SQL offers better support and security.

Part 3. The Hybrid Database Approach

What *Is* a Hybrid Database?



Different DB types accessed through a common abstraction layer

A pure software solution

Compromise solution to address all drawbacks (e.g. speed vs. consistency)

Targeted Design Optimizations can also help improve SQL DB's performance

Originally described in A Hybrid Database Approach Using Graph and Relational Database by H.R. Vyawahare, P.P. Karde and V.M. Thakare (2019)

Part 4. Graph Mining and Other Technologies

Graph Mining

Helps visualize data from different formats (e.g. XML) as a Graph

A pure software solution (again) for Data Science

Data Nodes and Clusters enhance decision-making and pattern-recognition

Image Source: https://www.facebook.com/zuck

Originally described in Graph Mining Frameworks for Finding and Visualizing Substructures using Graph Databases by Swapnil Shrivastava and Supriya N. Pal (2009)

Summary

Relational and NoSQL Databases: A Comparison and Introduction to Hybrid and Mining Frameworks

- Compared the two families of databases:
    - SQL
    - NoSQL (Graph)
    - Proved that neither is a complete winner
- Introduced some compromise solution:
    - The Hybrid Approach
    - Design Optimization and Database Tuning
- Presented Graph Mining
- Open-source Graph Mining and Hybrid DB Tools need to be developed
- Standards need to be established. More Research needs to be done.

Thank You!