
# Google Cloud vs Azure Cloud

## Advanced Databases project

**Simon Coessens, Valerio Rocca, Ludovica Caiola, Maria Camila Salazar**

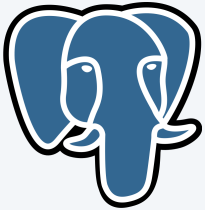
Advanced Databases
École polytechnique de Bruxelles
Université Libre de Bruxelles

December 15, 2023

# Overview

# Introduction

# Introduction to PostgreSQL



- RDMBS
- 1986 (Started as INGRES)
- Michael Stonebraker - Berkeley
- 1994 support SQL.

**OPEN CODE AND FREE**

# Why PostgreSQL?

## Support for Advanced Data Types

- JSON. Geomtric types

## Extensibility and User-Defined Functions

## ACID Transactions and Concurrency

- Atomicity, Consistency, Isolation, Durability

## Replication and High Availability

- Transaction - based replication

## Query Optimization and Query Planner

- Complex optimization

## Open Source License and Active Community

- Continuous development

# Technology Fundamentals

# What is Cloud Computing?

**Definition by Azure team**: Cloud computing is the delivery of computing services over the internet to offer faster innovation, flexible resources, and economies of scale.

Cloud Computing milestones:

- **1961**: definition of utility computing by J. McCarthy.
- **2006**: introduction of Amazon Web Services (AWS).
- **2009**: new actors in Cloud Computing.

# Actual situation

Worldwide market share of leading infrastructure services providers in Q2 2023, including platform as a Services (PaaS) and Infrastructure as a Service (IaaS) as well as hosted private services

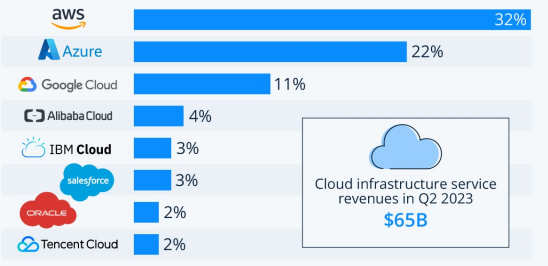
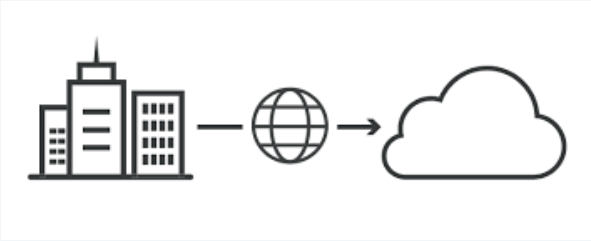



Figure: Cloud computing service

# Comparison with traditional technologies

- Migration to cloud for SMBs on-premise:
  - Why is it so important?
  - For which reasons?



# Secure and protected environment

SMBs can delegate security problems relying on big companies of cloud computing services.

Main threats:

- Malware can bring to a completely loss of data if local servers are attacked
- Ransomware can bring to a company failure in exchange of the data

Key components in cloud computing environment: **backups**, **cloud support**

Proper education: excessive privilege, strong password, MFA.

# Cost

- Cost maintanance are entrusted to cloud services
- Relevant cloud support service offered by cloud services
- In case of expansion of the company, being on cloud is a cheaper option

# Scalability

- Possibility to scale IT resources up or down:
  - increase or decrease resources
- Little focus on vertical scaling: increment of the server's computing power, adding more RAM or CPU cores, or adding storage capacities (SSD).
- Cost-efficient feature: local storage tends to cost more

**Business agility:**

- ability to adapt to both internal and external changes: fast-changing realm may involves scaling resources

# Introduction to the two tools

# Introduction to Microsoft Azure

- Launched in 2010 as response to AWS Cloud.
- Second biggest market share holder at 22% market share
- Adaptation to trends: recent years a great focus on AI

# Introduction to Microsoft Azure


## IaaS (Infrastructure as a Service):

- Azure Virtual Machines (VMs)

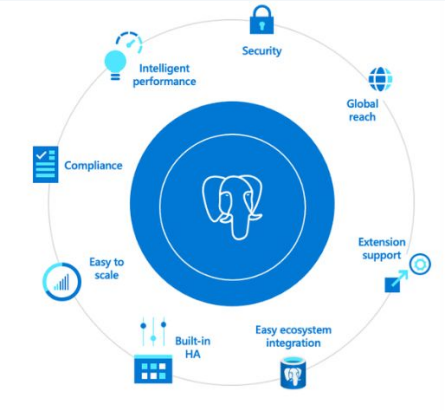
## PaaS (Platform as a Service):

- Azure SQL Database

## SaaS (Software as a Service):

- Microsoft 365 (formerly Office 365)

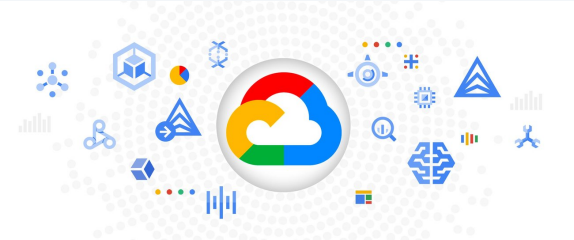
# Azure PostgreSQL Database




- Simplified Management (PaaS)
- Automatic Scalability
- Integrated Security
- Integration with Other Azure Services
- Development and Deployment Ease
- Flexible Billing

# Introduction to Google Cloud

- Google Cloud: a term for encompassing all of Google's Cloud-based services
- Google Cloud Platform: is included in Google Cloud and it is a series of cloud computing services



# Considered, but not chosen services

- **Cloud Spanner**:
  - distributed SQL database management and storage service
  - Horizontal scalability
- **BigQuery**:
  - data warehousing tool
  - geospatial analytics

# 1 - Google Cloud SQL

Fully managed relational database service for MySQL, PostgreSQL and SQL Server.
Features:

- **Fully managed**: database, backups, replication, encryption
- **Open and standards-based**: open-source and commercial engines
- **Easy migration**: with Google Cloud SQL's Database Migration Service
- **Integrated**: possibility of building an application with Compute Engine
- **Backups**: scheduled or on-demand

# 2 - Google Cloud SQL

Each Cloud SQL instance is powered by a Virtual Machine running on a host Google Cloud server $\Rightarrow$ flexibility, scalability, isolation of resources

- **Geospatial support**: Google Cloud SQL for PostgreSQL supports the popular PostGIS extension
- **Simplicity**: user-friendly tool
- **Vertical scalability**: increasing server's computational power, when it is necessary

# Introduction to the project

# Background and context

**Taxi Lecco**: organization that represents taxi drivers in the city of Lecco, in Italy. Taxi Lecco is planning to develop an **application** to request taxi rides in the city.

Figure: Lecco

# Project Objective

Decide whether it is convenient to migrate the organisation's **RDBMS** to the cloud and, in case, if **Microsoft Azure** or **Google Cloud** solutions are better.

Try to answer these questions:

- Is it worth it?
- In which terms?
- At what cost?
- On which tools?

# Scope and Limitation

We approximated the ”pay-as-you-go” costs as **costs per transactions**.

**Motivation**: we cannot predict apriori the exact amount of resources spent since:

1. Assigned resources are **not fixed**.
2. ”Pay-as-you-go” costs depend on a **series of factors**.

# Introduction to the Database

# Taxi Lecco's Data

- Requirement analysis for Taxi Lecco
- Essential for determining Cloud server size

| Season | Hour | Average number of rides per hour | Hours per day* |
|---|---|---|---|
| Touristic | Peak | 403 | 3 |
| Touristic | Other daytime | 184 | 13 |
| Touristic | Nighttime | 66 | 8 |
| Non-touristic | Peak | 103 | 1 |
| Non-touristic | Other daytime | 70 | 15 |
| Non-touristic | Nighttime | 15 | 8 |

Table: Taxi rides data. *: hours per day belonging to that category

# The Database

- Operational database
- Spatial data handled using `POSTGIS`

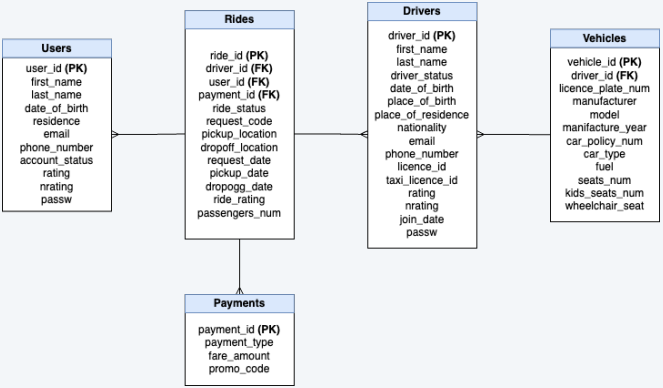



Figure: ER schema

# The Database

- Inspiration from online Datasets
- We made our own specific dataset using python's `Faker` library
- The scale factor decides number of drivers, 10 times number of users, 100 times number of rides.

| user_id | first_name | last_name | date_of_birth | residence | email |
|---|---|---|---|---|---|
| 1 | Terry | Nguyen | 1992-02-22 | East Gabrielaport | pclark@example.net |

| vehicle_id | licence_plate_num | manufacturer | model | manifacture_year | car_policy_num |
|---|---|---|---|---|---|
| 1 | AO-7811 | Green, Wells and Long | their | 2005 | DLGX31688437815593 |

# System Operations

1. Updates driver status after a ride is completed and when the driver stops or resumes working.
2. Inserts a new ride after a ride is completed.
3. Inserts a new payment after a ride is completed.
4. If done by the user, updates the rating of a driver after a ride is completed.
5. If done by the user, updates the rating of a user after a ride is completed.
6. When a client creates an account, inserts a new user in the Users table.
7. When a new taxi driver registers, inserts them in the Drivers table.
8. When a new vehicle is registered, inserts them in the Vehicles table.
9. At 3 A.M. every day, data in tables Rides and Payments is copied in a .csv file and then removed from the RDBMS (this operation is later referred also as ”daily ETL”).

# Cloud Configuration: Azure vs G-Cloud

- How to make a fair comparison ?
- Two tools and two different machines

| | Google Cloud SQL | Azure |
|---|---|---|
| DB version | PostgreSQL 15.4 | PostgreSQL 15.4 |
| Location | US Central 1 | Central US |
| vCPUs | 1 vCPU | 1 vCPU |
| Memory | 3.75 GB | 2 GB |
| Data Cache | Disabled | - |
| Storage | 32 GB | 32 GB |
| Backup | Automated | 7 days |
| Availability | Single zone | Single zone |
| Service | Enterprise | - |

Table: Configuration of cloud servers

# Cloud configuration Azure vs G-Cloud

- `POSTGIS` extension is directly available in Cloud SQL, needed to be enabled in Azure
- Connected to the Cloud instances with our local machines
- We used HammerDB software to perform benchmarks
- Data is generated locally and then pushed to the Cloud instances

# Benchmark

# Project's benchmark

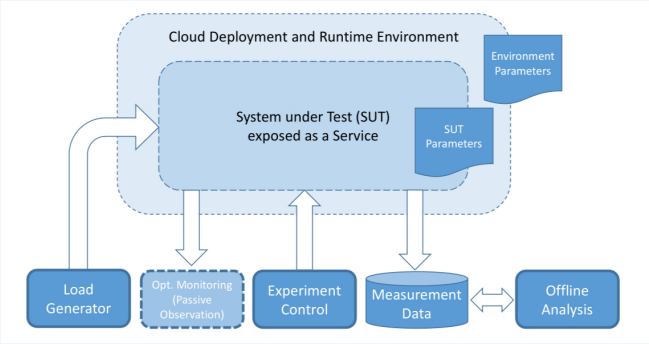



Figure: Standard components of Benchmarking

# Workload: Lecco's taxi case

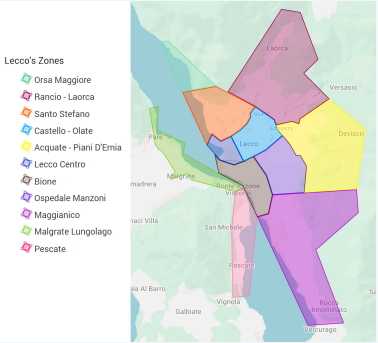



Figure: Satellite view of Lecco's chosen zones

Inspired by some databases from internet, we recreated the data with Faker library.

# Workload: Queries

```
1       INSERT INTO rides (
2       ride_id, driver_id, user_id, ride_status, request_code,
    pickup_location, dropoff_location, request_date, pickup_date,
    dropoff_date, ride_rating, payment_id, passengers_num)
3       VALUES (
4       %s, 1, %s, 'TEST', 27164,
5       ST_MakePoint(40.89911326259291, -74.19879968),
6       ST_MakePoint(40.89119552115688, -73.78466389),
7       '2023-05-04', '2023-07-07 20:03', '2023-08-23 07:54',
8       4, 1, 2);
9
```

Listing 1: Insert transaction

Update, Inserts, Aggregation and Delete queries were designed to strees the SUT's.

# Benchmark execution

## Experiment control:

- Scales: 2, 10, 100, 1000.
- 3 runs: Warming
- 7 runs: Analysis

## Measurement data:

- Save in .txt files

## Offline analysis:

- Python

# TPROC-C

- **What TPROC-C is**: open-source benchmark based on TPC-C.
  - **What TPC-C is**: OLTP benchmark where five different transactions of varying types and complexities are executed simultaneously.
  - **TPC-C Structure**: database with nine tables simulating the operations of a wholesale supplier.

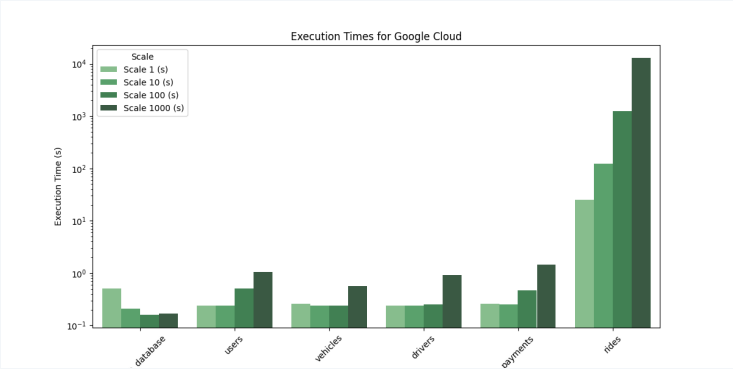
- **Why TPROC-C**:
  - **Reliability**: test on a industry-recognized benchmark.
  - **Cost computation**: since it outputs the number of transactions per minute, it allows the computation of the cost per transaction.
  - **Multiple users**.

# TPROC-C Execution

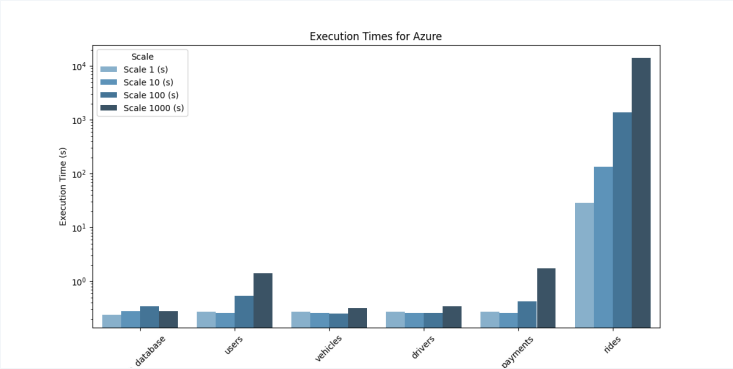
- Made through **HammerDB**, a benchmarking software.
- **Five scale factors**: 500, 5,000, 50,000, and 500,000 transactions per user.
- Six users to fill twelve Data Warehouses.
- Six runs for each scale factor (first one not recorded).

# Results

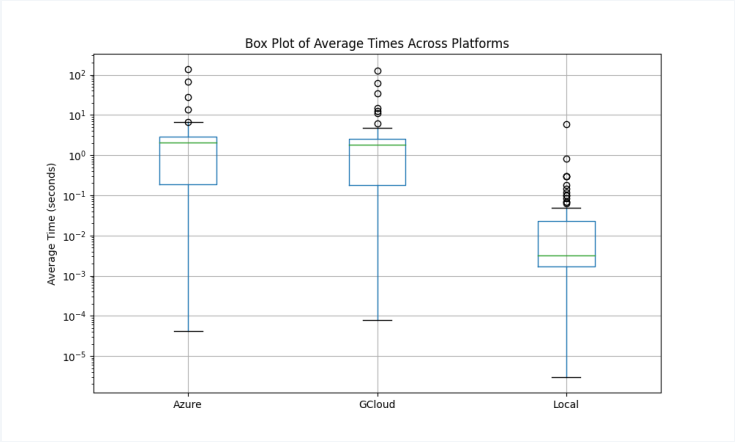
# Project Benchmark Results: Loading



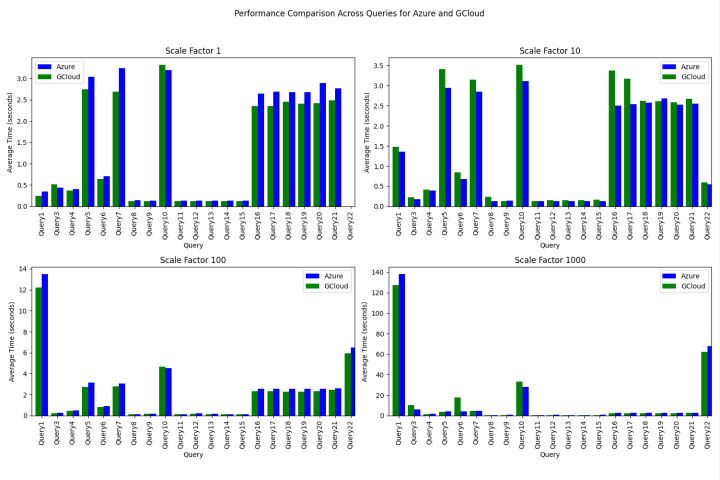
# Project Benchmark Results: Loading



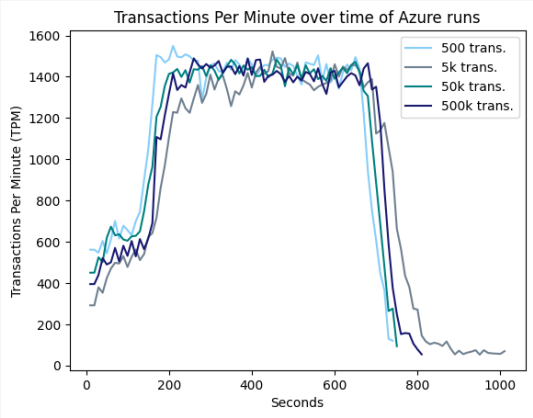
# Project Benchmark Results: Queries

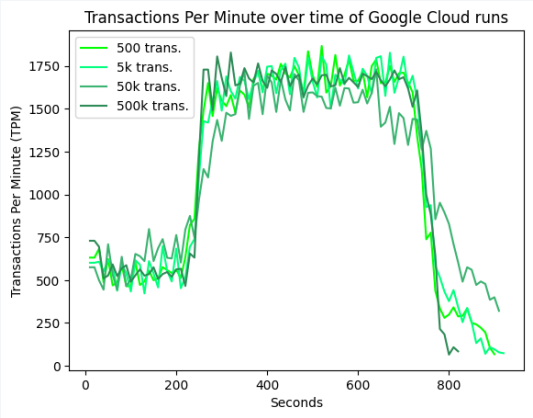


# Project Benchmark Results: Queries

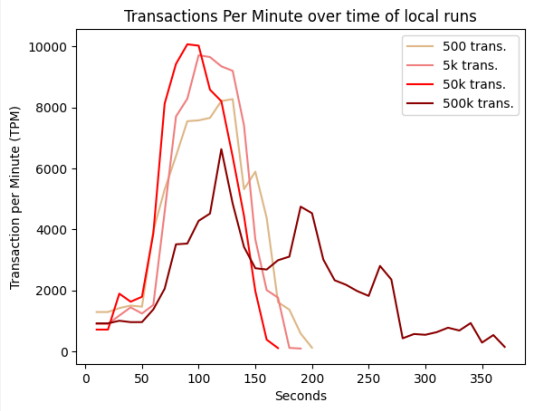


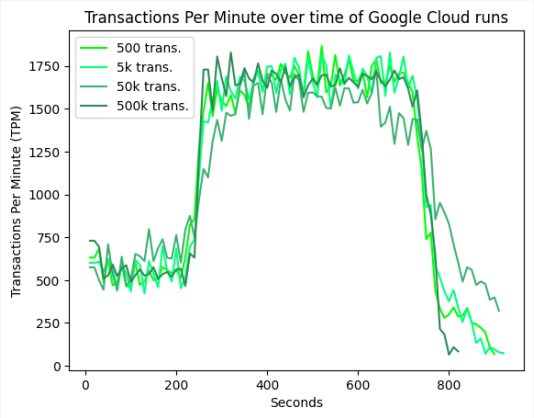
# TPROC-C Results - Azure VS G. Cloud



# TPROC-C Results - Local VS Cloud



# TPROC-C Results - Costs

| Provider | Tot. exec. trans. | Tot. cost | Cost per 10,000 trans. |
|---|---|---|---|
| Google Cloud | 2,354,322 | 11.27€ | 0.047869€ |
| Microsoft Azure | 3,002,478 | 20.53€ | 0.068377€ |

- 1 USD = 0.92 EUR

# Discussion and Conclusion

# Cost Metric

**MVC (Monthly Variable Cost)**

$$MVC = [(PARh \cdot PNh) + (OARh \cdot ONh) + (NARh \cdot NNh)] \cdot CPT \cdot 30 \cdot K$$

where:

- **PARh, OARh, NARh**: respectively Peak, Other daytime, and Nighttime average number of rides per hour;
- **PNh, ONh, NNh**: respectively Peak, Other daytime, and Nighttime hours per day;
- **CPT**: cost per transaction;
- **30**: number of days in a commercial month;
- **K**: coefficient which determines the number of transactions in the daily operations performed on the database. Assumed between 8 and 15.

# Results (K=8)

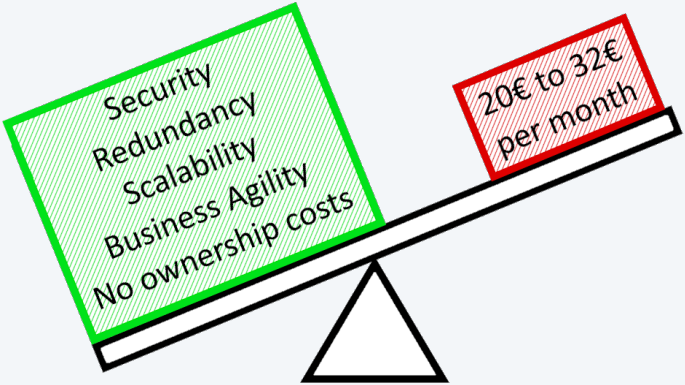
| Provider | Season | Zone | Fixed Costs | Variable Costs | Total cost |
|---|---|---|---|---|---|
| Google | Touristic | Single | ∼54€/mo | ∼5.05€/mo | ∼59€/mo |
| Google | Non-Touristic | Single | ∼54€/mo | ∼1.47€/mo | ∼55€/mo |
| Google | Touristic | Multi | ∼108€/mo | ∼5.05€/mo | ∼113€/mo |
| Google | Non-Touristic | Multi | ∼108€/mo | ∼1.47€/mo | ∼109€/mo |
| Azure | Touristic | Multi | ∼18€/mo | ∼7.21€/mo | ∼25€/mo |
| Azure | Non-Touristic | Multi | ∼18€/mo | ∼2.09€/mo | ∼20€/mo |

# Results (K=15)

| Provider | Season | Zone | Fixed Costs | Variable Costs | Total cost |
|---|---|---|---|---|---|
| Google | Touristic | Single | ∼54€/mo | ∼9.47€/mo | ∼63€/mo |
| Google | Non-Touristic | Single | ∼54€/mo | ∼2.76€/mo | ∼57€/mo |
| Google | Touristic | Multi | ∼108€/mo | ∼9.47€/mo | ∼117€/mo |
| Google | Non-Touristic | Multi | ∼108€/mo | ∼2.76€/mo | ∼111€/mo |
| Azure | Touristic | Multi | ∼18€/mo | ∼14€/mo | ∼32€/mo |
| Azure | Non-Touristic | Multi | ∼18€/mo | ∼3.92€/mo | ∼22€/mo |

# Conclusion

- We highly suggest Taxi Lecco to migrate their OLTP to **Microsoft Azure**.



ULB UNIVERSITÉ LIBRE DE BRUXELLES

# Thank you for your attention

**Simon Coessens, Valerio Rocca, Ludovica Caiola, Maria Camila Salazar**

Advanced Databases
École polytechnique de Bruxelles
Université Libre de Bruxelles

December 15, 2023