Database Administration
Lab 07: Backup & Restoring.

Andrés Calderón, Ph.D.

March 17, 2025

1 Introduction

In today's data-driven world, ensuring the availability and integrity of data is a critical aspect of database administration. Data loss can occur due to various factors, including accidental deletions, system failures, or catastrophic events. To mitigate these risks, organizations implement Backup & Restore (B&R) strategies to recover lost data and maintain operational continuity.

This lab focuses on applying real-time database management concepts in a simulated environment where students will:

- Utilize real-time data from the SUMO (Simulation of Urban MObility) traffic simulator.

- Set up a database to store and manage location data.

- Simulate a catastrophic event leading to data loss.

- Implement a Backup & Restore tool to recover lost data and assess its effectiveness.

Through this hands-on experience, students will gain a deeper understanding of data resilience techniques, best practices for database recovery, and the importance of periodic backups in real-world applications.

By the end of this lab, students will deliver a presentation on their approach, findings, and insights, with peer evaluation as part of the assessment process. This lab serves as a practical exercise in database administration, reinforcing theoretical knowledge with real-world implementation.

2 Getting Real Data

So, to obtain real-time data, we will use SUMO—Simulation of Urban MObility. It is an Eclipse Open Source project that provides a highly portable, microscopic, and continuous multi-modal traffic simulator.

In particular, we will use a scenario created a couple of years ago by DLR-VF[1] using TAPAS (Travel and Activity PAtterns Simulation) and SUMO for the city of Cologne in Germany. We will refer to this scenario as the TAPAS Cologne Scenario throughout this document. You can download the scenario from this link [`TAPASCologne-0.32.0.7z, 2018-04-10, 52.3 MB`].

To understand the details of the scenario and how we will use it in this project, you should watch this video.

3 Setting Up a Database for Location Data

As mentioned in the video, using the data generated by the TAPAS Cologne Scenario and SUMO, we will populate a database in the DBMS of your choice. You can accomplish this using a script written in your preferred programming language, Pentaho Data Integration, or any other tool suitable for handling the XML file that SUMO updates in real time.

---

[1] The German Aerospace Center - Institute of Transport Research

4   Simulating a Catastrophic Data Event to Use a B&R Tool

The goal of this project is to test the use of a Backup & Restore (B&R) tool to recover data and mitigate the damage caused by a hypothetical data loss. First, you need to choose a tool (either Open Source or Commercial) from this list. During the execution of the scenario and the database population process, you will simulate a catastrophic event—whatever form it may take—resulting in data loss.

Specifically, you will inject a 'DELETE' statement into the database to remove a section of the data. For example, around the 8000th second, you could execute an SQL sentence that removes all tuples recorded between seconds 4000 and 6000. It is expected that you configure the Backup & Restore tool to operate periodically within the same time window in which the scenario is running.

Finally, you will investigate and demonstrate how your chosen B&R tool can recover the lost data (or part of it) and mitigate any potential damage.

5   What We Expect

You will choose a DBMS and a B&R tool and prepare a presentation to share with the class how you handled real-time data ingestion, simulated data loss, and properly utilized the B&R tool of your choice.

We expect you to present your work in class on **March 31, 2025**. The presentation will be co-evaluated by your classmates.

Happy Hacking 😎!