

Data Studio Performance Tuning

Jinxiao (Jason) Song, William Kudsk, Michael Zhang,
Chuangbo Tong, Haofan Wang

Abstract

The goal of this project is to take Data Studio, a web-based notebook platform currently in development, and create a performance testing suite that can identify bottlenecks in queries to its database. Using Java Microbenchmarking Harness, we created a testing suite that extracts information related to how much time specific queries take to complete. This information is used to identify performance bottlenecks, and to guide development of future optimizations.

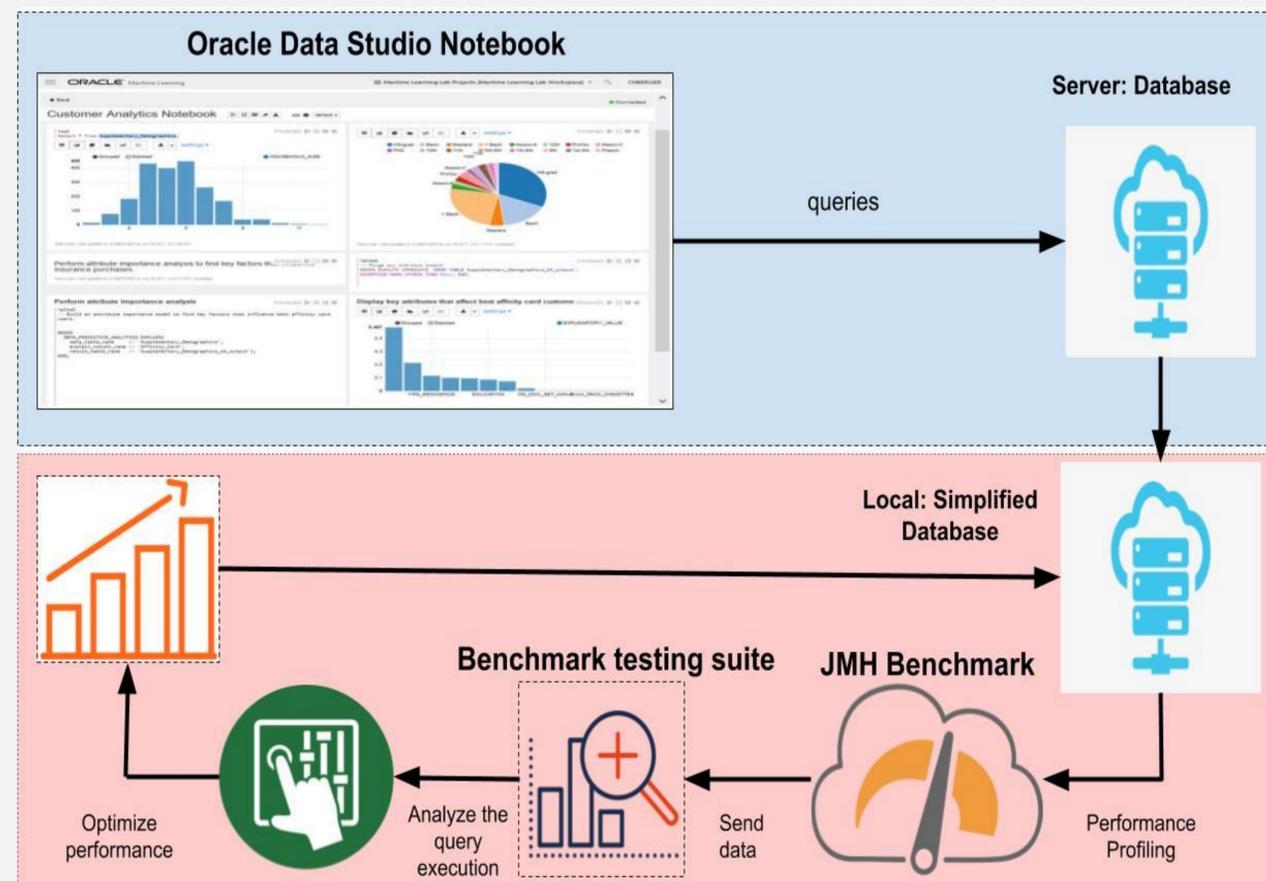
Approach

- Understand Oracle Data Studio's system and data structure
 - Set up Oracle Data Studio in the working environment with **Ubuntu**¹
 - Study **Spring Boot**², **JMH**, and **Hibernate**
 - Use **Hibernate** as the object relational mapping tool between Data Studio objects and our simplified version of Oracle database.
- Integrate benchmarking tool: **JMH**³
 - Add **JMH** dependency in **Gradle**⁴ and create custom Gradle task
 - Import **JMH** tools into benchmark testing suite
 - Use these tools to benchmark the database operations
- Develop benchmark testing suite
 - Write tests for database Create, Read, Update, and Read operations
 - Use **Benchmark** annotation from **JMH** to observe result.
 - Create tests for custom database functions made by Oracle developers
- Use the result of the benchmark to find bottlenecks

Overview

Web-based notebooks are word processing web-applications which can compile code to display numerical and graphical output using various programming languages. Data Studio is a web-based notebook platform that can help with data visualization for data scientists, who deal with many data daily. Because of this, Data Studio has to handle a lot of data, as users can create and share notebooks between each other. As Data Studio is still in development, Oracle developers do not have insight into the performance of database components, as they lack the right tool to do so. Our goal involved building a testing suite that could give us valuable performance related data, and to then identify potential bottlenecks to provide possible optimizations.

Architecture



Technology used:

1. **Ubuntu**: an open source linux system
2. **Spring Boot**: an open source Java-based framework that provide [Java Persistence API](#)
3. **JMH**: a lightweight java benchmarking tool
4. **Gradle**: an open-source build automation tool

Results

We benchmarked database queries for 6 different Data Studio entities:

- Notebook, Job, Comment, User, PermissionMapping and Permission
- For each entity, we benchmarked create, read, update, and delete operations.
- Additionally benchmarked custom operation specific to certain entities.
- **Example: Notebook entity**
Benchmarked functions involving the object's relationship with Users, Permissions, and filterNotebook with user tag.
- Used SQL scripts and different batch sizes of entities saved to the local database to best emulate performance in of data studio in practice
- Investigated ways to achieve optimizations especially for the costly findAll queries

Conclusion

With this project we created a benchmarking platform used to optimize database transactions in Data Studio. This platform is currently being integrated with Data Studio to give Oracle developers a tool for detecting bottlenecks and improving performance.

Acknowledgments

Thanks to:

Daniel Langerenken	Mentor
Dr. Richard Jullig	Instructor
Faeze Brahman	TA
Chandranil Chakrabortti	TA