

Scale out SaaS Analytical Databases with PostgreSQL_

Proposta de Bolsa/Estágio na Altice Labs



ID do Projeto	2024/25_N34
Departamento	SSO4
Proponente	Ricardo Santos Ferreira
Data de validade	31-Dec-2024

IDI - Key areas

Digital Services & Platforms

Subject / Title

Techniques to Scale-out Analytical Databases with SaaS Capabilities

Context

The combination of a SaaS architecture and horizontal scaling capabilities in analytical databases is an increasingly relevant challenge as cloud applications process real-time data from IoT or home networks. Sharding solutions are becoming more popular for handling the scalability of analytics and timeseries platforms. Sharding is a technique that involves dividing a database into smaller fragments, known as shards, which are distributed across different servers. With this technique, insertion and queries can be executed in parallel, significantly improving system performance.

However, implementing these techniques can be complex and present some challenges. One of them is ensuring data consistency across all shards, especially in systems that require high availability and fault tolerance. Furthermore, dividing data into shards may affect the multi-tenant architecture and the efficiency of some queries, particularly those involving operations across multiple shards. It is important to carefully evaluate the system's needs and choose a multi-tenant strategy that meets scalability, performance, and data consistency requirements.

Project Goals

The objective of this internship project is to explore and implement a combination of techniques to enable horizontal scalability and multi-tenant support for the analytical database of Altice Labs' Smart WiFi cloud platform.

This platform processes data from millions of CPEs and connected devices in Wi-Fi networks, aiming to optimize the quality of Wi-Fi network experience and enhance visibility and troubleshooting of customer premises issues.

The scalability of the platform is currently supported by a distributed processing architecture in the cloud, utilizing micro-services and a common message bus.

By the end of the internship, it is expected that the intern will have gained knowledge and experience in sharding and scalability of analytical databases, as well as practical skills in implementing and configuring sharding tools in PostgreSQL within the Altice Labs' Smart WiFi application. The outcomes of this project will be highly valuable for Altice Labs and other companies facing scalability challenges in analytical databases."

Keys

- High-performance and scalable platforms;
- Database sharding techniques and technologies;
- *Row-level security (RLS) techniques;*
- Micro-serviços de analytics.

Tools

Within the scope of this internship, several tools that enhance the sharding technique in PostgreSQL databases should be explored:

- Timescale: High performance PostgreSQL extension for time series. It has the possibility to create distributed hypertables by several PostgreSQL nodes;

- PostgreSQL Citus: PostgreSQL extension that allows the distribution of data in several nodes, using the sharding technique. Allows the user to create a scalable PostgreSQL cluster, transparently distributing data across multiple nodes;
- pg_shard: PostgreSQL extension that provides horizontal sharding. Distributes data across multiple partitions on different servers, providing horizontal scalability and high availability;
- Postgres-BDR: PostgreSQL-based bi-directional replication extension that allows distribution of data between different PostgreSQL database nodes. Supports both asynchronous and synchronous replication, making it suitable for high availability and scalability scenarios;
- Postgres-XL: Distributed database solution based on horizontal and vertical sharding.

References

- <https://www.timescale.com/>
- <https://postgrespro.com/blog/pgsql/5969681>
- https://docs.citusdata.com/en/v7.2/articles/designing_saas.html

Activities

- Study of the state of the art regarding techniques and technologies of sharding and multi-tenancy in analytical databases;
- Study of the architecture of the Smart Wi-Fi platform;
- Implementation of proof of concept (PoC);
- Report of comparative benchmark (tools and techniques);
- Be part of software development process with the smart Wi-Fi platform engineering team to incorporate changes and configurations necessary to support sharding and support for multi-tenancy.

Required Skills

- Software development;
- PostgreSQL;
- Database;
- JAVA.

Supervisor (name and e-mail)

Ricardo Santos Ferreira - ricardo-s-ferreira@alticelabs.com

Para concorrer podes enviar a tua candidatura, envia e-mail para o Programa GENIUS: genius@inova-ria.pt