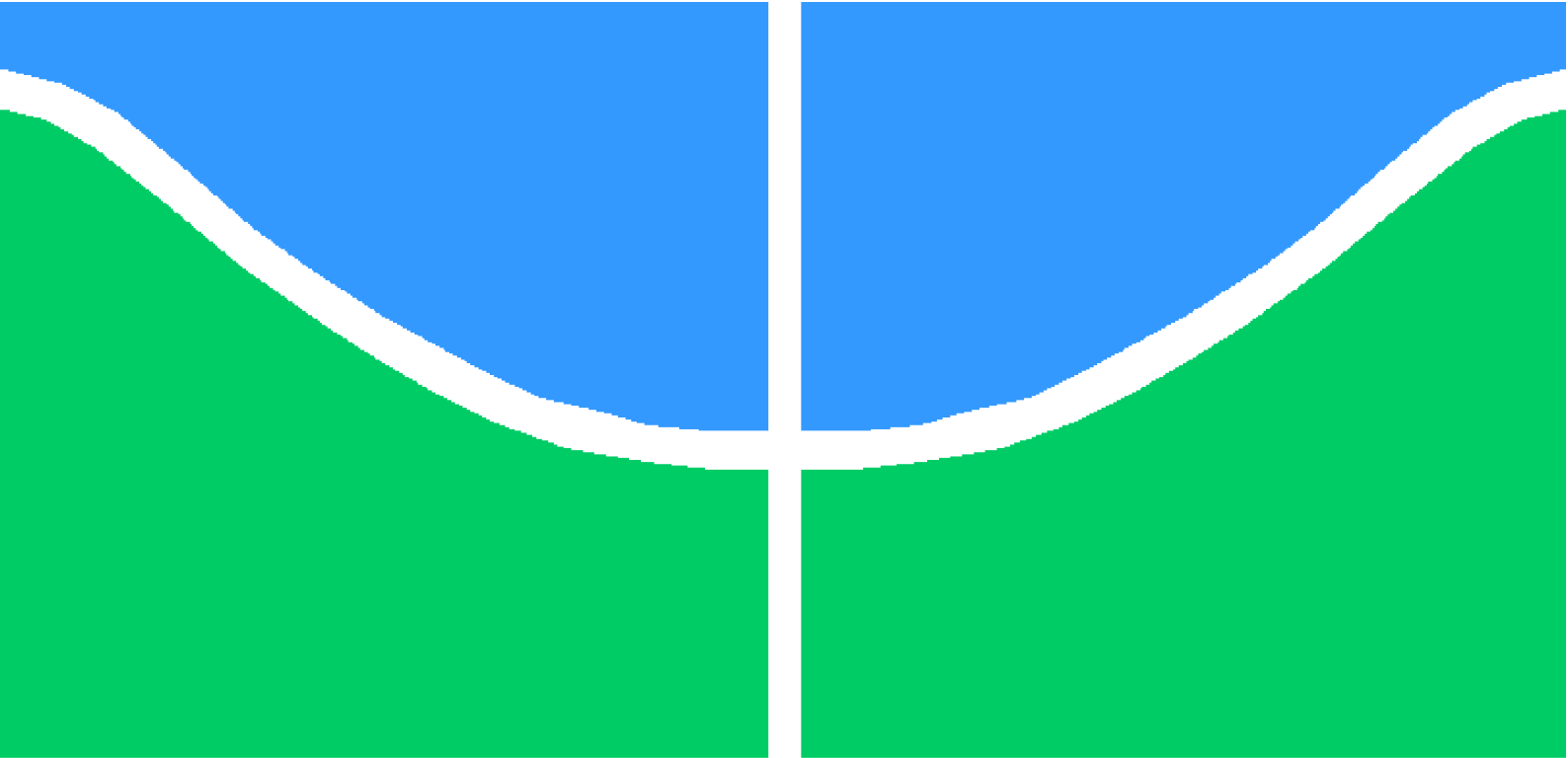




Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Engenharia Aeroespacial

**EP5 Engineer Internship Report: Development
of Python Applications in Snowflake**

Autor: Mateus de Alencar Moura Santos

Orientador: Dr. Sébastien R.M.J. Rondineau

Brasília, DF

2024



EP5 Engineer Internship Report

Development of Python Applications in Snowflake

Mateus DE ALENCAR MOURA SANTOS

Enterprise: Solution BI, 54 rue de Paradis, 75010 Paris

Internship Mentor: Arnaud COL

Academy Mentor: Abdellah HADJADJ

Academic Year: 2023/2024

Internship Period: 10/01 – 11/07/2024

September 2, 2024

Acknowledgments

I would like to begin by expressing my heartfelt gratitude to Solution BI Paris for their warm welcome, which has made my internship a delightful journey from start to finish. The supportive environment and encouraging team have been instrumental in my professional growth.

Special thanks are due to my mentor, Arnaud Col, whom has been a figure of inspiration to me since the beginning. His unwavering confidence in my abilities and constant recognition of my efforts have greatly contributed not only to the development of the projects I worked on but also in my professional career.

I am equally grateful to Alice Jadot, who not only recognized my potential during the interviews, but also facilitated my invaluable connection with my mentor at Solution BI. Her role in bridging the gap between my aspirations and the opportunities at the company has been pivotal.

Furthermore, I extend my sincere appreciation to Nathalie Bourrier, whose perpetual good humor and positive spirit made my integration into the company seamless and my time here thoroughly enjoyable.

This experience has been profoundly enriching, largely due to the excellent company culture that fosters learning and growth. It has paved the way for me to build a solid foundation in the field of data, and for that, I am immensely thankful.

GLOSSARY

- **Business Intelligence (BI):** Technologies, applications, and practices for the collection, integration, analysis, and presentation of business information to support better decision-making.
- **Cloud-Based Platform:** A service that provides computing resources over the internet, allowing for scalable and flexible data storage and processing.
- **Snowflake:** A cloud-based data warehousing platform that allows for the storage and analysis of large volumes of data.
- **Data Preparation:** The process of cleaning, transforming, and organizing raw data into a usable format for analysis.
- **ETL (Extract, Transform, Load):** A process in data warehousing that involves extracting data from various sources, transforming it into a suitable format, and loading it into a storage system.
- **Structured Query Language (SQL):** A programming language used for managing and manipulating relational databases, allowing for data storage, retrieval, and analysis.
- **dbt (Data Build Tool):** An open-source tool that uses SQL to transform data, focusing on the transformation part of the ETL process, and allows for managing and documenting data transformations.
- **Docker:** A popular platform for developing, shipping, and running applications using containerization.
- **Snowpark Container Services:** A Snowflake service that supports containerized applications, allowing for flexible and scalable deployment.

RÉSUMÉ

Ce rapport de stage présente l'expérience d'un Ingénieur de Données chez Solution BI, une société de conseil orientée données. Le stage faisait partie du projet innovant BIG AIR, le premier de son type dans le marché et qui se concentre sur le développement d'applications natives sur Snowflake en utilisant Python et d'autres outils. Chaque application a été conçue avec un objectif spécifique et un public cible à l'esprit, démontrant la puissance de la prise de décision basée sur les données dans divers secteurs en utilisant des applications simples mais efficaces. L'expérience a fourni des informations précieuses sur l'utilisation pratique de ces outils dans des scénarios du monde réel et l'impact qu'ils peuvent avoir sur les résultats commerciaux. Le rapport est organisé en sections qui présentent l'entreprise et le contexte du stage, expliquent les concepts et technologies clés utilisées, présentent les résultats du stage et concluent par des réflexions sur l'expérience.

Mots clés: Snowflake, Python, Streamlit.

ABSTRACT

This internship report presents the experience of a Data Engineer at Solution BI, a data-oriented consulting company. The internship was part of the innovative BIG AIR project, the first of its kind on the market and which focuses on the development of Snowflake Native Apps using Python and other tools. Each application was designed with a specific goal and target audience in mind, demonstrating and extending the power of data-driven decision making in various sectors with the use of simple but efficient applications. The experience provided valuable insights into the practical use of these tools in real-world scenarios and the impact they can have on business outcomes. The report is organized into sections that introduce the company and the context of the internship, explain key concepts and technologies used, present the results of the internship, and conclude with reflections on the experience.

Key-words: Snowflake, Python, Streamlit.