

## PERSONAL INFORMATION

---

**Name:** João Vitor Rodrigues Baptista

**Location:** Brasilia, Brazil

**Fone/Whatsapp:** +55 (61) 994031754

**Email:** [jvrbaptista@live.com](mailto:jvrbaptista@live.com)

**Social medias:** [GitHub](#), [Linkedin](#), [Credly](#), [Medium](#), [Portfolio](#).

## EXPERIENCE

---

### [Itaú Unibanco](#)

*Data Engineer - Sirius Platform*

(Remote) São Paulo, Br

Sep 2022 - Current

- **Achievements:** Successfully migrated over 4 SAS legacy jobs to AWS, achieving over 16x time faster in processing those jobs.
- Led the modernization of a legacy data pipeline, resulting in reduction in errors.
- Trained over 200 persons in data engineering, data mesh, and AWS Data Services for consumers.
- **Key Responsibilities:** Migrated SAS legacy jobs to AWS using AWS Athena and AWS Glue. Modeled dimensional and analytics tables to serve internal stakeholders and visualizations using lakehouse/mesh paradigms.
- Modernized legacy jobs using unit testing, clean code, standards templates, data quality and observability.
- Developed and delivered training courses on data engineering, data mesh, and AWS Data Services for consumers.
- **Skills:** AWS, AWS Glue, Apache Spark, Cloud Computing, AWS Athena, Terraform, CI/CD and Dimensional Modeling.

### [XP Inc](#)

*Data Engineer - Credit Risk Platform*

(Remote) São Paulo, Br

Aug 2021 - Sep 2022

- **Achievements:** Developed a ELT and ETL solutions and components that saved the up to 80% in time of development of new pipelines
- Manage over 30 spark pipelines in batch and streamming modes, orchestrated with Airflow DAGs.
- Reduced the time to deploy new data flows by 50% by using CI/CD pipelines using unit testing, SonarQube and Lints.
- Increased the number of automated tests by over 70% of coverage in data pipelines code.
- **Key Responsibilities:** Designed, developed, and implemented data flow solutions.
- Managed data quality and QA processes.
- Refined DevOps culture in Data Chapter.
- Used/Applied Lakehouse paradigms.
- **Skills:** Azure Databricks · PySpark · Apache Spark Streaming · Delta Lake · Apache Kafka · Airflow · Azure DevOps · Continuous Integration and Continuous Delivery (CI/CD) · Microsoft Azure · MongoDB

*Data Engineer - Compliance Offshore*

(Remote) São Paulo, Br

Aug 2020 - Aug 2021

- **Achievements:** Structured data flow solutions that saved the company annual costs by finishing a third party compliance company partnership.
- Created OnShore Compliance Solution that met all regulatory requirements.
- Created CI/CD pipelines that reduced the time to deploy new data flows by 50
- Created RESTful API that improved the performance of AML, Matrix, Analytics and Trade surveillance by 80% compare with the third party compliance company solution.
- **Key Responsibilities:** Structured data flow solutions (ELT and ETL in spark code).
- Created OnShore Compliance Solution intern web app.
- Created CI/CD pipelines for ETL jobs and Database schemas.
- Created RESTful API using ASP.NET Core with Clean Architecture.

- Created Monitoring alerts AML, Matrix, Analytics and Trade surveillance for costumers.
- **Skills:** Azure, Azure SQL and SQL Server, Azure Data Factory, Databricks, Python, PySpark.

### Leega Consultoria

(Remote) São Paulo, Br

Data Engineer

Jul 2020 - Aug 2020

- **Achievements:** Successfully created and deployed ETL stream/batch jobs in Dataflow with Apache Beam (Java and Python), processing over 100k/s events.
- Developed a solution with Cloud Pub/Sub, Cloud Spanner, Dataflow, and GCS that persist over 3 PB os data.
- **Key Responsibilities:** Created and deployed ETL stream/batch jobs in Dataflow with Apache Beam (Java and Python).
- Developed solutions with Cloud Pub/Sub, Cloud Spanner, Dataflow, and GCS.
- Source: <https://cloud.google.com/customers/cerc?hl=pt-br>
- **Skills:** GCP · GCS · Apache Beam · Cloud Spanner · Dataflow · Cloud Pub/Sub · Azure DevOps · Continuous Integration and Continuous Delivery (CI/CD) · Java · Python

### Qubo Tecnologia e Sistemas

Brasília, Br

Back-End Developer

Feb 2019 - Jul 2020

- Supporting pre-sales/technical team in data engineering and cloud issues;
- Supporting development team as a back-end developer;
- New management system to deal with new employees QHUB;
- RESTful API Development to connect with Quickbooks to QBOT;
- Development of AI technologies to [Qortex](#) project: (OCR e NLP).
- **Skills:** Python (main data stack libs), Django e Flask, PostgreSQL, MongoDB, Nginx, Docker, Git, Google Cloud(VISION api, AutoML e NLP api) e Digital Ocean.

## EDUCATION

### University of Brasilia

Brasília, Br

jun 2023 - Current

Master's in Applied Computing, Data science, Big Data, cloud computing and Artificial Intelligence

### University of Brasilia

Brasília, Br

Mar 2015 - Dez 2020

Bachelor of Electrical and Electronics Engineering with a focus on XAI techniques applied on convolutional neural network for skin lesion classification.

Graduated with honors from Universidade de Brasília in 2020. Completed a thesis on XAI techniques applied on convolutional neural network for skin lesion classification, which is available on [XAI techniques applied on convolutional neural network for skin lesion classification](#)

### George Washington University

Washington, D.C, EUA

Participated in the Innovative Leadership in a Transnational World exchange program at Georgetown University in 2014.

The program was divided into four themes: governance, civic engagement, development, and sustainability. Learned about the important economic and social issues facing our world. Gained a better understanding of what it means to be a thoughtful leader in a global society. [Innovative Leadership in a Transnational World](#)

Jun 2014 - Ago 2014

### Centro Interescolar de Línguas da Ceilândia - CILC

Brasília, Br

Specialization English language and Literature, General

Mar 2007 - Dez 2014

Completed an 8-year English language course with 1000 hours of classes. Received a certificate of completion. Developed advanced English language skills, including fluency in grammar, vocabulary, and pronunciation. Gained a deep understanding of British and American culture.

## ADVANCED COMPUTER SKILLS

Cloud Architecting:	GCP, AWS and Azure.
Data Engineering:	No-SQL, Apache Ecosystem, MongoDB, Relacional Databases.
Programming Language:	Python, Scala and SQL.
Frameworks:	FastAPI, Spark and PySpark, Python main libs, Airflow.
DevOps:	Azure Devops, GitHub, Agile, CI/CD, unit testing and clean code.

## CERTIFICATIONS AND AWARDS

---

Microsoft Certified: Azure Fundamentals	May 2020
Google Cloud Certified - Associate Cloud Engineer	Aug 2020 - Aug 2022
Microsoft Certified: Azure Administrator Associate (AZ-104)	Aug 2020 - Aug 2022
Microsoft Certified: Azure Data Engineer Associate	Nov 2020 - Nov 2023
Confluent Fundamentals for Apache Kafka	Feb 2021
Microsoft Certified: Azure AI Engineer Associate	Feb 2021 - Feb 2023
Microsoft Certified: DevOps Engineer Expert	Jul 2021 - Jul 2023
Astronomer Certification for Apache Airflow Fundamentals	May 2022
Academy Accreditation - Databricks Lakehouse Fundamentals	May 2022 - May 2023
AWS Certified Cloud Practitioner	Sep 2022 - Sep 2025
Google Cloud Certified - Professional Data Engineer	Jan 2020 - Nov 2024
Google Cloud Certified - Professional Cloud DevOps Engineer	Jan 2023 - Jan 2025