

THIAGO BAPTISTA

Vitória-ES | (27) 99925-6080 | thiagobaptista1981@outlook.com | linkedin.com/in/thiagob1981

PROFESSIONAL SUMMARY

Senior Data Engineer with 25+ years of experience designing, modeling, and building **Analytics**, **Big Data**, and **Data Lakehouse** solutions across various industries.

- In-depth knowledge of **Databricks**, **Spark**, and **Azure Cloud** for large-scale data processing, integration, and modeling.
- Strong background in **Data Warehousing**, **ETL/ELT**, pipeline orchestration, data governance, and performance optimization in both on-premises and cloud environments.
- Proven ability to lead complex data engineering initiatives, combining **Python**, **SQL**, and best-in-class architectural practices to leverage insights, Machine Learning, and Artificial Intelligence in high-impact projects.
- Excellent communication skills, working collaboratively in **multidisciplinary** teams to consistently deliver value and drive innovation.

WORK EXPERIENCE

Senior Data Engineer @ Prodesp (Almaviva Solutions)

12/2024 – Present (São Paulo – Brazil)

Working in GDAP (Digital Office of Public Administration), directly serving the Governor's Office of the State of São Paulo. Responsible for integrating, consolidating, and making state data available on a unified platform to support government decision-making, optimize public policies, and back strategic high-impact initiatives.

- **Databricks and Spark:** Built and optimized scalable data pipelines to integrate multiple sources using Data Lakehouse architectures to unify and democratize data access across the state government.
- **Azure Storage Account:** Managed data in a cloud environment, ensuring high availability, security, and performance for critical operations.
- **CI/CD with Azure DevOps and GIT:** Developed fully automated data pipelines for ingestion, transformation, and deployment, fostering robust governance and efficient version control.
- **SQL & Python/PySpark:** Created high-performance solutions to transform and model large-scale data, enabling the construction of Data Warehouses and the integration of structured and semi-structured data.
- **Power BI:** Designed semantic models and interactive dashboards focused on strategic use by government managers, enabling integrated views and near real-time analysis.
- **Kanban with Flow Metrics:** Employed agile practices for planning and tracking deliverables, ensuring predictability, efficiency, and tangible impact on operations.

Senior Database Administrator @ Prodesp (AlmavivA Solutions)

03/2024 – 12/2024 (São Paulo – Brazil)

Ensured the efficiency, integrity, and availability of critical databases supporting government programs focused on employability and economic development, directly assisting the State of São Paulo's Department of Economic Development.

- **Advanced SQL Server:** Developed and optimized views, complex queries, CTEs, stored procedures, scalable functions, triggers, and T-SQL scripts for systems like Qualifica-SP, Meu Emprego, and Via Rápida.
- **Systems Sustainability:** Worked across production, homologation, and development environments, performing proactive monitoring, implementing corrections through GMUD, and handling SLA-based tickets using IBM Maximo (SmartCloud Control Desk 2).
- **Process Automation:** Created Python scripts for data extraction, analysis, and integration, enhancing efficiency and eliminating repetitive manual tasks.
- **Collaboration with Agile Teams:** Participated in daily stand-ups and continuous alignment with development teams, promoting integrated and high-performance database solutions.
- **Data Integration and Governance:** Ensured integrity, security, and compliance in critical initiatives, focusing on relational modeling, query performance, monitoring, and solving complex issues.

Senior Data Engineer @ Elever Vision

04/2014 – 03/2024 (Vitória – Brazil)

For nearly 10 years, led Data Engineering initiatives for various clients—ranging from retail (supermarkets) and manufacturing (aluminum) to media and entertainment—focusing on the creation of data warehouses, ETL orchestration, and analytics in both on-premises and Azure cloud platforms.

- **Data Warehousing & Analytics:** Built architectures for Data Lakes and Data Warehouses using Azure Storage, Azure SQL Database, and on-premises (SQL Server, PostgreSQL) environments, enabling advanced, unified analyses for decision-making support.
- **ETL Pipeline Orchestration:** Developed data ingestion and transformation processes with Azure Data Factory, Apache Airflow, and SSIS, connecting multiple sources (including sales systems, ERPs, and digital marketing data via Google Tag Manager/Analytics).
- **Automation & Python Scripts:** Leveraged Python (pandas) for cleaning, enriching, and analyzing large datasets; created T-SQL and PL/pgSQL routines to optimize queries and stored procedures in SQL Server and PostgreSQL databases.
- **CI/CD & Data Governance:** Set up integration and continuous delivery pipelines using Azure DevOps and Git/GitHub, ensuring version control, automated testing, and governance throughout the data lifecycle.
- **Visualization & Agile Methodologies:** Developed Power BI dashboards to support business areas; adopted Kanban and JIRA for planning, execution, and progress tracking, focusing on efficiency and collaboration.

Data Engineer @ Vale (subsequently spun off to VLI)

03/2010 – 01/2014 (Belo Horizonte – Brazil)

Served as a Data Engineer and technical reference for a large-scale BI and Data Warehousing project involving railways and ports. Defined architectural standards, ingestion strategies, and data modeling for a multidisciplinary team of 15 professionals. The project consolidated data from multiple sources (SAP, Oracle ERP, Cognos, logistics operating systems, among others) to support FP&A (Financial Planning & Analysis) processes, focusing on demand forecasting, results calculation, and cost allocation.

- **Data Warehouse Construction in SQL Server:** Developed multidimensional modeling (fact and dimension tables), complex queries, stored procedures, and T-SQL functions, ensuring performance and reliability for critical analyses.
- **On-Premises ETL Orchestration:** Created extraction and transformation processes with SSIS, integrating data from internal and external railway and port systems, emphasizing automation and governance.
- **BI Solution Architecture:** Established standards for data ingestion, consolidation, and availability, enabling advanced financial analyses (including optimistic/pessimistic scenarios) for logistics capacity and budget forecasting.
- **Automation & VBA/Excel Advanced:** Created VBA tools for generating interactive dashboards, calculating results, and allocating costs, integrated with the Data Warehouse, optimizing workflow in commercial and planning areas.
- **Team Management & Project Methodologies:** Served as a technical leader for a multidisciplinary team, coordinating demands from various stakeholders (finance, commercial, operations) and monitoring timelines while adhering to best development and data quality practices.

Data Engineer @ Fundação Ceciliano Abel de Almeida

02/2000 – 03/2010 (Vitória – Brazil)

A nonprofit organization affiliated with the Federal University of Espírito Santo (UFES), active in project management, university extension, public exams, and graduate studies. In this role, I focused on creating a Business Intelligence layer over the Sapiens ERP, leveraging SQL Server to deliver advanced analytics and managerial reports in an on-premises environment.

- **BI Architecture & SQL Server:** Built and maintained data structures (tables, views, procedures, and T-SQL functions) for integrated analytics, ensuring performance and data security.
- **On-Premises Database Administration:** Capacity planning, implementing backup routines, performance monitoring, disaster recovery, and optimizations for high availability.
- **Process Automation with Visual Basic:** Created routines for data extraction, transformation, and loading (ETL) from the Sapiens ERP, minimizing manual interventions and reducing failure risks.
- **Data Modeling & Analysis:** Organized and normalized financial and operational data, enabling reliable reporting and strategic insights for various areas of the Foundation.

- **Interactive Reports with Crystal Reports:** Developed dashboards and control panels to measure KPIs, providing visibility into financial, managerial, and academic performance.

Developer @ Multi - Cia da Informação

10/1996 – 12/1999 (Vitória – Brazil)

Worked on the development of commercial applications and financial modules for multiple clients, using Visual Basic and relational databases (Access and SQL). Although the main focus was on transactional systems (accounts payable/receivable, billing, etc.), I gained valuable experience in practices now associated with Data Engineering.

- **Requirement Gathering:** Conducted client visits to identify functional needs and propose improvements to ERP modules, covering finance processes to commercial control.
- **Database Modeling & Architecture:** Defined database schemas, performed data normalization, and wrote SQL queries, ensuring integrity and efficiency in data storage and retrieval.
- **Commercial Application Development:** Implemented billing systems via bank slips for Internet Service Providers, in addition to accounts payable/receivable functionality and management reports.
- **Process Automation:** Built Visual Basic routines to automate repetitive tasks, streamlining workflows and reducing manual errors.
- **Support & Maintenance:** Fixed bugs, enhanced applications based on new business requirements, and facilitated go-live processes with end users.

EDUCATION

Bachelor of Computer Science @ Universidade Federal do Espírito Santo (UFES)

2004 – 2010 (Vitória – Brazil)

- **Strong academic foundation in:** Mathematics, Data Modeling, Algorithms, and Optimization, combining theory and practice to solve large-scale and complex problems.
- **Solid background in applied mathematics and quantitative methods:** Linear Algebra, Calculus I–IV, Probability & Statistics, Numerical Methods, Graph Theory, Queueing Theory, Linear Programming, Non-Linear Programming and Network Flow Optimization.
- **Extensive training in:** Algorithm Design, Data Structures, and Computational Complexity, with advanced coursework in: Algorithm Analysis, Search and Sorting Techniques, Formal Languages, Compiler Construction and Combinatorial Optimization.
- **Strong expertise in:** Data Systems and Distributed Computing, including: Database Systems, Parallel Computing, Computer Architecture, Operating Systems, Software Engineering, Project Management, Interfaces and Peripherals.
- **Additional training in:** Data Communication and Networking, Digital Logic, and Societal Impact of Computing.

CERTIFICATIONS

- **Databricks Certified Data Engineer Associate**
Databricks – Issued Jul 2025 · Expires Jul 2027 | [Verify](#)
- **Databricks Fundamentals Accreditation**
Databricks – Issued Jul 2025 · Expires Jul 2026 | [Verify](#)
- **Microsoft Certified: Fabric Data Engineer Associate**
Microsoft – Issued Jun 2025 · Expires Jun 2026
Credential ID: 4DA97B9D069C64D4 | Certification Number: 852DFE-25358B | [Verify](#)
- **Microsoft Certified: Azure Data Fundamentals**
Microsoft – Issued May 2025
Credential ID: E614C62FF1DD0B67 | Certification Number: 75741A-F85EXA | [Verify](#)
- **GitHub Foundations | [Verify](#)**
GitHub – Issued June 2025 · Expires Jun 2028
- **EF SET English Certificate – C1 Advanced**
EF Standard English Test – Issued Sep 2024 | [Verify](#)

COURSES

Datacamp - [Verify](#)

- Associate Data Engineer in SQL (29h - 2025)
- Associate Data Analyst in SQL (39h - 2025)
- SQL for Database Administrators (16h - 2025)
- SQL Fundamentals (26h - 2025)
- Understanding Data Topics (10h - 2025)
- dbt Fundamentals (10h - 2025)
- GitHub Foundations (9h - 2024)
- Financial Reporting in Power BI (18h - 2024)
- Data Analyst in Power BI (50h - 2023)
- Power BI Fundamentals (19h - 2023)

LANGUAGES

- **English:** Advanced
- **Portuguese:** Native