



SENIOR DATA ENGINEER · DATA PLATFORMS & AI INFRASTRUCTURE

Vitor Bordini Garbim

19+ years architecting petabyte-scale data platforms on AWS, now building governed agentic AI systems on Amazon Bedrock.

 [linkedin.com/in/vitorgarbim](https://www.linkedin.com/in/vitorgarbim)  vitorgarbim.me

PROFESSIONAL SUMMARY

Multilingual (English, Spanish, Portuguese) Senior Data Engineer with **19+ years** designing scalable, high-impact data platforms across the insurance, financial, and technology sectors. Proven ability to lead end-to-end architecture of data lakes, real-time and batch pipelines, and governance frameworks on the AWS ecosystem. Currently architecting enterprise agentic capabilities on Amazon Bedrock: RAG knowledge bases, Neptune knowledge graphs, MCP tooling, and hybrid deterministic and LLM automation.

Built a **70+ PB** centralized enterprise data lake certified for Red Security, enabling governed analytics for **6,000+ users**. Delivered real-time and batch ingestion using AWS Glue, Redshift, Kinesis, Airflow, and Lake Formation, and defined self-service governance with DataZone across hundreds of domains.

At Amazon, led large-scale initiatives across HR, Risk (P&C insurance), and central Operations, including the design of **Veritas**, a claims and actuarial data lake supporting reserve modeling and ingestion from third-party providers such as Sedgwick. Contributed to the open-source AWS Insurance Lake project, shaping cross-platform standards for insurance data.

Known for mentoring senior engineers and cross-disciplinary peers (data engineers, BIEs, and data scientists) while driving secure, cost-efficient data architecture through modern design patterns and operational best practices.

BEYOND WORK: HOME LAB AS PROVING GROUND

The same enterprise architecture runs at home on owned hardware, rebuilt in open source.

Platform: multi-node Proxmox cluster under GitOps running Airflow, Trino, Superset, MinIO, and PostgreSQL with graph and vector extensions, with full observability.

Agents: a fleet of cooperating agents on an agent-to-agent protocol and NATS event bus, calling in-house MCP servers that bridge hundreds of infrastructure and product APIs.

Governance: per-agent OAuth/OIDC grants from a control portal, human approval gates on destructive actions, and a private citation-backed RAG knowledge base for memory.

PROFESSIONAL EXPERIENCE

Senior Data Engineer

Amazon, OTS DataTech

May 2019 – Present

Austin, TX / Miami, FL

- Built a governed **RAG knowledge base** on Amazon Bedrock Knowledge Bases for document and code corpora, with ingestion, chunking, and retrieval pipelines delivering grounded, retrieval-backed answers over enterprise data assets.
- Designed a custom **SQL reverse engineering pipeline** that tokenizes SQL artifacts to capture tables, columns, and derived metrics, enriching each node with business and lineage definitions and materializing a transversal property graph in Amazon Neptune backed by OpenSearch semantic search.
- Developed **MCP (Model Context Protocol) tooling** on Amazon Bedrock Gateway, exposing graph queries, catalog and lineage search, and pipeline health APIs as secure, authenticated tools that any LLM agent can call.
- Stood up the **data mesh metadata plane** with AWS DataZone and an OpenLineage proxy, capturing table-level lineage, dataset freshness, and job metrics from Airflow and Glue ETL per domain and streaming them into the knowledge base for continuous enrichment.
- Designing **agentic automation** on Amazon Bedrock Runtime (in flight), combining deterministic prompt schemas, structured outputs, and evaluation and approval gates with LLM reasoning; the flagship agent explains table and pipeline metrics in plain language for internal business stakeholders, sourced from the OpenLineage metadata stream.
- Led the architecture and delivery of a **70+ PB centralized data lake** for Amazon Operations using AWS Glue, Redshift, Lake Formation, and Athena (Presto). Replaced a legacy ETL model, achieved Red Security certification, and **cut storage and compute costs by 40%**.
- Built **Veritas**, a Property & Casualty insurance data lake managing third-party claims ingestion (e.g. Sedgwick) and reserve modeling for Global Risk Management. Joined an open-source away team with AWS to help define **Insurance Lake**, standardizing cross-platform P&C data models.
- Defined source-of-truth governance for business-critical datasets, metrics, and dashboards across hundreds of teams, using AWS DataZone and SageMaker Studio to deliver a federated catalog for self-service analytics at scale.
- Built and maintained real-time and batch ETL pipelines with Airflow, Glue Streaming, Lambda, and Kinesis, integrating data from RDS, DynamoDB, OpenSearch, and other upstream systems.
- Drove data-product operations, defining contracts, aligning roadmaps, and tracking KPIs with product managers and domain leads in Operations and Risk.
- Mentored senior engineers and cross-functional peers (data engineers, BIEs, data scientists), elevating technical standards and operational best practices across Amazon's data community.

BUSINESS DOMAINS

Career Choice (HR): data infrastructure for a company-wide upskilling initiative for warehouse employees.

GRMC (Insurance & Risk): domain-specific P&C pipelines and actuarial support.

OTS DataTech (Central Operations): reusable, governed infrastructure for analytics at enterprise scale.

Senior Project Manager & Data Strategy

Jul 2012 – May 2019

AIG

Miami / Ft. Lauderdale & São Paulo (LATAM, Europe, US)

- Led enterprise-wide P&C insurance transformation across **7+ countries** in Latin America, Europe, and the U.S., focused on risk, claims, and underwriting operations.

- Designed and implemented a future-state data warehouse for the AIG Shared Services Center, unifying system interfaces and manual inputs to improve data quality, traceability, and automated reporting across **1,400+ reports**.
- Developed insurance analytics and KPI dashboards for HR, Finance, Operations, and Claims using Tableau, QlikView, SQL, Excel, and MS Access, enabling leadership to track performance and claims cycle times.
- Streamlined underwriting and claims workflows across Europe with productivity tools and automation, significantly reducing policy-issuance cycle times and manual workload during peak demand.
- As Strategic Planning Coordinator, built global dashboards for LAC and U.S. regions, delivering unified HR, Operations, Customer, and Claims KPIs for executive decision-making. As System Architect, integrated multiple insurance systems and built ETL/ODS processes using SAS, Oracle, and Microsoft tools.
- Implemented Lean Six Sigma methods and trained regional executives and managers on structured problem-solving, SOPs, and continuous improvement.

* Earlier roles and full project history are on LinkedIn: [linkedin.com/in/vitorgarbim](https://www.linkedin.com/in/vitorgarbim)

EDUCATION

B.S. in Computer Information Systems, Minor in Business

Jan 2008 – Dec 2012

Universidade Presbiteriana Mackenzie, São Paulo, Brazil

Computer Science, 3.5 years of coursework completed

Universidade Presbiteriana Mackenzie, São Paulo, Brazil. Transferred to Computer Information Systems.

Technologist Diploma in Systems Development (2-year program)

FECAP, São Paulo, Brazil

TECHNICAL SKILLS

AWS

Glue (Spark)

Redshift

Lake Formation

Athena (Presto)

EMR (Hadoop)

SageMaker

Lambda

CDK

ECS Fargate

S3

DataZone

AI & AGENTIC

Bedrock Knowledge Bases

Bedrock Gateway

Bedrock Runtime

Kendra

Neptune

OpenSearch

RAG

Knowledge Graphs

MCP

Agent-to-Agent Protocols

LLM Evaluation Gates

Structured Outputs

REAL-TIME & BATCH

Kinesis

Glue Streaming

Airflow

DynamoDB Streams

NATS

OpenLineage

PROGRAMMING

Python

TypeScript

JavaScript

SQL

PySpark

WEB & BACKEND

Flask

FastAPI

Node.js

ReactJS

HTML

CSS

