



Oguzhan

Senior Data Engineer

Summary

Senior Data Engineer with 6+ years of experience delivering scalable, production-grade data platforms across financial services, mobility, telecom, and real estate. Strong expertise in Python, SQL, Apache Spark, PySpark, Databricks, Kafka, Flink, Airflow, dbt, AWS, Azure, ETL/ELT pipelines, Data Mesh, DataOps, and enterprise data modeling.

Experienced in building high-throughput streaming and batch pipelines, regulatory and sustainability reporting systems, data quality and governance frameworks, and self-service analytics capabilities. Hands-on experience with GenAI/LLM, RAG, vector embeddings, prompt engineering, and AI-augmented engineering, with end-to-end ownership across architecture, data modeling, performance optimization, and delivery.

TECHNICAL SKILLS:

- Cloud & Infrastructure: AWS (S3, EC2, Redshift), Azure (Data Factory, SQL DB), Docker, CI/CD, Git
- Data Processing: Apache Spark, PySpark, Databricks, Hadoop, HiveQL
- Streaming: Apache Kafka, Apache Flink
- Orchestration & Transformation: Apache Airflow, dbt, Oracle Data Integrator, ETL/ELT Pipelines
- Data Modeling: SAP Power Designer, Data Vault, Dimensional Modeling
- AI & ML: GenAI/LLM, RAG, Prompt Engineering, Vector Embeddings
- Data Strategy & Governance: Data Mesh, DataOps, Data Governance, Data Quality, Data Lineage
- Regulatory: PCAF, PACTA, TCMB, BDDK, ECB, GDPR
- Programming: Python, SQL, Bash

Professional Career

Senior Data Engineer

January 2024 – Present | Ankara, Turkey

Project Info:

Company provides organizations with technology infrastructure, software development on different platforms, internet applications, integration, system management, security management, project management and office application services

Responsibilities:

- Architected and fully automated PCAF and PACTA sustainability reporting pipelines for BBVA from scratch - covering loan portfolio carbon assessment and climate transition analysis - delivering standardized CSV reports to the holding with consistent on-time regulatory compliance
- Built reusable, parameterized ETL workflows using Oracle Data Integrator serving 10+ financial and regulatory reporting domains - including credit cards, consumer credits, banking operations, legal, accounting, TCMB, BDDK, BBVA, and ECB - enabling domain teams to run reports independently and eliminating recurring ad-hoc data engineering bottlenecks
- Designed and built an AI-powered metric discovery platform using LLM reasoning over vector-embedded table metadata — enabling non-technical stakeholders to instantly find existing metrics or identify exact source fields to build new ones, transforming schema discovery into a self-serve capability
- Led the first PySpark ETL proof-of-concept for BBVA sustainability reports at Garanti Teknoloji, benchmarking performance across multiple cluster and partitioning configurations and producing a quantified migration roadmap that makes the technical and business case for replacing legacy Oracle batch infrastructure with Spark-native pipelines
- Owned enterprise data modeling for the CDD team in SAP Power Designer - designing Data Vault and dimensional models across 10+ analytics domains, enforcing naming standards and model consistency to provide full historical auditability for TCMB, BDDK, and ECB submissions
- Owned the definition and tracking of automated Data Quality rules with real-time threshold alerting and SLA monitoring across regulatory pipelines - reducing data quality incidents by ~70% and eliminating late-stage corrections on regulatory reports
- Served as primary data engineering owner for BBVA's sustainability reporting obligations within Garanti Teknoloji, coordinating with holding-level data teams across Turkey and Spain to align on schema standards, delivery SLAs, and evolving regulatory report formats
- Mentored junior data engineers on regulatory data engineering practices, ODI pipeline design, and data modeling standards - accelerating onboarding and raising team capability in delivering compliant, audit-ready reports across the CDD domain

Software and technologies:

Python, SQL, PySpark, Apache Spark, Oracle Data Integrator, SAP Power Designer, Data Vault, Dimensional Modeling, GenAI/LLM, Vector Embeddings, Data Quality, PCAF, PACTA, TCMB, BDDK, ECB

Senior Data Engineer

January 2022 – January 2024 | Hamburg, Germany

Project Info:

Data platform engineering, supporting marketing, sales, finance, product, and performance teams. The work included Data Mesh migration, real-time pricing, DataOps and governance, scalable ELT pipelines, geospatial analytics, and AI-augmented engineering

Responsibilities:

- Partnered with teams across marketing, sales, finance, product, and performance to design and deliver data pipelines they needed - integrating internal systems and external sources into the data warehouse and enabling self-serve analytics across 8+ consuming teams
- Architected domain-oriented data products for company-wide Data Mesh migration, decoupling payments and pricing domains from a centralized monolith into independently owned AWS-based federated products - reducing cross-team data dependencies by 40%
- Engineered a real-time dynamic pricing pipeline processing 500K+ events/sec using Kafka and Flink with Python, achieving sub-100ms end-to-end latency across 27 European cities and directly powering pricing algorithms that improved revenue optimization by 8%
- Implemented a DataOps and governance framework using Great Expectations and dbt across 50+ production pipelines, introducing automated quality validation, data lineage tracking, and GDPR compliance controls - reducing data incidents by 65% and cutting mean time to detection from hours to minutes
- Adopted early AI-augmented engineering practices - leveraging GitHub Copilot and GPT-based tooling for PySpark code generation, test scaffolding, and documentation - reducing development time by 40% and improving code review turnaround across the team
- Built scalable ELT pipelines on Databricks PySpark (AWS) integrating Salesforce, Zendesk, HubSpot, Braze, AppAnnie, and Mixpanel into S3 Data Lake and Redshift, replacing 6 third-party connectors with custom Python operators and Kafka CDC - saving \$70K/year in data transfer costs
- Built a geospatial analytics platform integrating OpenStreetMap data across all 27 cities, generating vector embeddings for location intelligence and powering ML-driven demand forecasting models - improving supply- demand prediction accuracy and enabling more precise driver dispatching across urban mobility markets

- Served as technical contributor on company's Data Architecture Committee, evaluating modern data stack tooling, defining pipeline optimization standards, and advocating for AI-assisted development and observability practices adopted organization-wide
- Mentored junior data engineers on Data Mesh principles, PySpark development patterns, and DataOps best practices - supporting knowledge transfer across the data platform team and contributing to a culture of engineering excellence

Software and technologies:

Python, PySpark, Databricks, AWS, Amazon S3, Amazon Redshift, Apache Kafka, Apache Flink, dbt, Great Expectations, Data Mesh, DataOps, GDPR, Vector Embeddings, GitHub Copilot

Data Engineer

January 2021 – January 2022 | Istanbul, Turkey

Project Info:

Big Data engineering for company supporting analytics for a platform with 50M+ monthly active users. The work focused on scalable ETL pipelines, data quality, anomaly detection, performance optimization, and operational reliability

Responsibilities:

- Architected scalable Big Data solutions for company - 50M+ monthly active users - implementing ML- based data quality validation and automated anomaly detection to ensure platform analytics reliability at global scale
- Designed and automated end-to-end ETL pipelines across ODS-to-ADS layers using Spark, HiveQL, Hadoop, and Airflow, applying DataOps monitoring and validation frameworks that achieved 99.8% operational efficiency across the analytics stack
- Delivered a transformative performance optimization by rewriting a legacy HiveQL profitability pipeline for gaming apps to Spark, reducing runtime from 8 hours to 1 hour (87.5% improvement) and enabling next-day regional decision- making
- Established data quality frameworks with automated validation rules and real-time monitoring dashboards, enforcing data contracts between pipeline stages and reducing incidents by 45%
- Performed systematic performance profiling and root cause analysis across multi-region analytics pipelines using automated query debugging tooling, resolving high-severity data integrity issues and institutionalizing postmortem processes that prevented recurrence - improving overall pipeline reliability across the AppGallery analytics stack

Software and technologies:

Apache Spark, HiveQL, Hadoop, Apache Airflow, ETL, DataOps, Data Quality, ML-based Validation, Anomaly Detection

Data Engineer

June 2020 – November 2020 | Amsterdam, Netherlands

Project Info:

Data engineering for real estate clients, integrating multiple data sources into Azure-based data lake and data warehouse environments. The work included ETL architecture, API integrations, automation, and technical requirements definition

Responsibilities:

- Architected ETL pipelines using Python, SQL, and Azure Data Factory integrating multiple real estate data sources into a schema-driven Azure SQL Data Lake and DWH, improving scalability by 40% and query performance by 3x for clients managing 10TB+ datasets
- Automated API integrations and data workflows via Azure Data Factory with SQL stored procedures and Python orchestration, eliminating 80% of manual intervention and delivering €15K in annual operational cost savings
- Led technical requirements and roadmap definition for Northern Europe's largest real estate clients, translating business analytics needs into data lake and DWH architecture decisions - improving project delivery speed by 30%

Software and technologies:

Python, SQL, Azure Data Factory, Azure SQL, ETL, Data Lake, Data Warehouse, REST APIs, Stored Procedures

Education

Izmir Institute of Technology

Bachelor's Degree, Electrical, Electronics and Communications Engineering

Foreign Languages

English: Fluent

Turkish: Native