

Hiring software engineer as easy as calling a taxi.

Nikita

Data Engineer

SUMMARY

A seasoned Data Engineer with over 6 years of experience in the field of software and big data engineering. Holds a strong academic background in Computer Science and Software Engineering, certified as a Google Cloud Professional Data Engineer. Demonstrates deep expertise in high-load system design, performance optimizations, and domain-specific solutions for Healthcare, Fintech, and E-commerce. Proficient in Python and SQL, with significant exposure to data engineering tools such as Apache Hadoop, Apache Spark, and Apache Airflow, and cloud technologies from AWS and GCP. Adept at working with various databases and message brokers, excelling in data modeling, BI, and data visualization using tools like Looker, Power BI, and Tableau. Enhanced system efficiencies through SQL and data pipeline optimizations, driving significant improvements in processing speed and system performance. A collaborative engineer with a strong grasp of DevOps practices, committed to best-in-class data governance and security standards.

SKILLS

Main Technical Skills	Python (6 yr.), SQL (6 yr.), Apache Airflow (5 yr.), JMeter (6 yr.), PySpark (6 yr.)
AI & Machine Learning	AWS ML (Amazon Machine learning services)
Programming Languages	Python (6 yr.)
Salesforce Ecosystem	Apex DataLoader
Python Libraries and Tools	PySpark (6 yr.)
Java Libraries and Tools	Spring Data
Data Analysis and Visualization Technologies	Apache Airflow (5 yr.), Apache Spark Streaming, Data Analysis, Data Quality, Data visualization, ELT
Databases & Management Systems / ORM	Apache Spark Streaming, Data Warehousing, SQL (6 yr.)
Amazon Web Services	AWS Kinesis, AWS ML (Amazon Machine learning services), AWS Security Groups

Azure Cloud Services	Microsoft Azure API
Deployment, CI/CD & Administration	CI/CD
QA, Test Automation, Security	JMeter (6 yr.)
SDK / API and Integrations	Microsoft Azure API
Other Technical Skills	3D Modelling, Cloud solutions, Code Review, Data backups, Illustration and icons, Lakehouse, Privacy Compliance & Data Governance

WORK EXPERIENCE

Data Engineer, Data Mesh Implementation for Healthcare Data

Duration: 08.2021 - till now

Summary:

- A data-driven startup leveraging the concept of data mesh to process and transform massive healthcare data
- Promoting data democratization within the organization

Responsibilities: Implement secure storage solutions, collect and preprocess data for ML models, develop data integrity checks, configure Kafka brokers, develop data breach response procedures, establish audit trail mechanisms, create data conduits with Kafka, configure AWS DMS, optimize Kafka configurations, enforce data integrity constraints, optimize SQL queries and workflows, integrate Amazon MQ, implement security controls, develop logical data models, integrate Spark with AWS Glue, monitor data quality, collaborate with data scientists, architect data streaming pipelines, maintain Tableau dashboards, configure Pub/Sub access controls, set up monitoring and logging for pipelines, utilize CDKTF for Terraform configurations, create Bash scripts, engage in GitHub Actions YAML file configuration, design CI/CD pipelines, code review.

Technologies: Python, SQL, Apache Spark, PySpark, Apache Airflow, AWS, Kafka, Redis, Oracle, Pandas, NumPy, Tableau, Bash scripting, CI/CD, Docker, Docker Compose, Kubernetes, GitHub Actions, GitHub

Data Engineer, Financial Data Management and Analytics Platform

Duration: 07.2019 - 07.2021

Summary:

- Revolutionizing financial data management and analysis with a unified platform that brings together disparate data sources and facilitates advanced analytics
- A combination of traditional DWH and Data Lake features for overcoming data silos, incomplete insights, and inefficient analysis methods



Responsibilities: Implement data governance policies, patch and update GCP Cloud SQL databases, design real-time streaming data pipelines with Apache Spark Streaming, normalize data models, identify data quality issues, migrate data with GCP Dataflow, build DWH on GCP BigQuery, integrate data with Google Cloud BigTable, manage workflows with Airflow, monitor Tableau dashboards, fine-tune ML models with Scikit-learn, prepare datasets with LookML, create dashboards with Looker, maintain metadata repositories, clean data with GCP Dataprep, orchestrate workflows with GCP Composer, transfer data with GCP Transfer Service, develop CI/CD with Jenkins, code review and refactoring.

Technologies: Python, SQL, Apache Airflow, Apache Spark, PySpark, GCP, Redis, MongoDB, PostgreSQL, Pandas, NumPy, Tableau, Matplotlib, Scikit-learn, Jenkins, CI/CD, Bash scripting, Docker, Docker Compose, GitHub

Data Engineer, Sales Analysis and Business Performance Improvement

Duration: 12.2018 - 07.2019

Summary:

- Sales analysis project to improve business performance through insights into customer behavior by utilizing data visualization and statistical analysis to identify trends and opportunities for growth
- Optimized pricing, promotions, and marketing campaigns

Responsibilities: Develop AWS Lambda for logic execution, maintain Kafka configurations, manage Amazon RDS databases, process large-scale data with AWS Databricks, monitor Kafka clusters, implement data ingestion pipelines, integrate Spark Streaming, perform data analysis with Databricks notebooks, design ETL pipelines, monitor Lambda functions, implement data models in Power BI, provision EC2 instances, analyze data with Apache Hive, follow DWH concepts, code review.

Technologies: Python, SQL, Kafka, Apache Spark, PySpark, Apache Hadoop, AWS, MS SQL, Redis, MongoDB, Power BI, Pandas, NumPy, Bash scripting, CI/CD, Docker, Docker Compose, Kubernetes, Bitbucket

Data Engineer, E-commerce Platform for Healthy Lifestyle Products

Duration: 01.2018 - 12.2018

Summary: E-commerce platform for a healthy lifestyle online store, providing product selection, ordering, delivery, and advice in sports and nutrition, with a focus on convenience and customer engagement.

Responsibilities: Provide cloud solutions with AWS, resolve RabbitMQ issues, secure Spark Streaming applications, automate data processes in Power BI, develop RabbitMQ messaging components, configure EC2 performance, orchestrate containers with Kubernetes, implement AWS Lambda serverless parts, write SQL queries, monitor RabbitMQ clusters, implement data backup and recovery strategies, troubleshoot Apache Spark, assist with data modeling, optimize SQL queries.

Technologies: Python, SQL, RabbitMQ, Apache Spark, PySpark, AWS, MongoDB, PostgreSQL, Power BI, Pandas, NumPy, Docker, Docker Compose, Kubernetes, CI/CD, Bash scripting, GitLab

EDUCATION

- **Computer Science and Software Engineering**



CERTIFICATION

- **Google Cloud Certified – Professional Data Engineer**
2023

