**Joshua Dev**

(413)463-7878 | devjoshua75@gmail.com

**Summary**

Python Developer with 7+ experience in developing and optimizing ETL processes for operational and analytical applications. Demonstrates expertise in constructing high-performance data pipelines using Apache Airflow, Spark, and Kafka to streamline data processing tasks. Partnered with cross-functional teams for successful implementation of complex data architectures and workflows. Eager to apply data engineering expertise to enhance data infrastructure and analytics capabilities in dynamic organizational settings.

**Skills**

**Technical Skills:** Data Analysis, Data Engineering, Data Mining, Data Warehousing, Regression Analysis, Data Integration, Data Modeling, Data Mapping, ETL Design, MapReduce, Databases, Machine Learning, Statistics, Tableau, Oracle, UNIX, Stata, Strategic Planning, SQL, Power BI, Python 3.6.3.8, Java, Net, Spark SQL, Groovy, GoLang Spark, Scala, Kafka, SQL Server, Oracle 9i/10g/11g, MySQL, PostgreSQL, SQLite, MongoDB, Web Sphere, Apache Tomcat, Django, Flask, PySpark, AJAX, HTML5, Node JS, CSS3, Jupyter Notebook, JavaScript, jQuery, Web Services (Soap, Restful), Amazon Webservices like EC2, S3, EBS, PyUnit, ETL, Kubernetes, React JS, Golang, JIRA, Rally, MS Azure, Heroku, Bootstrap, CherryPy, web2py, Pyramid, Robot, Selenium, SVN, CVS, Git, Gitlab, Agile, SDLC, Waterfall model, AWS Redshift, Mulesoft, SnapLogic

**Interpersonal Skills:** Strong Internet Research, Excellent Verbal Skills, Excellent Written Skills, Excellent Quantitative Skills, Time Management and Prioritization Abilities, Effective Presentation, Team Player, Leadership, Negotiation Skills

**Industry Knowledge:** Financial Services, Healthcare Services

**Work Experience**

**CVS Health** Chicago

Python Developer Oct 2022 - Present

* Developed robust microservices architecture for scalable backend operations, ensuring efficient data processing.
* Processed transactions per second using Kafka to enhance throughput, though not claimed here.
* Created RESTful APIs using FastAPI to facilitate seamless interoperability between services.
* Containerized applications with Docker and optimized CI/CD pipeline configuration for improved deployment.
* Utilized Kubernetes for application deployment and managed data solutions on Azure cloud infrastructure.
* Handled Azure Service Bus for consistent and efficient application communication.
* Implemented Python scripts for data transformations, enhancing data profiling and processing for analytics.
* Monitored systems using Grafana, ensuring performance standards and enabling swift troubleshooting.
* Streamlined ETL processes in Snowflake, bolstering data retrieval and supporting analytics.
* Employed SQL Server tools for querying and optimization to meet application requirements.
* Leveraged platforms like Apache Airflow and Jenkins to automate data workflows and deployment cycles.
* Enhanced system monitoring using Jaeger UI for service tracing, providing deeper insights into system dynamics.
* Utilized Docker to maintain environment consistency and improve process efficiency.
* Collaborated with data teams using AJAX to tailor customer engagement strategies, boosting satisfaction.
* Engaged with SQL and Python for transforming complex requirements into robust data solutions.

**TCS – Google** India

Python Developer March 2021 - Aug 2022

* Collaborated on developing AWS-based Data Engineering and Analytics solutions to improve data access.
* Crafted and maintained ETL processes ensuring optimal data flows and alignment with business goals.
* Designed data models for storage and retrieval, adhering to business and functional specifications.
* Analyzed data system elements and facilitated data modeling to enable robust ETL pipeline workflows.
* Implemented end-to-end **real-time data ingestion pipelines** using **Cloud Pub/Sub** and Dataflow for event-based architecture.
* Employed Pandas and NumPy for data quality improvement through feature engineering and analysis.
* Constructed data pipelines using Apache Airflow and Spark SQL, advancing data processing capabilities.
* Created data visualizations in Python to facilitate strategic business decision-making processes.
* Enabled real-time analytics by feeding streaming data into **BigQuery** and **Looker Studio** dashboards.
* Defined robust ETL frameworks using tools like Airflow, amplifying data processing workflow efficiency.
* Built scalable data exchanges via REST APIs, emphasizing JSON structuring for universal accessibility.
* Developed solutions using Python and Spark, integrating diverse data sources for streamlined processing.
* Designed and orchestrated complex data workflows in **Cloud Composer**, automating ETL tasks across multiple environments.
* Automated deployment of GCP infrastructure and data pipelines using **Terraform** and **Cloud Build**, reducing manual intervention.
* Designed and optimized data warehouses using **BigQuery**, reducing query cost by 40% through partitioning and clustering.
* Assembled complex datasets for business rule execution, meeting functional data requirement thresholds.
* Lead initiatives for efficient data integration and utilized analytics for strategic opportunities.
* Integrated DevOps tools such as GitHub and Kubernetes, ameliorating deployment and collaboration efficiency.

**Xerox – Septa, Banglore India** India

Python Developer April 2018 – Mar 2021

* Collaborated with cross-functional teams to design and develop Hadoop-based applications, using tools like HDFS, MapReduce, and Hive.
* Processed and analyzed large datasets with Hadoop ecosystem tools, while actively adopting new technologies to meet project needs.
* Spearheaded the integration of emerging technologies to enhance the efficiency and effectiveness of Hadoop-based applications.
* Assisted in developing strategies for data optimization, resulting in streamlined processes and improved performance outcomes.
* Built feature engineering pipelines in **Vertex AI Pipelines**, with source data from BigQuery and GCS.
* Used XML for dynamic display of options in select box and description on web page, enhancing user interaction and interface flexibility
* Enabled self-service analytics by exposing curated datasets in **BigQuery** to **Looker** and **Data Studio**.
* Wrote sub-queries, stored procedures, triggers, cursors, and functions on MySQL, PL/SQL, and PostgreSQL databases, improving database efficiency and reliability
* Established database connections for Python by configuring packages like JDBC and MySQL-Python, enhancing data processing capabilities
* Resolved user problems in a timely and accurate manner as a member of the End-User support team.
* Created environment-specific settings for new deployments and updated deployment-specific conditional checks in the codebase, ensuring smooth and error-free software releases
* Implemented RESTful web services for sending and receiving data between multiple systems, facilitating seamless data integration and communication
* Optimized batch processing workloads using **Flex Templates in Dataflow**, enabling parameterized runs and better resource utilization.
* Handled client-side validation using jQuery, improving user input accuracy and reducing errors
* Sent text data through US sockets using various APIs and JSON, enhancing data transmission efficiency
* Prepared and analyzed reports using Python libraries such as Pandas and NumPy, contributing to informed decision-making and successful environment setup
* Used Python and Django to interface with the jQuery UI and manage the storage and deletion of content.

**Education**

**Clark University, Worcester, MA**

Master of Science, Computer Science May 2024

**Projects**

**Project on Stock Market Prediction using Python** Jan 2022

* Developed predictive systems employing Random Forest, Logistic Regression, and SVM algorithms to ensure highly accurate predictions.
* Applied advanced machine learning to analyze stock market trends, conducting a comprehensive comparison of algorithm performance.

**Certification**

**Microsoft Certificate of Azure Data Fundamentals:**

**Stanford Online Certificate of Cryptography 1:**