Kunj Patel . Jarvis Consulting

With over two years of experience as a results-driven Data Engineer and Developer, I bring a strong academic background and a proven track record in the field. Armed with a Master's degree in Applied Computer Science from Dalhousie University, I have developed a deep understanding of data engineering principles and techniques. Specializing in designing and implementing ETL pipelines and optimizing database schema and SQL queries, I have consistently delivered improved query performance and enhanced data accuracy. Proficient in Airflow, DBT, and Snowflake, I have successfully processed millions of records in real-time, ensuring data integrity and quality. My expertise extends to programming languages such as Python, Java, and SQL, as well as cloud technologies like AWS and Google Cloud. Committed to agile development methodologies and armed with strong problem-solving skills, I am dedicated to delivering high-quality solutions that align with business needs.

Skills

Proficient: Java, Python, SQL, Linux/Bash, Docker, RDBMS, Git, Amazon Web Services, Agile/Scrum

Competent: PySpark, Pandas, JavaScript, Node, Snowflake, Hadoop, Databricks

Familiar: Spark, Scala, Kubernetes, Google Cloud Platform, React

Jarvis Projects

Project source code: https://github.com/Jarvis-Consulting-Group/jarvis_data_eng-KunjPatel03

Cluster Monitor [GitHub]: A user-friendly solution to efficiently manage a 10-server cluster. The solution allows server users to easily monitor hardware and usage information, ensuring smooth operation and maintenance. Implemented with bash scripts, it seamlessly integrates into existing workflows. PostgreSQL (psql) is chosen as the database technology for reliable data storage. The project is implemented within a Docker container for consistent execution and compatibility. Using Git version control, development progress is tracked. This monitoring solution empowers users to optimize resource allocation, enhancing stability and efficiency of the server cluster for seamless operations.

Java Applications [GitHub]:

- JDBC App: The JDBC application is a robust system for performing essential CRUD operations (Create, Read, Update, Delete) on a database, utilizing the powerful JDBC API. It integrates a PostgreSQL database within a Docker container environment, with Maven as the package management and dependency resolution tool, and Git as the chosen version control system.
- Grep App: Grep is a Java-based application offering two efficient approaches for pattern searching within text files. The Stream-Based Approach utilizes Java Streams and functional programming to process large datasets with minimal memory usage, ideal for handling massive files. The Traditional Approach uses conventional file I/O and looping for moderate-sized files. Comprehensive unit testing with JUnit and Mockito ensures application correctness. Dockerizing the application facilitates easy distribution and deployment.

Springboot App [GitHub]: Designed and developed a comprehensive retail stock trading platform for individual investors using Spring Boot technology. Leveraged Spring Data Access Object (DAO) to efficiently manage and persist trader data within a PostgreSQL database. Implemented RESTful web APIs to seamlessly deliver real-time stock quotes and trading functionalities to end-users. Ensured the application's reliability and robustness through rigorous testing methodologies, including integration tests and unit tests utilizing JUnit and Mockito. Additionally, performed extensive manual testing of endpoints using the user-friendly Swagger UI. Enhanced deployment efficiency by containerizing the application using Docker, streamlining the deployment process and ensuring consistent performance across various environments.

Python Data Analytics [GitHub]: Utilized advanced data analytics techniques to analyze a company's transactional data and gain insights into customer shopping behavior, contributing to the development of targeted marketing strategies. Employing Python as the primary programming language, I harnessed the power of Pandas, Numpy, Matplotlib, and SQLAlchemy libraries to perform comprehensive data analysis. The entire analysis was conducted within a Jupyter notebook environment, where I seamlessly accessed data from both CSV files and a PostgreSQL database, sourced from the company's data warehouse. Operating on a Jupyter server, the notebook generated detailed visualizations, including plots illustrating monthly sales trends, order placements, customer counts, and, significantly, Recency, Frequency, and Monetary segmentation of the customer base. This project showcased my proficiency in data analysis, software tools, and database integration, enabling data-driven insights to drive strategic decision-making.

Hadoop [GitHub]: Actively contributed to an ongoing Python data analytics project centered on optimizing core Hadoop components, including HDFS, MapReduce, and YARN, to efficiently process and analyze substantial datasets. This initiative provided exposure to enhancing Hadoop's performance within a distributed ecosystem. Leveraging Apache Hive and Zeppelin Notebook on the Google Cloud Platform, the project addressed real-world business challenges by executing powerful HQL statements for data exploration. Additionally, responsibilities encompassed strategic setup of a master node and two worker nodes, forming a robust Hadoop cluster to efficiently manage datasets exceeding 21 million records. Proficiency was gained in utilizing cutting-edge technologies such as Hadoop, Hive, YARN, and MapReduce, enhancing skills in optimizing data storage, processing, and analysis.

Spark [GitHub]: Spearheaded the development of a data analytics solution targeting revenue growth for a retailer. Leveraging Apache Spark, critically evaluated Databricks (on Azure) and Zeppelin (on Hadoop) for optimum data strategy. Executed in-depth analytics in both environments, calculating Recency, Frequency, and Monetary Value (RFM) scores for tailored marketing, and dissecting monthly sales data for revenue insights. Differentiated new and existing users for comprehensive customer behavior understanding. Crafted efficient architectures for Databricks (workspace, DBFS, PySpark, Scala) and Zeppelin (notebooks, Hadoop, Hive Metastore, PySpark), optimizing data processing. Envisioned future enhancements, including predictive modeling, personalized recommendations, and real-time analytics, aimed at boosting marketing and customer satisfaction for sustained revenue growth. Demonstrated prowess in data engineering, analytics, and innovative technology application to drive business success.

Highlighted Projects

Relational Database Management System [GitHub]: Developed a robust RDBMS from scratch using Java, providing support for SQL queries, CRUD operations, and transactions. Implemented a sophisticated parsing and validation mechanism utilizing complex regex expressions for accurate data processing. Designed a custom file format to efficiently store data and metadata. Incorporated multiuser concurrent execution capabilities to ensure data consistency during simultaneous interactions. Utilized reverse engineering techniques to generate an Entity-Relationship Diagram (ERD) for visual representation of the database structure. This project showcased my ability to create advanced software systems, empowering users to effectively manage their data and leverage the power of SQL queries.

Chess game [GitHub]: A console-based chess game application was developed in Java, offering an engaging experience for users to play against the computer. The application adhered to SOLID principles and employed the factory design pattern to ensure maintainability and extensibility. A Test-Driven Development (TDD) approach was applied, with extensive JUnit test cases written to validate the code and modularize its components. The development process followed an Agile methodology, allowing for iterative and collaborative development. To streamline the development workflow, a CI/CD pipeline was implemented to automate monitoring and deployment processes. This comprehensive approach resulted in a well-structured and thoroughly tested chess game application, delivering a seamless and enjoyable user experience.

Professional Experiences

Data Engineer, Jarvis (2023-present): Led the development of user-friendly solution for efficient cluster management, utilizing bash scripts, PostgreSQL and Docker. Created efficient and robust databases, ensuring smooth operations and optimizing performance. Collaborated closely with clients to understand their database requirements and provided customized solutions. Crafted robust Java applications, including a comprehensive JDBC system and a versatile Grep application for pattern searching. Leveraged advanced data analytics techniques to glean insights from transactional data, employing Pandas, Numpy, and Matplotlib within a Jupyter notebook environment. Significantly contributed to an ongoing initiative focused on optimizing core Hadoop components, utilizing Apache Hive and Zeppelin Notebook on Google Cloud Platform, and strategically configuring a master node and two worker nodes for streamlined data processing.

Data Engineer, Wonolo (2022-2023): Designed and implemented a customer lifecycle tracking data asset, improving end-user accessibility and enabling faster insights. Optimized DBT incremental models for improved query performance and data accuracy. Implemented automated tests throughout the pipeline, significantly enhancing data quality and reducing error rates. Optimized SQL queries to improve query performance and ETL processing time. Collaborated with cross-functional teams to gather requirements and design data solutions that meet business needs. Coauthored an enterprise-level Data Warehouse Architecture Narrative. Designed and created data models using ERD diagrams to ensure efficient data storage and retrieval. Communicated effectively with team members and stakeholders to ensure alignment and transparency in project goals and timelines. Demonstrated strong problem-solving skills to troubleshoot and resolve technical issues.

Jr. Product Support Analyst, Exxat Systems (2019-2021): Generated client-specific reports through the development of SQL scripts, resulting in a 40% reduction in report generation time and enhancing overall efficiency. Analyzed software issues to identify root causes and implemented technical resolutions, resulting in a 13% performance improvement. Conducted thorough product testing, promptly communicating faults to the QA team and devising effective workarounds for customers. Implemented cost-effective technology solutions by assessing requirements and ensuring clear communication with stakeholders. Contributed to streamlined processes, optimized performance, and improved customer satisfaction through proficient report generation, problem-solving, and effective communication.

Education

Dalhousie University (2021-2022), Master of Applied Computer Science, Computer Science - GPA: 3.91/4.3

Gujarat Technological University (2015-2019), Bachelor of Information Techonology, Information Techonology - GPA: 8.44/10.0

Miscellaneous

- Anime Enthusiast
- Travel and exploration