Rishi Pandey

📞 +1 343-558-4690 — ✉ Pandeyrishi760@gmail.com — 🔗 Linkedin — 🐙 Github — 🌐 Portfolio

As a passionate **Data Engineer** with over **3 years of experience**, I specialize in transforming raw data into valuable insights through innovative **data pipelines** and robust **database solutions**. Armed with a **Master's in Systems Science and Engineering** from the **University of Ottawa**, I've led projects that automate **ETL processes**, optimize **database performance**, and design scalable **data architectures**. My expertise spans **SQL**, **Python**, and cloud platforms like **Azure**, **AWS**, and **GCP**, allowing me to tackle complex data challenges.

Technical Skills

Programming Languages:	SQL, NoSQL, Python (Pandas, NumPy, PySpark)	**Data Warehouse:**	Azure SQL Databases, Azure Blob Storage, Amazon Redshift, Snowflake, Google Big Query, Amazon S3
Databases:	MySQL, PostgreSQL, MongoDB	**Workflow Management:**	Apache Airflow
ETL Tools:	Azure Data Factory, AWS Glue, Databricks	**Visualization:**	Power BI, Matplotlib, Seaborn
Big Data:	Apache Spark, Apache Kafka, Hadoop	**Testing:**	Docker, Jenkins, Github Actions

Experience

eHospital, University of Ottawa   **Jul 2024 – Present**
Data Engineer

- Creating a **RESTful API** using **Flask** to facilitate secure access to the **MySQL database**, allowing multiple users to interact with the database without requiring direct admin access.
- Implemented **user registration and authentication** features, enabling users to sign up, log in, and receive authentication tokens for accessing protected **API endpoints**.
- Designed **API endpoints** for **CRUD operations**, allowing users to manage patient records, schedule appointments, and communicate with healthcare providers.
- Integrating **Azure Data Factory (ADF)** to fetch healthcare data from the API using **HTTP-linked services** for **ETL pipelines**.
- Processing and storing transformed data into **Azure Data Lake** and a **data warehouse** for advanced analytics and reporting.
- Building **Power BI dashboards** offering real-time healthcare insights, improving decision-making and healthcare delivery.

ETL Developer   **May 2023 – Dec 2023**
University of Ottawa

- Analyzed **business data** for a **healthcare web app** to deliver immediate **preventative care information** to patients.
- Implemented **data ingestion pipeline** using **AWS Glue** to extract and transform patient data by using **Amazon S3**.
- Utilized **Amazon Redshift** to store and manage the transformed data, enabling efficient **querying and analytics**.
- Built **Power BI dashboards** and monthly reports to visualize core data, reducing manual reporting work by hours each week.

Azure Data Engineer   **Jan 2021 – Aug 2021**
Infosys Limited

- Developed and maintained end-to-end operations of **ETL data pipelines** and worked with large data sets in **Azure Data Factory**.
- Researched and implemented various components like **pipeline**, **activity**, **mapping data flows**, **data sets**, **linked services**, **triggers**, and **control flow**.
- Performed extensive **debugging**, **data validation**, **error handling mechanism**, **transformation types**, and **data cleanup analysis** within large datasets.
- Collaborated with **DevOps Engineers** to develop automated **CI/CD** and **Test-driven development pipelines** using **Azure** as per client standards.

Data Engineer   **Jul 2018 – Aug 2020**
Promofox Digital

- Gathered and **analyzed** data from various **digital marketing channels** to **optimize marketing strategies**, **improve campaign performance**, and achieve **business goals**.
- Responsible for **data integrity**, **completeness**, and **reliability** by implementing **data validation processes**, **cleaning datasets**, and **resolving data discrepancies**.
- **Optimized** MySQL and PostgreSQL performance using advanced **indexing** and **partitioning**, improving data retrieval speed.

- Used **Apache Spark**, **Azure Data Lake**, and **Databricks** to **transform** and **aggregate** marketing data, supporting the development of prebiotic suspensions and calcium tablets.
- Streamlined data **loading** and **organization** in **Azure Synapse**, delivering actionable insights for product innovation.
- Configured **Azure Monitor** to identify and resolve pipeline issues, enhancing overall **operational efficiency**.
- Collaborated with senior engineers to translate business needs into **data-driven solutions**, improving product development and workflows.

Education

University of Ottawa
Masters in Systems Science and Engineering — GPA: 4.0/4.0
Major: Computer Engineering

Certifications

- PL 300: Microsoft Certified Power BI Data Analyst Associate
- Ongoing DP 203: Microsoft Azure Certified Data Engineering Associate

Projects

Ongoing Project- Jason Moving

- Integrating and modeling data from **Excel** into **Power BI**, creating a unified and accurate **data model** to support **reporting and analysis**.
- Working with **DAX (Data Analysis Expressions)** for measuring **KPIs** such as **delivery times**, **route optimization**, **fuel efficiency**, and **customer satisfaction scores**.
- Collaborating with the **operations manager** to create interactive **dashboards** that provided real-time insights into **delivery status**, **fleet performance**, and **package tracking**.

Azure Data Factory Data Pipeline using Databricks, Azure SQL Database and Azure Data Lake

- **Used Azure Data Factory** to fetch **ECDC Covid-19 data**, which is then transformed through **Azure Data Flow**, **Databricks**, and **HDInsight**. The processed data was loaded into **Azure SQL Database**, where it was analyzed and visualized using **Power BI**.

Uber Big Data using Google Big Query, Python, Mage AI ETL Tool

- Developed **ETL pipelines** with **Mage AI** for **automation** and **efficiency**.
- Leveraged **Python** for **data extraction**, **transformation**, and **loading**.
- Optimized **data storage** and **querying performance** using **Google BigQuery**.

Exploratory Data Analysis using Python's (NumPy, Pandas, Matplotlib and Seaborn)

- **Leveraged Python libraries** such as **Pandas** for **data manipulation**, **NumPy** for **numerical operations**, and **Matplotlib/Seaborn** for **advanced data visualization** to identify patterns and insights that could enhance marketability.

Canadian Tire Retail Sales Exploratory Data Analysis

- Integrated multiple **data sources** to ensure comprehensive and accurate reporting.
- Implemented **data models** and created complex **measures** using **DAX**.

Scalable Web Application Deployment with NGINX, Docker and CI/CD

- Used **NGINX** as a reverse proxy and load balancer and **Docker** for containerization across development, testing, and production environments.
- Implemented a **CI/CD pipeline** to automate build, test, and deploy processes using tools such as **Jenkins** and **GitHub Actions** for CI/CD automation.