Piyush Badule
Data Engineer

✉ piyush.badule30@gmail.com

📞 +91- 9322479517

📍 Pune, Maharashtra

🔗 piyushbadule.github.io

🐙 github.com/piyushbadule

in linkedin.com/in/piyushbadule

Data Engineer with 5 years of experience, specializing in designing and optimizing ETL pipelines using Azure, AWS, and Databricks. Proficient in Python and SQL for data processing, querying, and manipulation, with expertise in tools like Pandas, Numpy, and PySpark. Skilled in managing data workflows with Apache Airflow and ensuring efficient large-scale data processing. AWS Certified Data Engineer, AWS Certified Developer, and Azure Data Engineer Certified, with strong experience in cloud deployments on both Azure and AWS. Proficient in version control with Git and containerization using Docker, delivering high-performance, scalable data solutions.

🧠 Skills

Skill	Level
Python	●●●●●
Pandas & Numpy	●●●●○
Pysparks	●●●●○
AWS	●●●●○
Microsoft AZURE	●●●○○
Databricks	●●●●○
SQL	●●●●○
Airflow	●●●●○
GIT	●●●●○
Docker	●●●○○
CI/CD	●●●●○
FastApi	●●●●○
Flask	●●●●○
Django	●●●●○
HTML, CSS & JS	●●●●○

💼 Professional Experience

T - Systems ICT India Pvt Ltd, *Software Engineer* ⧉
Jun 2022 – present | Pune, India

∅ **OneBI- Lakehouse:**
- **Description:** OneBI is designed to provide a unified Business Intelligence (BI) solution for T-Systems, supporting both internal and external customers. The project focuses on consolidating data into a single enterprise data platform, enabling efficient data preparation, orchestration, and seamless integration of data for reporting, analytics, and planning. It supports business processes across multiple departments including finance, sales, and human resources. The OneBI platform leverages state-of-the-art BI tools and techniques to deliver reliable and scalable solutions that enhance decision-making and business performance.
- **Technologies:** Azure, DataBricks, DBT, PySparks, SQL
- **Responsibilities:**
  - **Data Engineer**: Worked on developing scalable data pipelines and automated data workflows using PySpark, Azure, and Databricks.
  - **Dashboard Development**: Created interactive dashboards and visualizations in Databricks to support business reporting and data analysis needs.
  - **Notebook Development**: Wrote and maintained PySpark notebooks for data ingestion, transformation, and orchestration across various data sources.
  - **Data Consolidation**: Integrated data from multiple sources into a centralized platform to ensure consistency and availability for reporting and analytics.
  - **Collaboration with Stakeholders**: Collaborated with teams across finance, sales, portfolio, and human resources to develop custom data solutions and ensure business needs were met.
  - **Consultation and Support**: Provided technical support and consultation for the development of predefined reports, dashboards, and semantic models.

∅ **Space-ECB (European Central Bank):**
- **Description:** The SPACE European Central Bank (ECB) Data Engineering Project is focused on migrating and modernizing the management of economic time-series data across various frequencies (Daily, Weekly, Monthly, Quarterly, Annually). This data, which includes critical economic indicators like interest rates, income, and GDP, will be ingested, processed, and stored in a cloud environment on AWS. This modernization aims to establish high-quality, scalable, and reliable data pipelines.
- **Technologies:** Python, PySpark, Pandas, NumPy, AWS Glue, AWS S3, Amazon Redshift, Amazon CloudWatch, Amazon EMR, Docker.
- **Responsibilities:**
  - **Data Migration & Integration:**
    - Migrated legacy data scripts to the AWS cloud environment. Utilized **AWS Glue** to transfer data from on-premises systems to **Amazon S3** for structured storage and easy retrieval.
  - **Data Processing & Transformation:**
    - Developed scalable data processing workflows in AWS Glue ETL Jobs using PySpark for cleaning, validating, and harmonizing time-series data.
    - Created modular Python functions and PySpark User Defined Functions (UDFs) to manage transformations and data alignment across multiple time frequencies, ensuring consistency and accuracy.
  - **ETL Pipeline Design & Orchestration:**
    - Built and scheduled ETL workflows with AWS Glue, and AWS Step Functions, automating the ingestion, transformation, and loading of data without third-party orchestration tools.
    - Monitored pipeline health with Amazon CloudWatch logging and alerting features, proactively troubleshooting issues to maintain data reliability.
  - **Data Quality Assurance:**
    - Designed test cases within AWS Glue to verify data integrity and accuracy at each processing stage.
    - Conducted code reviews to ensure best practices in maintainability, efficiency, and scalability, reducing processing time and improving performance.
  - **Data Storage & Access Optimization:**
    - Organized and loaded transformed data into Amazon Redshift for warehousing, implementing partitioning and indexing to optimize for time-series queries.
    - Implemented role-based access control (RBAC) using AWS IAM to manage user permissions to ensure secure, compliant data access.

🎓 Education

Bachelor of Engineering - Electronics & Telecommunication,

G. H. Raisoni College of Engineering

2016 – 2019 | Nagpur

CGPA: 8.73

Diploma in Electronics & Telecommunication,

Government Polytechnic, Nagpur

2013 – 2016 | Nagpur

Percentage: 72.31

Grade - X, *K. John Public School*

2012 – 2013 | Nagpur

CBSE

CGPA: 8.8

📇 Certificates

- AZURE Certified Data Engineer
- AWS Certified Data Engineer ⧉
- AWS Certified Developer ⧉

🏅 Awards

Certification Of Appreciation

Oct 2024

T - Systems ICT India Pvt. Ltd.

Certification Of Appreciation

Jun 2024

T - Systems ICT India Pvt. Ltd

Pat on the Back

Mar 2024

T - Systems ICT India Pvt. Ltd

Stay Curious & Grow

May 2023

T - Systems ICT India Pvt. Ltd

Pat on the Back

Jan 2023

T - Systems ICT India Pvt. Ltd

🌐 Languages

English ●●●●●

Hindi ●●●●●

Marathi ●●●○○

Ø **DTSE Translation:**
- **Description:** DTSE Translation is a product by T-Systems that enables seamlesstranslation across various languages. It supports text, document, image,website, and live screen translation. The product handles multiple images anddocuments, including embedded images. It offers a versatile solution fordiverse translation needs."
- **Technologies:** Python, Object-Oriented, FastApi, Docker, Azure.
- **Responsibilities:**
  - **Coding and Development:**
    - Led the coding efforts alongside four backendcolleagues, implementing text, image, website translation functionalitiesusing Python Fastapi and integrating Azure cognitive services.
  - **Code Review and Quality Assurance:**
    - Conducted thorough code reviews,ensuring adherence to high coding standards and best practices.
  - **Unit Testing:**
    - Developed comprehensive unit test cases to validate thefunctionality and robustness of the translation services.
  - **Documentation and Guidance:**
    - Authored Docker files, test cases, andprovided guidance to junior developers, fostering their growth andensuring project consistency.

MindNerves Technology Services Pvt. Ltd., *Software Engineer* ⧉

Dec 2019 – Jun 2022 | Pune, India

Ø **Vistabee**
- **Description:** Vistabee is a digital marketing platform tailored for real estate agents and businesses, providing them with tools to create and manage high-quality digital content. My role involved designing and implementing data engineering solutions to enable comprehensive insights and optimize data flow within the platform, thus enhancing content performance analytics and supporting scalable content distribution.
- **Technologies:** Python, Django, PostgreSQL, AWS (S3, Redshift), Apache Airflow, Celery, SQL, Docker.
- **Responsibilities:**
  - **Data Modeling and Pipeline Development:** Designed and implemented data models and ETL pipelines to capture and process real-time user interaction data. These pipelines ensured that all content interaction data was ingested, cleaned, and transformed for analytical use.
  - **Requirement Analysis for Data Needs:** Analyzed business requirements to translate them into robust data engineering solutions, ensuring the data architecture met both functional and analytical needs.
  - **Business Logic for Data Aggregation and API Development:** Developed the business logic to aggregate, filter, and present key metrics on content engagement and conversion rates. Created APIs to expose these insights for integration with other services.
  - **Data Warehousing and Analytics:** Implemented a data warehousing solution using AWS Redshift to store and analyze user interaction data, enabling advanced analytics on content performance, user engagement, and conversion tracking.
  - **Data Processing and Task Orchestration:** Leveraged Celery and Apache Airflow for task orchestration, automating data processing jobs and ensuring the reliability of the data pipeline.
  - **Performance Optimization and Debugging:** Utilized the Django Debug Toolbar and optimized SQL queries to improve the performance and reliability of data retrieval, particularly for analytics dashboards.

Ø **Trumov**
- **Description:** Trumov is a versatile video creation platform designed to assist users in producing engaging videos tailored for various industries such as real estate, news, automotive, and more.
- **Technologies:** Python, Object-Oriented Programming (OOP's), HTML, CSS, JavaScript, Django, PostgreSQL, AWS.
- **Responsibilities:**
  - **Video, Audio, & Image Processing:** Spearheaded the implementation of critical video, audio, and image processing functionalities, enhancing the platform's multimedia capabilities.
  - **Email Configuration:** Successfully implemented email configuration to streamline communication and user engagement.
  - **Responsive Design Resolution:** Collaborated with team members to resolve responsive design issues by leveraging HTML, CSS, and Python, ensuring a seamless user experience. Trumov's commitment to providing a user-friendly and feature-rich video creation solution empowers businesses across different sectors to create captivating video content for their specific needs.