

# Rishi Agrawal

---

+91 9607633657 • Pune • rishiagra110@gmail.com • <https://www.linkedin.com/in/rishi-agrawal-a76a2b339/>

## Professional Summary

---

- Having experience in all stages of the project including requirements gathering, designing
- documenting, development, performance optimization, data extraction, cleaning and reporting.
- Exposure to Azure cloud technologies such as Azure Data Factory, Azure Data Lake gen 2, Data Bricks, Azure Synapse, Snowflake.
- Having experience in design, implement and support different bigdata applications by using Apache Spark, Pyspark, Azure.
- Having experience on Spark SQL, include integration of the Spark with applications.
- Having experience in designing Data Pipelines and Data Handling.
- Experienced in branching, tagging and maintaining the version across the Environments using tools like Git, GitHub.
- Knowledge of programming language Python and its libraries related to bigdata.
- Knowledge of Spark using python, spark core and Spark-SQL for faster processing of data.
- Worked on structured, semi-structured data frame.
- Also having knowledge of Snowflake.
- Familiar with PyCharm and Jupiter notebook IDE.
- Exposure to MySQL, Teradata, Oracle, SQL Server.
- Experienced in Kafka data streaming technology.
- Experienced working with Airflow data as an orchestration tool.
- Experience working with different Spark API like Scala, Python
- A good team player, good communicator.
- Proficient in communicating technical information to non-technical audience.

## Technical Skills

---

- **Operating Systems:** Unix, Windows
- **Databases:** MySQL, Oracle, PostgreSQL
- **Hadoop Ecosystem:** HDFS, Hive, Spark-Core, Spark-SQL, PySpark
- **Integrated Development Environments (IDEs):** PyCharm, Jupyter Notebook, Databricks
- **Programming Language:** Python
- **Cloud Technologies:** Azure (Azure Data Factory (ADF), Azure Synapse Analytics, Databricks)
- **Orchestration Tool:** Azure Data Factory
- **Other Tools and Technologies:** Jira, Git, GitHub

## Work Experience

---

**Organization:** E-Mech Solutions Pvt Ltd (Client Company – Enixo Corporate Solution Pvt. Ltd.)

**Designation:** Data Engineer

**Location:** Bengaluru , Karnataka.

**Duration:** January 2022 – Present

## Industrial Projects

---

### Project #1:

#### Data Pipeline Automation with Azure Data Factory, Databricks, and SQL Server

##### Roles and Responsibilities:

1. Developed ADF data pipelines to extract data from SQL Server and load it into the Bronze layer following Medallion Architecture.
2. Configured self-hosted integration runtime and parameterized connection strings for seamless data ingestion.
3. Performed data transformations using Azure Databricks (PySpark, SQL) for loading data into the Silver layer.
4. Collaborated with DBT and Synapse Analytics for further data refinement in the Gold layer.
5. Supported Power BI teams by ensuring clean, structured data for reporting and business insights.

**Environment:** Azure, Databricks, ADF, ADLS, Key Vault, DBT & Synapse Analytics

### Project #2:

#### Data Migration and Transformation

##### Description:

1. Migrated data from **SQL Server** to **Snowflake**.
2. Performed migration of data in various file formats and from **RDS to Snowflake**.
3. Transformed data as per ETL specifications and business logic

##### Roles and Responsibilities:

1. Developed **PySpark** scripts for data extraction and transformation.
2. Proficient in **Spark architecture**, including **Spark Core** and **Spark SQL**.
3. Built optimized data pipelines using AWS services such as **S3, Glue, Athena, and Redshift**.
4. Designed and implemented batch processing workflows, integrating solutions for both structured and unstructured data.
5. Applied **data ingestion**, transformations, and performance tuning techniques.
6. Ensured data quality by cleaning inaccurate and corrupted data, improving overall data reliability.
7. Worked extensively with **Parquet file formats** and applied compression techniques to optimize storage.
8. Filtered data and performed operations to generate meaningful insights aiding client decision-making.
9. Collaborated on upstream source analysis and developed solutions aligned with business cases.
10. Utilized **Airflow** as an orchestration tool for workflow management.
11. Experienced in working with **Spark APIs**, including Scala and Python.

**Environment:** PySpark, AWS (S3, Glue, Athena, RDS), SparkSQL, DMS, Snowflake

## Educations

---

**Degree : BTech**

**University :** Dr. Babasaheb Ambedkar Technological University (DBATU), Maharashtra

**Graduated Year:** 2022

## Personal Information

---

**Date of Birth:** 11th September 2000

**Gender:** Male

**Nationality:** Indian

**Mobile No.:** +91 9607633657

**Current Place :** Pune

**Languages Known:** English, Hindi, Marathi

I hereby declare that the above information given is true and correct to the best of my knowledge and belief.

Place: Pune

(Rishi Sunil Agrawal)