

Gaurav Sahadev

📍 Kolkata

✉ gauravsahadev@outlook.com

☎ +91-7979788681

🌐 <https://gauravsahadev.github.io/>

Work Experience

Cognizant

Associate (Data Engineer)

Kolkata, West Bengal

December 2020 - Present

- Developed scalable applications for ETL and Analytics for one of the largest insurance companies in North America.
- Work on design and development while remaining concentrated on client needs.
- Worked on technologies like Informatica, PySpark, Hadoop, SQL, and have good knowledge of Azure cloud and ETL technologies.
- Contributed to the delivery of each project (analysis, development, and testing)
- Onboarding new developers into existing projects

Education

Siliguri Institute of Technology

Bachelor of Technology Computer Science

Siliguri

August 2016 - July 2020

Skills

Programming Languages:	Java, Python, Shell, SQL
Architectures & Concepts:	ETL (Extract Transform Load), Data structures & algorithms
Databases:	MySQL, Cassandra, MongoDB, Hbase
Frameworks:	PySpark, Hive, HDFS (knowledge of Kafka, Sqoop), Pandas
Other Technical Skills:	Data wrangling, Data Analysis using Python, Airflow
Azure Services:	Azure Data Factory, Azure Data Lake Storage, Azure Databricks, Synapse, Cosmos DB, Key Vault

Personal Projects

Real-Time ETL Pipeline:- PySpark, Databricks, Azure Data Lake Storage, Azure SQL database

- Built a Real-Time ETL Pipeline using PySpark and Azure Data Lake Storage.
- Used JDBC connection string to fetch data from the Azure SQL database and created PySpark Dataframe.
- Cleaned the data to remove duplicate and null values from dimension and fact tables.
- Applied transformations on selected fields of Dataframe exported the cleansed data to ADLS container in CSV format.

API for Docx to PDF conversion:- Python, Flask, Shell

www.pdfbanao.com

- Built API using flask and python to convert Docx file to pdf.
- API is further integrated with a web application and is deployed to production (2.5k - 3k monthly users).

Achievements

- Explore ML Facilitator (Speaker) - Google
- Software Innovator program (Speaker) - Intel

Certifications

- **Azure Data Engineer (DP 203)** - Microsoft
- **Azure Data Fundamentals (DP 900)** - Microsoft
- **Azure Fundamentals (AZ 900)** - Microsoft