

DHRUBAJYOTI CHATTERJEE

Azure Data Engineer @ DISYS India (Jan 22 to Jan 2023)

+919434221699 | dhruuba1008@gmail.com

CAREER HIGHLIGHTS

- Passionate software Engineer with overall 20+ years of experience, with proficiency in platform/solution architecture, developing robust data pipelines and software systems.
- Enthusiast of open-source software, contribute back to community whenever possible by writing code and answering Bigdata (Apache Spark and Hadoop) questions in global forum.
- Quick at acquiring of modern technologies and ability to work in a team as well as individual.

PROFESSIONAL SUMMARY

- **Overall 19 + Years** of professional experience in Indian Army as **Senior Engineer** (2002-2021).
- Over **9+ years** of expertise in IT designing and implementing solution, **4.5 years exp data strategy execution, Azure cloud, AWS S3, Big Data , Data Lake, Data Warehouse, and Agile** with strong resource management & support for diverseresolutions and technical platforms.
- Hands-on in apps architecture design for projects, ensuring the architecture is appropriate across the business domain to fulfill actual business needs.
- **Azure Data Cloud experience 4.5 years**
- Proficient in (**Azure DataLake, Data Bricks ,Synapse analytics, Azure Data factory, Event Hub**).
- Hands-on experience on Azure synapse analytics for big data analytics, data warehousing.
- Good on implementation of data pipelines using **Azure data factory & SSIS**, data modelling & Power BI.
- **2.5+years** of experience in **Bigdata, Kafka, Event Hub, Azure Streaming Analytics, Data Explorer, SQL & T-SQL, Spark,Python, Airflow, Snowflake, Power BI**.
- Experience in working on Hive external/internal table, knowledge in portioning and bucketing.
- Worked on **PySpark using RDD** and Data frames, functions which were introduced in PySpark (1.Xand 2.X)with Hadoop (2.0 and 3.0)
- Good handson experience and understanding of implementing large scale data warehouse solutions in **Snowflake, Azure Synapse , AWS**.
- Performed data ingestion from **Snowflake, MongoDB, Oracle, MySQL, SAP, Salesforce, IoT Sensor** data .
- Experience to use data to make **better-informed, fact based decision**.
- Helping business to **better understand their customers & markets**.
- Focused on **improving and automating business process** with data.
- Experience to **work with business directly and solutions** within specified time frame.

TECHNICAL EXPERTISE

Big Data (Framework)	: Spark, Hadoop, Hive, Kafka
Databases	: MySQL server, Hbase , HDinsight
Data Pipeline	: Azure Data factory, ELT/ETL
Scheduling	: Airflow
Azure Technologies	: Azure Data factory, Data lake, Blob , HDInsight, Event Hub
Cloud Data warehouses	: Azure Synapse , Snowflake
ETL Tool	: Azure Data Factory
Programming Languages	: Python
Cloud	: Azure Stack, Azure DevOps, Azure Databricks
DevOps	: Agile, CI/CD
Data Analytics	: Power BI

PROJECTS

Project #1: Oil & Gas (USA) (Nov 2022- Jan 23)

Environment : Snowflake, CRM, SQL server, Web API, Azure Data Lake, Azure Data Factory & Power Apps

Responsibilities:

- Data ingestion to curation Architecture design.
- Preparing Data Factory pipelines for loading data from ingestion to curation.
- Ingest data from Snowflake to Dataverse.
- Working in team for development of needed scripting & procedures for data import and manipulation
- Work in Mapping data from source to destination including extraction, transformation, and load processes.
- Hands on to create lookup table with filter function in Dataflow and pipeline.
- Data ingestion from Snowflake DW to Azure Data lake & Azure to Powerapps.
- Deployed CI/CD process for each change to the GitHub repo that starts a build in Azure DevOps, and a CD pipeline runs a deployment to Azure.
- Continuously working with client need & simultaneously replying clients email for any changes.

Project #2: Retail (UK) (Sep 2022- Oct 22)

Environment : Apache Spark, Synapse , Spark core & Data Frame, Spark SQL, Python, HDFS, ADLS .

Responsibilities :

Worked on all types of data like semi-structured and structured.

Developed data pipe lines using hive and spark.

Used Azure Data Factory Pipeline for Databricks linked service.

Applied several transformations and actions to process the data and saved it as CSV and JSON files which will be used in data visualization.

Experience in optimizing queries for better performance.

Written spark job to join different types of data and created final table and saved it in hive.

Description : Reporting and presenting the expenses. Reporting the profitability and transactions for improved financial forecasting and budget planning to identify and manage financial planning risks. To gather accurate insights on branch performance, sales performance, profitability attribution and track credit pipelines.

Project #3: HealthCare (UK) (Jul 2022- Aug 22)

Environment : Snowflake(SAP), CRM, SQLServer, Azure Data Factory, Databricks, ADLS, Power BI

Responsibilities:

- Used Data Factory pipelines for loading data from Cloud data warehouse.
- Data processed & migrated to ADLS. Used T-SQL.Used Custom scripts and stored procedures for data import and manipulation.
- Developed Pipelines in Azure Data factory for ETL.
- Map source system data elements to target system and develop, test, and support extraction, transformation, and load processes. Create lookup table with filter function in Dataflow.
- Create schedule automated refresh using gateway (differ by pipeline).
- Create pipeline for Powerapps & worked with data scientists for data manipulation.

Project #: Oil & Gas (Jan 2022- Jul 22)- Two Projects

Environment : Snowflake(SAP), SQL Server, Azure Synapse , Data bricks, Azure Data Factory & Data Science.

Responsibilities:

- Model & Architect DW design, direct and/or execute the technical characteristics of the overall strategy of the data warehouse and ETL process .
 - Lead a team on Data Factory pipelines processing & loading from SQL db to destinations. Used Azure Synapse Analytics.
 - Migrated all packages from on premises to Azure Data factory. Custom scripting and stored procedures for data import and manipulation. Developed Pipelines in Azure Data factory using Multiple Activities.
 - Map source system data elements to target system and develop, test, and support extraction, transformation, and load processes.
 - Lead a team to Create schedule automated gateway for pipeline to curation zone (Dataverse).
 - Use GitHub repo for CI/CD.
 - Lead a team to Designed data centric Power BI dashboards.
- Continuous working with client for gathering requirement and pass to team for improvement

Project #5: Defence Project Management {Healthcare, Finance(2016-21)}

Role : Senior Data Engineer

Environment: SQL Server(SQL & T-SQL), ADLS,
Azure Data Factory, Azure Synapse , Power BI.

Responsibilities:

- Leading a team of proficient engineers, developing Azure Data Factory Pipelines for data ingestion from SFTP or On-Prem.
- Loading data into Azure Synapse DW.
- Making data strategy for simple segmentation from the variety of data.
- Data transformation & create and deployment of cube in Tabular model. Used T-SQL in query to fetch rows.
- Design the Dashboards using different types of Power BI visualizations like Charts, Bubbles Maps, Tree Map, Slicer and KPI. Creation of Bar Chart, Line Chart, Area Chart and Slicer etc.
- Charts are used in making reports more user-friendly in analysis purposes.
- Performed daily data queries and prepared reports on daily, weekly, monthly, and quarterly basis.
- Creating Incident Management Dashboard Providing Click throws from summary to Detailed Level.
- Installed and Configured Data Gate ways to Refresh and Schedule Refresh the Dataset.
- Managed access of Reports, Dashboards and Data for in dividable users using roles.

Project #6 : Defence Project Management {Healthcare, Finance(2014-18)}

Role : Senior Data Engineer

Environment : ADLS, Azure Data Factory, Databricks, Power BI

Responsibilities:

- Model & Architect DW design, direct and/or execute the technical characteristics of the overall strategy of the data warehouse and ETL process.
- Created Data Factory pipelines processing data loading from source to destinations.
- Data migration from on premises to Azure Data factory using Custom scripts and stored procedures for import.
- Extensively using Azure Analysis Services to create semantic models
- Developed Pipelines in Azure Data factory using Multiple Activities.
- Write a required DAX functions using aggregates, date, logical and table functions.
- Created Tabular cube and published to Azure Analysis Services & schedule automated refresh using to gateway.
- Designed Power BI dashboards in Online by taking Data from Azure DW.

Certification/independent courses

- **MBA (Pursuing)- NMIMS- 2022**
- **Graduated from University of Calcutta (B.A)-2009**
- **“First class” Diploma in Electronics Engineering** (2nd highest on Project) from **Military College of EME**, Secunderabad, Telengana.
- **Microsoft certified Azure Data Engineering Associate (DP-203)**
- **Snowflake SnowPro Core certified**
- **Microsoft Databricks Accredited Lakehouse Fundamental**
- **Certified Scrum Master**
- **German Language** Level A1, A2 (pursuing).

Awards

- Nominated for **Chief of Army Staff cash award** in Defence secured data migration for two consecutive years in New Delhi 2018, 2019.
- Nominated for best Technical Trainer award for given Data engineering live project training to more than 7 batches of Army Officers. Participated and organized various **strategic, technical** events in **Defence with USA, Russian, Israel conglomerate**.