

Tanush Aggarwal

Data Engineer

github.com/TanushAgg | candidates.intervues.club/u/tanush-aggarwal

SUMMARY

Data Engineer with experience building scalable ETL pipelines, cloud data platforms, and medallion architecture solutions. Skilled in PySpark, Databricks, Azure Data Factory, Snowflake, and Spark performance tuning to optimize data workflows. Proven track record in migrating legacy Hadoop systems to modern lakehouse environments.

EXPERIENCE

Data Engineer - Associate Software Engineer

Intelibliss

Apr 2025 – Mar 2026

- Engineered PySpark Databricks notebooks ingesting data across 5+ source formats including cloud, API, CSV, JSON, and Parquet.
- Built enterprise solutions using Databricks, Azure Data Factory, and ADLS Gen2 adhering to a 3-layer medallion architecture.
- Optimized Delta Lake performance using OPTIMIZE, Z-Ordering, VACUUM, Liquid Clustering, and compaction, reducing query latency by 20%.
- Tuned Databricks Spark configurations via broadcast joins, salting, caching, and persistence, reducing execution time from 20 to 12 minutes.
- Migrated legacy Hadoop HDFS Spark pipelines to Databricks and Snowflake orchestrated with Apache Airflow.
- Implemented AIOps Metric Intelligence and Predictive Intelligence models, automatically routing 35% of tasks.
- Automated IT and HR workflows using Virtual Agent and Flow Designer, reducing overall ticket volume by 25%.

Data Engineer - Associate Engineer

Emerson (Contractor)

Dec 2024 – Mar 2025

- Optimized Apache Spark jobs using broadcast joins, salting, and caching, improving ETL pipeline performance by 30%.
- Analyzed 10,000+ employee records using SQL for data validation and recurring reporting.
- Developed Power BI dashboards to track 25+ HR and payroll KPIs.
- Conducted root-cause analysis on reporting defects and collaborated with 5+ cross-functional stakeholders.

Data Analytics Intern

Shree Adisoft Technology Pvt. Ltd.

Jun 2023 – Dec 2023

- Developed 10+ Power BI dashboards consolidating 25+ KPIs to enhance operational visibility.
- Maintained Power BI reporting workflows supporting 30+ monthly report refreshes.

Software Engineer Intern

Water And Power Consultancy Services (Wapcos)

Jul 2022 – Sep 2022

- Updated 10+ official web pages using HTML and CSS to ensure consistent layouts and smooth site functionality.

EDUCATION

Guru Gobind Singh Indraprastha University · Bachelor of Technology - BTech · 2020 – 2024

Udemy · Course Certificate · 2023 – 2023

Coding Ninjas · Course Certificate · 2022 – 2022

SKILLS

Languages and Frameworks: CSS, HTML, JavaScript, PLpgSQL, Python, Spark SQL, SQL, Apache Spark, Delta Lake, Hadoop HDFS, PySpark

Databases: Delta Tables, Relational Databases, Snowflake, SQL Server

Tools: ADLS Gen2, AIOps Metric Intelligence, Apache Airflow, Azure Blob Storage, Azure Data Factory, Azure Data Lake Storage Gen2, Azure Databricks, Azure Synapse Analytics, Broadcast Joins, Bronze-Silver-Gold Architecture, Caching, Data Modelling, Data Warehousing, Databricks, File Compaction, Flow Designer, Git, GitHub, Lakehouse Architecture, Linked Services, Liquid Clustering, Mapping Data Flows, Medallion Architecture, Microsoft Azure, OPTIMIZE, Parameterized Datasets, Parent-Child Pipelines, Partitioning, Persistence, Power BI, Predictive Intelligence, Salting, Shuffle Optimisation, Spark Performance Tuning, Storage Event Triggers, VACUUM, Virtual Agent, Z-Ordering

PROJECTS

Modular ADF ELT Pipeline

Built a modular ADF ELT pipeline with parent-child orchestration across 10+ datasets. Designed and implemented modular Azure Data Factory ELT pipelines with parent-child orchestration across 10+ datasets.

Azure Data Factory

Additional projects: End-to-End Automated Data Pipeline

AWARDS

Rising Star Award - Intelibliss, Spot Award - Intelibliss, Rising Star Award, Spot Award, Python(Basic), Data Manipulation In Python, Python Foundation with Data Structures and Algorithms