

Anurag Singh

48 Allston Street, Boston, MA 02134, USA

+1(980)287-8012 | singhanurag2194@gmail.com | www.linkedin.com/in/anurag-singh-68810710b

EDUCATION:

University of North Carolina at Charlotte, Charlotte, NC, USA

[Dec 2020]

Master's in information technology (Conc: Advanced Data Science & Knowledge Discovery)

University of Mumbai, Mumbai, India

[May 2016]

Bachelors in Electronics & Telecommunication Engineering

Data Engineering /Cloud Big Data Engineering Experience/Certification:

Data Engineering: Azure Data Factory, Databricks, Data Vault 2.0, dbt, Spark SQL, BigQuery, MLOps

Cloud Services: Azure, Azure ML, Logic Apps, AWS, Redshift, Glue, SageMaker, GCP, Kafka [Confluent], Spark, Scala, Hive

Services: Delta Live Tables (DLT), Unity Catalog, Snowflake, Airflow, Collibra, GitHub Action, SonarQube

Development Skills: Data Platform, Data Pipeline, Data Modeling, Machine Learning Integration, DQ, Automation

Certification: Databricks Certified Data Engineer Associate (CredentialID: 101934589)

WORK EXPERIENCE:

Data Engineer II (Consultant), Cervello Inc. (Boston, MA, USA)

[May 2021– Present]

- Design & developed an efficient ETL data pipeline architecture, optimized the load processes to extract client customer's data from the distributed source architecture into the blob & loaded it into Snowflake stages, resulted in time reduction by **35%**.
- Implemented *Data Vault 2.0* methodology for business analytics platform resulting in continuous development and deployment with scalability, flexibility, and enhanced data governance. Onboarded **500+** active users on the day of release and onboarded new users by using Salesforce CRM platform through Row Level Security process.
- Extracted historical data from **50+** client production servers and engineered an incremental loading mechanism, facilitating daily consumption by the analytical framework. Implemented a data ingestion strategy to ensure efficient and up-to-date analytics capabilities with Structured and Semi-Structured data.
- Designed a secure Lakehouse platform using a multi-hop architecture for dev/test/prod environment. Optimized Spark jobs for performance & efficiency by tuning configurations, partitioning data, & leveraging caching.
- Created Python APIs for the Data Quality teams to runs inbound/outbound dataflow process as Daily Jobs, pull Collibra KPIs and load to Snowflake and into Salesforce CRM. Resulted with **20%** decrease in data redundancy & **15%** increase in DQ.

Data Engineer I (Analyst)

[May 2021– July 2022]

- Developed standard for warehouse design in MPP (Massively Parallel Processing) engines such as snowflake while collaborating with analysts and data scientist in the creation of efficient data models for deployment with MLOps.
- Deployed the CI/CD process for Databricks, Data Factory, Airflow & Snowflake for Development/QA/ Production environments. Leveraging GitHub Action to automate the CI/CD workflows in the same GitHub Repository.
- Automated client's weekly reports using Power BI to keep track of general ledger data. This helped in analyzing the trends related to logistics and supply chain management.

Internal Initiatives:

- Provide key contribution to the Cervello Data Engineering & Data Science combine research group to setup the stock trading pipeline using AWS services such as Kinesis, Sagemaker & Glue. Adopted Lambda Architecture.
- Currently acting as mentor and providing operational guidance for Cervello Center of Excellence for Advanced Analytics. Responsibilities include training initiatives and marketing blogs aimed for promoting firm development and expansion.

Data Analyst, Deloitte (Charlotte, NC, USA)

[April 2021]

- Assisted in creating Dynamic Python Scripts, optimized SQL queries: Stored procedures for Dashboard and achieving **30%** improvement in Dashboard performance. Managed scheduled triggers & jobs for data analytics and predictive modeling.

Graduate Assistant, University of North Carolina (Charlotte, USA)

[Sep 2020 – Dec 2020]

- Worked with Batch & Streaming Data for flights arrival & departure delay and provided business solution to understand it's impact on the customers itinerary. Performed real-time analysis with Managed cloud GCP platform to perform ELT.

System Engineer, Infosys Private Limited (Pune, India)

[Mar 2017 – July 2019]

- Ingested streaming banking data into the Snowflake DB for predictive analytics. Built ML model to analyze the transaction is legit or fraudulent & build a predictive model that would perform Multiclass classification.
 - Reported business requirements for 'Bank of America' & created Technical Design Documents to ensure all system support s initiatives. Parallely worked extensively using Databricks analytics tools to publish various time series insights.
-