

Anudeep Kolluri

(682) 391-6787 • upskill.anudeep@gmail.com

Arlington, Tx • linkedin.com/in/kolluri-anudeep • github.com/Anudeep-Kolluri

PERSONAL PROFILE

Data Engineer with a strong vision for building scalable data solutions and expertise in relational databases. Skilled in problem-solving, decision-making, and effective communication, with a focus on delivering data-driven insights and innovations. Committed to continuously learning and solving complex data challenges.

EXPERIENCE

Data Engineer intern, Arlington, Tx

Aug 2024 - Present

University of Texas at Arlington

- Automated reporting processes for university lab incidents by developing ETL workflows in Apache Airflow, improving pipeline efficiency by 40%.
- Built scalable analytics pipelines on Hadoop to store and process university lab data, leveraging Snowflake for high-performance querying and AWS infrastructure for cloud-based solutions.
- Analyzed incident and shift data to identify patterns in lab operations, providing insights that informed decision making by the Office of Information Technology (OIT).
- Delivered insights through dynamic dashboards in Tableau, enabling leadership to make data-driven decisions.

Associate Software Engineer, India

June 2022 - Jan 2023

Accenture

- Built complex ETL solutions for data warehousing, data integration, and analytical and predictive solutions using Informatica Power
- Developed 200+ optimal Informatica Power Center mappings for slowly changing dimensions (SCD) Type 1, 2, and 3 scenarios, catering to diverse source systems like SAP, Salesforce, Flat file, SQL Server, and Teamcenter application.
- Used various Informatica transformations like Filter, Expression, Sequence Generator, Update Strategy, Joiner, Stored Procedure, and Union to develop complex mappings in the Informatica Designer.
- Automated these Informatica data flows using the scheduling tool BMC Control-M.

PROJECTS

YouTube Data Pipeline: Integrating ETL, ML, and Analytics for Trend analysis

Oct - Nov 2024

- Developed an ETL pipeline to scrape and analyze YouTube data, storing results in DuckDB for fast querying.
- Used **Apache Spark** for distributed data processing to ensure scalability and speed while handling large datasets.
- Integrated a **YOLO** machine learning model for image recognition on thumbnails, yielding insights on how thumbnail content impacts views.
- Leveraged **Dagster** for workflow orchestration, managing data extraction and transformation tasks.
- Tools: DuckDB, Python, DBT, Dagster, TensorFlow, YOLO, Apache Spark, scala, java
- Project Link: <https://kolluri-youtube-dashboard.streamlit.app/>

Automated Company Scraper: ELT and Snowflake Analytics

Sept - Oct 2024

- Designed and implemented an ELT pipeline to scrape data from Y Combinator's website and store it in **Snowflake**.
- Utilized DBT for data transformation and performed analytics directly in Snowflake.
- Automated the pipeline with Airflow, ensuring regular updates on a weekly schedule.
- Tools: Snowflake, DBT, Airflow, Selenium, Python
- Project Link: <https://github.com/Anudeep-Kolluri/automated-company-scraper>

EDUCATION

University of Texas, Arlington

Dec 2024

Master of Science in Data Science **GPA : 3.89**

Coursework: Distributed Systems, Cloud Computing, Machine learning, Data Visualization, Database Systems

Bharath Institute of Higher Education and research, India

June 2022

Bachelor of Technology in Computer Science

Coursework: Data Structures and Algorithms, Neural Networks, Big Data, Cloud Computing, Project Management