VANGALAPUDI DHANYA

dhanyav026@gmail.com | +91 8921 777737 | [Portfolio](#)

Summary:
Dedicated and highly skilled Data Engineer with 2.5+ years of experience in designing scalable data architectures, optimizing ETL workflows, and implementing cloud-based solutions. Proficient in **Python, SQL, BigQuery,** and **GCP tools**, with expertise in frameworks like **Airflow** and **Dataform**. Demonstrated success in automating large-scale data migrations, developing efficient pipelines, and delivering data-driven insights. Known for solving complex problems through innovation and leveraging cutting-edge technologies to enhance performance and collaborating across teams to achieve organizational goals.Enthusiastic about applying expertise to build data-driven solutions that enhance business decision-making and efficiency.

Educational Qualifications:
- B.Tech, Production Engineering, National Institute of Technology, Calicut, 2018-2022 (GPA: 8.43/10)
- Class XII, Sri Chaitanya Junior College, 2015-2017 (Score: 977/1000)
- Class X, Siddhartha High School, 2015 (Score: 9.5/10)

Work Experience:

Data Engineer at NewsCorp (Nov 2023 – Present)
1. **Developed and Optimized Data Pipelines:** Designed and deployed 16+ **DataFusion pipelines** for seamless Salesforce and database file ingestion, consistently meeting project deadlines.
2. **Enhanced Engineering Standards:** Created Python modules and optimized codebases, significantly improving the efficiency and reliability of data engineering processes.
3. **Airflow Operator:** Developed a custom **Airflow operator** to optimize GCS file composition, boosting system performance and robustness.
4. **Customized Airflow Dataform Operator:** Developed a tailored **Airflow operator** by merging the **DataformCreateCompilationResultOperator** and **DataformQueryWorkflowInvocationActionsOperator** into a single class, enhancing process effectiveness by enabling compilation and invocation in a unified workflow
5. **AI-Powered Automation for Migration**: Automated the migration of 400+ SQL scripts from a Git repository to Dataform using Vertex AI Prompt Management and the Gemini-1.5 Flash Model, reducing migration time by 70% and increasing process efficiency by 95%, significantly enhancing quality and speed of migration.
6. **Data Transformation Optimization:** Leveraged **Dataform** to streamline and enhance SQL code transformations, improving workflow efficiency and ensuring consistency.
7. **Thought Leadership:** Delivered impactful presentations on **data architecture evolution** and **IoT initiatives**, earning recognition and positive feedback from organizational stakeholders.

Data Engineer at Searce (June 2022 – Nov 2023)

1. **SQL Server to BigQuery Migration:** Designed and implemented an end-to-end solution to migrate **SQL Server databases** to BigQuery, reducing migration costs and improving performance. Developed a robust mechanism to handle **5 schema drift scenarios** during migration using **Change Data Capture (CDC)** to capture and synchronize incremental changes hourly.
2. **Snowflake to BigQuery Migration:** Led the development of a fully automated migration script, transferring approximately **5TB of data** from Snowflake to BigQuery within a **2-week timeframe**, ensuring a seamless and efficient transition.
3. **Data Mesh Workshop for Google Cloud Meetup:** Worked on the ideation and preparation of the presentation for a Data Mesh Implementation workshop in GCP, which led to an opportunity to present to Google Cloud leadership and received commendation for the insights shared.
4. **IoT Data Streaming Solution:** Built a real-time streaming solution using **Pub/Sub** and **Dataflow**, enabling efficient ingestion and loading of IoT data into BigQuery.
5. **Email Alerting Mechanism:** Developed and implemented an automated **email alerting mechanism** using **BigQuery** and an **SMTP server**, enhancing monitoring and notification capabilities.

Skills:

1. **Programming Languages:** Python, SQL, PySpark
2. **Cloud & Data Tools:** GCP,BigQuery,Dataflow,Dataform, Pub/Sub, Cloud Storage,Airflow,ETL
3. **Architectures & Frameworks**:
   a. **Data Warehousing:** Expertise in designing, implementing, and optimizing cloud-based data warehouses (e.g., BigQuery)
   b. **Data Lake**: Building scalable and secure Data Lake solutions using GCP tools for raw data ingestion and storage
   c. **ETL**: Design and development of robust ETL pipelines for data extraction, transformation, and loading across cloud platforms

---

Certifications:

- Google Professional Data Engineering Certification(**Active: Dec 2022 – Dec 2024)**
- Google Professional Cloud Database Engineer Certification.(**Active: Apr 2023 – Apr 2025)**
- Google Professional Machine learning Engineer Certification.(**Active: July 2024 – July 2026)**

Academic Projects:

1. **Assessment and Prediction of Stressors among Healthcare Workers during COVID-19 using Machine Learning Techniques**
   - Collected data through offline and online surveys.Conducted Exploratory Data Analysis (EDA) and implemented five machine learning algorithms for classification and prediction of stressors among healthcare workers.
2. **Perform 'Exploratory Data Analysis' on Dataset 'Sample Superstore'**
   - Performed Data Cleaning, Data Visualization, and Data Analysis to identify areas for profit maximization.

Positions of Responsibility:

1. **Class Representative at National Institute of Technology, Calicut** (Sep'20 – May'22)
   - Organized and facilitated class meetings and events, ensuring effective communication among members.

Academic & Co-Curricular Achievements:

- **Employee Spotlight Award (FY'24):** Recognized for outstanding performance, teamwork, and contribution to organizational goals.
- **Star Performer of the Month (April 2024):** Awarded for exceeding targets and demonstrating exceptional dedication to project success.
- **Team Lead for Final Year Project:** Demonstrated exceptional leadership by guiding the team to successfully deliver a complex project within deadlines.
- **4th Position, All India Inter-NIT Kho-Kho Tournament:** Showcased dedication, resilience, and teamwork in a highly competitive national event.
- **"Most Entrepreneurial Team," Climate Data Hackathon:** Recognized for innovative solutions in "Climate Resilient Agriculture," emphasizing creativity and problem-solving.