ANUPRIYA KAUSHIK

813.573.5803 | ANUKAUSHIK2201@OUTLOOK.COM | HTTPS://WWW.LINKEDIN.COM/IN/ANUPRIYAKAUSHIK

SENIOR DATA ENGINEER

Adept in orchestrating sophisticated data infrastructures using tools like Azure, Airflow, and PySpark, excel in automating operations to enhance efficiency and data integrity, facilitating insightful business decisions. With a strong foundation in machine and deep learning, can tackle complex data analyses and foster multidisciplinary collaborations to exceed strategic goals.

PROFESSIONAL EXPERIENCE

VARIS   2022 – 2024
Sr. Data Engineer

- Orchestrated deployment pipeline of **Azure services** and an **Azure Data Factory (ADF)** instance to establish a catalog pipeline automating supplier catalog processing for marketplaces using **Azure Blob and SFTP** for storage services; **cut processing time by 14%**, enhancing operational efficiency.
- Curated data mapping algorithm using supplier data to internal VARIS's taxonomy with help of **Python** and **Protégé**, elevating marketplace item classification **accuracy by 20%**.
- Developed a comprehensive status service using **FAST REST APIs** in **Python** for a catalog pipeline, streamlining process tracking, and **reducing idle time by 10%** through proactive team and supplier notifications.
- Pioneered an ML Ranking model integrating with **SOLR using LTR**, refining search relevancy by 28% through meticulous data preparation and feature engineering.
- Crafted multiple **Azure Dashboards**, using **Azure Web API** logs and **Cosmos DB** data showcasing **KPIs** like processing time reduction, error rate, and throughput, to provide daily operational insights to stakeholders helped in enhancing decision-making efficiency and driving measurable improvements resulting in **productivity and process optimization**.
- Guided and mentored junior team members within the scrum team, **enhancing overall team performance and knowledge** through targeted coaching sessions.

APPLE   2021 – 2022
Sr. Data Engineer

- Designed and implemented **Python** scripts using **PyMongo** and **FLASK APIs** for data preprocessing of Apple SIRI data stored in **MongoDB**, enhancing language model **performance by 12%**.
- Created and maintained robust data pipelines using **Python** to manage data mart for Apple Siri NLP model to store it into **MongoDB** to be used for analytics and visualization enhancing stakeholder's decision-making.
- Developed and deployed **Tableau** dashboards to continuously measure and report data availability, utilization, and capacity trends, enhancing operational efficiency through proactive tuning and optimising the **review process by 20%** through improved visual analytics.
- Orchestrated an **ETL** process in **Python** to consolidate data from diverse sources such as **PostgresDB**, enhancing data utility for strategic reporting. Optimized **SQL** queries to improve performance for large-scale data processing.

ANTERIAD   2020 – 2021
Data Engineer II

- Automated key data processes with **Airflow** DAGs running on EMR clusters, significantly improving data handling efficiency with **quicker turnaround time and high productivity**.
- Designed and implemented **ETL** processes by leveraging **SQL** to manage **Slowly Changing Dimensions** (SCD) in **data warehousing**, enabling accurate tracking and historical data analysis of evolving business entities, significantly **enhancing data integrity** and reporting capabilities.
- Integrated notification tasks into **Airflow DAGs** to automate alerts for completion, success, and errors to designated Slack channels, enhancing operational **efficiency by 20%**.

- Optimized existing data flow configurations in **Airflow, reducing duplicate data by 30%** in **multi-TB data** sets and enhancing data access speed and cleanliness for analysts and clients. Implemented rigorous **Airflow tests** to ensure the reliability and efficiency of these improvements.
- Developed and deployed **PySpark** scripts to transform and efficiently deliver reporting data to stakeholders via **data marts**, enhancing decision-making and **operational efficiency by 14%**.

UNIVERSITY OF SOUTH FLORIDA   2019 - 2020
Data Engineer

- Designed an ML model to classify human emotions into low/high valence & low/high arousal classes using a human emotion dataset and **Random Forest ML** algorithm which improved existing emotion classification **accuracy by 20+%**.
- Devised a multi-modal **deep learning model** to classify human emotions, integrating EEG and physiological data such as respiration rate, heart rate, and skin conductance. This approach not only enhanced the classification accuracy but also surpassed the existing state-of-the-art networks **by 4+%** and improved upon baseline studies **by approx 26%.**
- Evaluated a machine learning model's performance in classifying human emotions across brain lobes and hemispheres using **Heatmaps**, **Confusion Matrices**, and **ROC Curves**. Identified high performance in the frontal and temporal lobes, achieving a **15% increase in overall accuracy** compared to baseline.

TATA CONSULTANCY SERVICES   2016 – 2018
Data Engineer

- Implemented **REST APIs** with **Spring** framework to integrate Bing Maps for displaying Tata group office locations and detailed building data, including occupancy and energy consumption, from an **ElasticSearch** Database. This switch to Bing Maps enhanced data access speed and **reduced costs by 15%**.
- Designed and implemented automated data loading system using **SSIS** packages, **C#** scripts, and **Tableau** reports for Bridgestone, transitioning from manual to automated inventory data processing, accelerating **processing time by 55%**.
- Spearheaded the development of an inventory data analysis system using **ETL** tools and **Tableau** reports, enabling real-time operational monitoring of products. This initiative contributed to a **12% increase** in monthly revenue.

EDUCATION & CERTIFICATIONS

Master of Science in Computer Science - University of South Florida
Bachelor of Engineering in Computer Science and Engineering - Rajiv Gandhi Technical University

TECH STACK

PYTHON | PANDAS | NUMPY | JAVA/J2EE | PYSPARK | SQL | SPARKSQL | FASTAPI | SSIS | ETL | TABLEAU | AZURE ML | SPARQL | APACHE AIRFLOW | DAGSTER | AZURE DATA FACTORY | MONGODB | HADOOP | MULESOFT | UNIX/LINUX SHELL SCRIPTING | AZURE (BLOB STORAGE, DATAFACTORY, COSMOS DB) | AWS (S3) | AWS EMR | GIT

SOFTWARE SKILLS

APPLICATION SERVERS/ENVIRONMENTS: AZURE WEB APP | TOMCAT | **CLOUD**: AWS| AZURE | **DATABASES**: COSMOS DB | MONGO DB | ORACLE | MYSQL | POSTGRES | SQL SERVER | **VISUALIZATION TOOLS**: TABLEAU | AZURE DASHBOARDS | OPEN PIM | **ML TOOLS**: AZURE ML | **DEVELOPMENT ENVIRONMENTS**: VISUAL STUDIO | ECLIPSE | JUPYTER NOTEBOOK | **VERSION CONTROL**: GITHUB | BITBUCKET | **BUILD TOOL**: APACHE AIRFLOW | DAGS | WORKFLOW ORCHESTRATION | DAGSTER | TALEND STUDIO | **SECURITY TOOLS**: SNYK | **CODE ANALYSIS:** CODACY | **TESTING TOOLS:** POSTMAN | **PROJECT MANAGEMENT:** JIRA | **METHODOLOGIES**: AGILE | SCRUM | DEVOPS | WATERFALL | KANBAN

AREA OF EMPHASIS

STRATEGIC SOLUTION-FOCUSED INITIATIVES| DATA MIGRATION & IMPLEMENTATION | DATA PIPELINE DEVELOPMENT| BIG DATA TECHNOLOGIES| DATA WAREHOUSING| DATA MODELING | DATABASE MANAGEMENT| ETL (EXTRACT, TRANSFORM, LOAD)| CLOUD PLATFORMS| DATA QUALITY AND GOVERNANCE| DATA ANALYTICS AND VISUALIZATION| MACHINE LEARNING AND AI| PROJECT SCOPING & ESTIMATION| LEADERSHIP &

COACHING| RELATIONSHIP BUILDING WITH INTERNAL & EXTERNAL CLIENTS| RESEARCH & ANALYSIS| SOFTWARE DESIGN/QUALITY ASSURANCE | COLLABORATION AND COMMUNICATION