Prasuna Challa

challap@usf.edu • 813-705-3576
LinkedIn • GitHub • Tampa, FL, 33613

Experienced Data Engineer with 4+ years in advanced data solutions and ETL optimization. Led implementation, increasing productivity by 30% and cutting costs by 20%. Proficient in AWS, Azure and GCP, driving data-driven insights and operational efficiency.

Technical Proficiencies & Certifications

- **Languages & ML:** Python (Libraries-pandas, NumPy), ASP.Net, ADO.Net C#, PySpark, R Language, Bash Shell Scripting, Web scraping (Selenium, BeautifulSoup, ScraPy), statistical modeling & machine learning (TensorFlow, Keras, PyTorch, Scikit-Learn, Langchain).
- **Tools & Visualization:** Tableau, Power BI, MS SQL Server, PostgreSQL, MS Office, Azure synapse, SAS Enterprise Miner, Databricks, Git, Bitbucket, DBT SQL, Snowflake, Data Bricks, Azure Datafactory, AWS services- Redshift, S3, Lambda, Google Cloud Platform Services - GCS, Bigquery, Firestore, DataFlow, Pub/Sub, Kafka, Artifact registry, Airflow, IAM, Cloud Functions, Jira, Confluence.
- **Databases and Bigdata Technologies:** MySQL, MongoDB, Cassandra, Hadoop, Spark, Spark SQL, and Hive.
- **Certifications:** AWS Cloud Practitioner, GCP Associate Cloud Engineer, Astronomer Certification for Apache Airflow Fundamentals.

Education

Master in Business Analytics and Information Systems	University of South Florida	Tampa, FL	**Aug 2022- May 2024**.
Bachelor of Electronics & Communication Engineering	Sree Vidyanikethan Engineering College	Tirupati, IN	**Jul 2015 - Apr 2019**.

Career Experience

Graduate Assistant | Data Engineer, University of South Florida, Tampa   **Nov 2022 – May 2024**
- Architected and deployed an efficient ETL system integrating Canvas course data using Azure services like **ADL Gen2, ADF, DataBricks**, and **SODA** API. Managed 4 courses, resulting in a 40% boost in data accuracy and a 30% decrease in processing duration.
- Implemented automation to detect and filter inactive students, resulting in a 20% increase in student engagement and targeted email notifications to instructors.
- Harnessed **Canvas LMS API** to extract detailed student data and analytics, facilitating intervention strategies and personalized support initiatives. This resulted in a 25% increase in student retention and a 30% improvement in course completion rates.

Data Engineer | Freelance   **Jul 2022-Oct 2022**
- Designed an ETL pipeline for a multi-location restaurant group using **Azure Data Factory**, automating data extraction and loading from SFTP to Azure Synapse, ensuring 100% data availability within a 7-day window.
- Implemented error-handling mechanisms with automated retries & alerts, ensuring data integrity. Developed a real-time monitoring dashboard, tracking ETL process status & failures, enhancing operational efficiency.
- Devised Google Cloud Functions to process daily CSV/json uploads, appending data to BigQuery, achieving 99% pipeline reliability with real-time failure monitoring.

Data Engineer, Infosys, Hyd   **Oct 2019 – Jun 2022**
- Orchestrated the automation of over 70 data pipeline jobs with Python and **Airflow DAGs**, enhancing operational efficiency; Leveraged **Firestore, Pub/Sub, Dataflow**, GCS, and **Composer, Stack Driver** for advanced Meta Data Management and seamless data processing
- Optimized SQL logic through **DBT**, boosting cost-effectiveness by 90% and enhancing data lineage, accessibility, and system performance. Proficient in Azure technologies, including **Data Bricks, Event Hubs, Cosmos DB, Data Factory, Azure Functions**. Experienced in Python development frameworks like **Flask**, and **Fast API**; skilled in writing and utilizing **Databricks** Notebooks.
- Enforced automated anomaly detection and **Six Sigma** analysis via **Cloud Functions**, validating proofs-of-concept for onboarding 3 new regions and 2 e-retailers.
- Pioneered the creation and execution of **KPIs** such as CTR, conversion rates, ROAS, and cart abandonment rates; employed data dictionaries to ensure precise SQL logic, enabling data-driven decisions and contributing to a 15% increase in e-commerce campaigns.
- Built Cloud **Data Fusion** pipelines to transform API responses, exporting over 10TB of data to external servers, streamlining data synchronization, and enhancing data availability by 40% for downstream processing.
- Directed CI/CD deployments via GCP **Artifact Registry** for 5 significant data infrastructure ventures, guaranteeing smooth updates with zero downtime and achieving a 30% enhancement in deployment frequency.
- Deployed **Docker** images using **Cloud Run**, achieving 99.9% uptime and reducing deployment time by 40%.
- Restructured processes and leveraged automation tools & drove a significant 25% enhancement in deployment efficiency. Earned 'The Insta Rise Award' and garnered client praise for adept task execution.

Associate Data Engineer, Tata Consultancy Services, Chennai, TN   **Jun 2019 – Sep 2019**
- Designed and maintained scalable data pipelines using AWS services (including **S3, EC2, IAM, Athena, Glue, Lambda** functions, **SNS**, and **Redshift**) and Databricks, optimizing data flow and storage to reduce processing time by 30%.
- Engineered and optimized ETL processes using PySpark on **Databricks** and **Kafka** for real-time analytics and led successful migration of legacy systems to AWS, resulting in a 20% cost savings and enhanced data availability for data-driven decision-making.

Academic Projects
- Detect AI Generated Text.   **Jan 2024 - Mar 2024**
- ChatBot for MSBAIS Program – University of South Florida.   **Sep 2023 – Dec 2023**
- Object detection for exam monitoring Application.   **Sep 2023 – Dec 2023**