

**CERTIFICATIONS**[Data Science Professional Certification](#)[AWS Certified Data Engineer – Associate](#)[AWS Certified Cloud Practitioner certificate](#)**WORK EXPERIENCE****The University of South Florida (USF)**

Tampa, Florida

*Teaching Assistant (TA)*

Aug 2024 - Dec 2024

- Guided over 150 learners through advanced topics in vulnerability assessments and ethical hacking methodologies; feedback from participants indicated increased confidence levels by an average of 2 skill rating points after each session.
- Mentored over 60 students in software development, delivering tailored feedback on project designs and debugging challenges that directed coding skills and improved overall comprehension of best practices across diverse projects.

**Research Assistant –The University of South Florida (USF)**

Tampa, Florida

*Advisor: Dr. Yicheng Tu*

Aug 2024 - Dec 2024

- Developed innovative vector embeddings, which enhanced data representation accuracy across biological datasets ranging from 3.5 million to 10 million records, optimizing the overall analytical workflow for diverse research applications.
- Executed GPU-accelerated workflows utilizing CUDA to streamline genetic pipelines, resulting in a significant 40% reduction in data processing time while maintaining high accuracy across large biological datasets.

**Tata Consultancy Services (TCS)**

Hyderabad, India

*Data Engineer*

Aug 2022 - Jul 2023

- Engineered automated ETL pipelines using AWS Glue, Apache Spark, and Delta Lake integration, optimizing data ingestion processes for over 1TB of data daily, improving scalability, and reducing manual tasks by 30%.
- Deployed cost-effective data storage solutions with Amazon S3, AWS Lambda, and other serverless architectures, achieving a 20% cost reduction while maintaining 99.9% availability and reducing retrieval times by 15%.
- Analyzed and optimized batch and streaming data pipelines leveraging Python (Boto3), PySpark, and AWS Glue, improving processing speeds by 35% and reducing pipeline execution times from 8 hours to under 3 hours.
- Pioneered a scalable, real-time data processing framework on AWS services, including Amazon Kinesis and AWS Step Functions, processing over 100GB of data per hour and delivering insights within 5 minutes.
- Implemented a data governance strategy using AWS Lake Formation, AWS IAM policies, and encryption protocols, reducing access request resolution time by 25%, improving compliance accuracy to 99%.

**SKILLS****Programming Languages & Scripting:** Python (including Boto3), Java, SQL (T-SQL, PL/SQL), Scala, R, JavaScript, Bash**Data Engineering & Processing:** Apache Spark, Kafka, Airflow, NiFi, Informatica, ETL/ELT Pipelines, Batch & Real-Time Data Processing, Data Warehousing (Snowflake, Redshift), Data Lake Architecture, Data Infrastructure, Data Architecture, Hadoop**Cloud Platforms & Services:** AWS (S3, Glue, Redshift, Lambda, Kinesis, Athena, DynamoDB, CloudFormation), Azure (Databricks, Data Factory, Synapse Analytics), Snowflake**Data Management & Analytics Tools:** Tableau, Power BI, Presto, Hive, Delta Tables, data modeling (star and snowflake schemas), databases, data quality, validation tools, Relational/non-relational Databases, Data Extraction, and Data Manipulation, datadog**Development & Deployment Tools:** Git, Docker, Kubernetes, Jenkins, Terraform, CI/CD pipelines, Cloud Monitoring and Logging Tools**Machine Learning & AI Integration:** TensorFlow, PyTorch, Scikit-learn, Feature Engineering, Predictive Analytics, Advanced Analytics Algorithms, kubernetes**EDUCATION****Master of Science in Computer Science:** The University of South Florida (USF)

Aug 2023 - Present

**Bachelor of Engineering in Computer Science:** IIIT Nuzvid (GPA: 3.69 )

Aug 2018 - May 2022

**PROJECTS****PySpark Project: End-to-End Data Analytics and Visualization**

Tech Stack: PySpark, Databricks, Python, Big Data Analytics, KPI Calculation, Dashboard Development, Azure Data Lake Storage, Delta Lake, Power BI, Delta Tables, Presto, Hive

- Crafted flexible pipeline frameworks leveraging PySpark, Delta Lake, and Azure Data Lake Storage to process over 2 million entries, reducing pre-processing time and improving performance metrics on 10GB datasets.
- Designed interactive Databricks dashboards integrated with Power BI, modernizing KPI computations and enabling faster data-driven decisions with a 30% increase in reporting efficiency.

**Azure Databricks Project: End-to-End Data Pipeline with Delta Lake Integration**

Tech Stack: Azure Databricks, Delta Lake, PySpark, Power BI, Data Lake Storage, Big Data Analytics, Azure Data Factory, Azure Synapse Analytics, Azure Event Hubs, Azure Blob Storage, Azure Functions, Azure Key Vault

- Devised streamlined ETL processes leveraging advanced automation in PySpark and Azure Data Factory for handling large volumes exceeding 5 million records; slashed required preparation time by approximately 50%.
- Rationalized Delta Table queries using Azure Synapse Analytics, achieving a 60% reduction in execution times; developed comprehensive Power BI dashboards that enhanced analytics workflows and boosted reporting efficiency by 40%.