Saideep Taddi
Data Analyst

Denton, Texas | (940) 231-4270 | https://www.linkedin.com/in/saideep-taddi-49a8b12a2/ | taddisaideep@gmail.com

SUMMARY

- Accumulated as a **Data Analyst** having over 3+ years of experience in data analysis, leveraging **Python** and **SQL** to automate workflows, build 3+ ETL pipelines, and data extraction.
- Proficient in migrating legacy systems to **AWS Cloud,** deploying scalable pipelines and integrating **CNN/LSTM** models to improve forecasting accuracy for business use cases.
- Expertise in handling large datasets and real-time streaming with **Apache Spark, Hadoop, Kafka, MapReduce,** and **PySpark**, processing data.
- Developed interactive dashboards and visualizations using **Tableau** and **Power BI**, enabling data-driven decision-making and improving reporting efficiency.
- Engineered **ETL workflows** (**Informatica, Apache Nifi**) to process datasets, reducing latency by 2x via **Hadoop/Spark** optimizations across **PostgreSQL, MySQL**, and **NoSQL** databases.

TECHNICAL SKILLS

Programming & Scripting:	Python, SQL, R
Methodologies:	Agile, Waterfall
Data Visualization:	Power BI, Tableau, Microsoft Excel, Looker Studio
Version Control Tools and Others:	Git, GitHub, Docker, Jenkins, Terraform, Visual Studio Code
Databases:	MySQL, SQL, Oracle, MongoDB, NoSQL, Cassandra
ETL Tools:	Informatica, Apache Airflow, SSIS, Snowflake, Talend, Apache Nifi, Spark
Data Analysis:	Data Mining, Reporting, A/B Testing, Data Warehousing
Big Data & Machine Learning:	Hadoop, Spark, Cassandra, Kafka, HDFS, MapReduce, ML (CNN, LSTM)
Libraries and Frameworks:	NumPy, Pandas, SciPy, PySpark, Matplotlib, Flask, Django, Bootstrap
Cloud Platforms:	AWS (S3, EC2, Redshift, IAM, Glue, Athena, QuickSight), Azure (Databricks, Data Lake, Data Factory, Azure Synapse Analytics)

PROFESSIONAL EXPERIENCE

MERCK   **Jul 2024 - Current**
Data Analyst   TX, USA

- Analyzed electronic health records (**EHR**) and claims data with **Python** (Pandas, NumPy) to identify workflow bottlenecks, reducing claim processing time by 8 hours weekly.
- Built and maintained **Azure** (**Azure VMs, Azure Blob Storage**) data pipelines to process healthcare data, cutting query response time from 5 minutes to 2 minutes.
- Secured 500M+ patient records and claims data in **HIPAA-compliant Azure Storage systems** with encryption and compliance safeguards to ensure privacy and regulatory adherence.
- Facilitated **Power BI** dashboards integrating data from EHR systems, billing platforms, and other sources to monitor KPIs like patient wait times and ER overcrowding.
- Streamlined **ETL** workflows using **Azure Data Factory** reducing manual processing time from 12 hours to 4 hours weekly, ensuring seamless data ingestion and transformation.
- Created visualizations (**Seaborn**) to track population health trends across 10,000+ patient records.
- Deployed SQL queries for **Azure Database** for **MySQL** to generate reports on metrics like patient readmission rates, reducing report runtime from 10 minutes to 3 minutes.

Cognizant Technology Solutions   **Mar 2021 – Dec 2022**
Programmer Analyst/ Data Analyst   INDIA
Client: CVS Health

- Developed and implemented automated solutions using **Automation Anywhere 360**, resulting in a significant reduction in manual processing time and operational costs.
- Designed, configured, and maintained **RPA bots**, achieving a 30% improvement in workflow efficiency and accuracy.
- Conceptualized batch processing applications using **Spark APIs** and in-memory computing to automate healthcare data workflows, managing **5+ TB** of patient records stored in AWS S3-based Data Lakes.
- Crafted 10+ scalable data pipelines on **AWS (EMR, EC2, Glue, Lambda, Redshift, RDS)** to streamline member eligibility and provider network analytics, improving data accessibility for clinical analytics.
- Executed ETL/ELT workflows using **Redshift** transforming 10+ TB of raw healthcare data into structured insights supporting cost analysis and population health management.
- Built Python-based **ETL pipelines** to consolidate fragmented healthcare data into centralized **HIPAA-compliant** storage, ensuring seamless access for downstream analytics and reporting.

- Collaborated with **cross-functional teams** to identify and assess automation opportunities, leading to the successful deployment of **15+ RPA solutions** tailored to business needs.
- Provided **training and support** to **end-users**, facilitating a smooth transition, and ensuring effective utilization of automated systems.
- Coordinated with **stakeholders** to gather requirements and ensure alignment with project goals, delivering solutions that met client expectations.

Cipla   **Jul 2020 – Feb 2021**
Data Analyst   INDIA
- Pioneered **AWS Glue ETL pipelines** to automate **data integration** across clinical trials and supply chain systems, cutting manual processing time by 4 hours per dataset and ensuring uninterrupted data flow.
- Executed **predictive analytics** (**SQL**) on 6,000+ anonymized patient records to uncover drug efficacy patterns, informing strategic decisions for drug development initiatives, including dosage optimization.
- Leveraged legacy **SQL queries** and **Tableau** data sources, boosting report generation efficiency by enhancing usability for 12+ global stakeholders in regulatory and supply chain teams.
- Streamlined **Salesforce data extraction** workflows using Django Celery tasks, improving report accuracy and timeliness for 50+ healthcare partners, reducing discrepancies in monthly compliance audits.
- Formulated **Tableau dashboards** to visualize real-time sales performance and patient adherence metrics, enabling stakeholders to resolve operational inefficiencies and save 2.5 hours weekly in the process.
- Collaborated in **Agile sprints** to align analytics deliverables with **FDA** compliance timelines, reducing task completion time by 1.2 hours per sprint through improved backlog prioritization.
- Devised **AWS Redshift** data warehouses to consolidate pharmacovigilance data, achieving <500 ms query latency and maintaining HIPAA-compliant storage for 1M+ patient records.

EDUCATION

University of North Texas   Denton, TX
Masters in **Advanced Data Analytics**   **Jan 2023 – May 2024**

Lovely Professional Institute   Pune, INDIA
Bachelors in **Computer Science & Technology**   **Aug 2017 – Jul 2021**

PROJECTS

California House Price Prediction with R Shiny:
- Formulated interactive **R** Shiny web app for predicting California house prices.
- Conducted comprehensive **data exploration** and implemented interactive map functionality.
- Utilized **machine learning** for **predictive modeling** and created intuitive user interface.

Amazon Product Sales Analysis with GCP and Looker Studio:
- Analyzed Amazon sales data using **GCP** and **Looker Studio**.
- Leveraged **Big Query** for data extraction and developed visualizations in **Looker**.
- Enabled strategic decision-making, leading to improved profitability and customer satisfaction.

Analysis of CMS Data for Healthcare Insights:
- Analyzed **CMS** data for over 3,000 U.S. hospitals, covering more than seven million discharges categorized by **Medicare Severity Diagnosis Related Group** (DRG).
- Identified **heart-related conditions** (e.g., heart transplants, heart assist systems) as the most frequent and costly diagnoses.
- Provided actionable insights, recommending increased production of **heart-related drugs**, specialized doctors, and medical equipment.
- Suggested preventive measures, such as **healthy lifestyle** choices and regular monitoring of risk factors, to reduce the prevalence and cost of heart-related conditions.
- Utilized **AI** and **machine learning techniques** to predict future medical concerns, improving drug accessibility for a wider population.

CERTIFICATIONS
- C and C++: Internshala Trainings
- SQL: Udemy
- HTML & CSS: Udemy
- JavaScript: Udemy
- R-Language for statistical analysis: Data Camp