

Akshay Raj Pallerla

akshayraj.pallerla@gmail.com | (860)-218-7219 | [linkedin](#)

Summary:

- Graduated in **May 2023** from the University of Connecticut with expertise in **data engineering, analytics, and machine learning** gained through rigorous coursework and real-world projects.
- Data professional with 7+ years of experience and expertise in cloud technologies, streamlining ETL processes and optimizing data processing for seamless deployment. Successfully delivered production-ready projects, meeting all requirements and timelines, resulting in enhanced efficiency.
- Skilled in database schema modeling and design across relational databases, including Oracle, SQL Server, and PostgreSQL. Proficient in migrating to NoSQL databases like MongoDB and Cassandra, ensuring flexibility and scalability.
- Experienced in implementing ETL processes using Apache Airflow, SSIS, and AWS services, with expertise in Synapse Analytics data integration, for data workflows and integration.
- Proficient in using Python to implement backend configurations and automation of workflows using various libraries like Pandas, NumPy, Matplotlib, TensorFlow, and scikit-learn, Plotly
- Managed projects using industry-standard tools such as Jira, Bit Bucket, GitLab, and Confluence, fostering seamless collaboration and achieving exceptional results.
- Expertise in creating data streaming using **PySpark/Spark Streaming, Kafka, and Apache Airflow**
- Demonstrated expertise in big data technologies like Hadoop, hive, Sqoop to provide data solutions for managing data processing systems, data storing and processing predicaments
- Extensive experience in PySpark and Python, leveraging techniques to optimize data processing, improve performance, and drive efficient data transformations.
- Proven ability to train and mentor junior peers on application flows, team expectations and standards, cloud technologies, fostering a culture of continuous learning and driving innovation in data engineering practices.
- Skilled in data migration, data cleansing, and data profiling, ensuring data accuracy, integrity, and quality throughout the data warehousing process.
- Briefly worked with Azure services (Azure VMs, Blob Storage, Azure functions, Azure Synapse, Azure Data Factory, Azure Data Factory and Data Lakes)

Technical Skills:

- *Programming languages* : Python, R, SAS, MS SQL, PL/SQL, MySQL, Linux, Vertica
- *Databases* : Snowflake, Oracle, SQL Server, Cassandra, MongoDB, Postgres
- *ETL and Analytics* : SSIS, Airflow, AWS, SSRS, Power BI, Tableau, Terraform, JMP
- *Methodologies* : Agile, Scrum, Waterfall, SDLC
- *Big Data* : Hadoop (Hive, HDFS, Pig, MapReduce), PySpark, Databricks
- *Design Tools* : Microsoft Visio, Lucid, MS-Excel

Education:

- MS in Business Analytics & Project Management, **University of Connecticut** (Completed: May 2023)
- Bachelors in Electronics and Computer Engineering, **JNTU-H** (Completed: May 2017)

Work Experience:

University of Connecticut

Graduate Assistant Data Steward (Feb 2022- Apr 2023)

- Streamlined data integration processes by implementing automated ETL workflows, reducing data processing time by 40% and improving data accuracy by 25%.
- Developed and maintained financial data models and databases, enabling efficient storage and retrieval of financial data for budgeting, forecasting, and financial analysis purposes.
- Collaborated with cross-functional teams, including finance analysts and administrators, to gather data requirements and designed and implemented robust data pipelines for data availability in financial reporting.
- Led the implementation of a data governance framework for financial data, ensuring compliance with regulatory requirements and improving data security and privacy controls.
- Created interactive financial dashboards and reports using data visualization tools, such as Tableau,

enabling finance stakeholders to gain actionable insights from financial data, resulting in a 15% improvement in data-driven decision-making.

- Enhanced financial data quality assurance processes, implementing data validation checks and anomaly detection mechanisms, resulting reduction in data errors.

Accenture (Client: Charter Communications)

Sr. Data Engineer (Dec 2018- Sep 2021)

- Orchestrated and developed ETL pipelines using AWS Glue for automated data cataloging, and Python in the backend for data loads in downstream, reducing 10% the DOJ call-in rate for field technicians in the last quarter.
- Applied transformations like harmonization, deduplication, Change data capture using Python and lambda functions on usage data extracted from Athena ensuring accurate dataloads to AWS S3 buckets and AWS redshift.
- Implemented tech-mobile project re-architecture, providing guidance to the team throughout the process in creating intermediate tables for ad-hoc analysis with good scalability to allow stakeholders to query historic data based on a UID.
- Developed 9+ Tableau reports for digital platform insights analyzing call rates, work-order volumes, and other Tableau dashboards such as Techmobile score card for analyzing metrics on quantified issues and activation attempts.
- Utilized Prism to perform statistical analysis, identifying trend patterns, and API tagging based on event actions and API definitions used in data products while defining metrics for performance reports of ad campaigns through collaboration with cross-functional teams
- Delivered recurring production reports and Tableau dashboards on viewer metrics, engagement metrics, region-wise content performance, and financial analytics for Connectivity Insights product and collaborated with the team for feature enhancements deployed on a 30 million internet customer base.
- Automated infrastructure configurations using Terraform for the Data Science team to handle tech-Mobile application data and support workforce planning initiatives.
- Trouble-shooted data issues through monitoring, log analysis, data exploration, and analysis on large volumes for Techmobile and digital-ad insertion product lines using Elastic search in the Kibana tool.
- Supervised a team of 7 ETL developers in an Agile-scrum setup and engaged with stakeholders for supporting 5 product lines (IP video, tech-mobile app, set-top box, digital ad insertion, spectrum news)
- Leveraged technical expertise to support data-related aspects of the ad-insertion process in Business Intelligence applications by integrating the data from different sources before the ETL process.

Accenture (Client: Volkswagen: VW Credit, Inc)

Data Engineer (Jun 2017- Nov 2018)

- Developed stored procedures, functions, and views in SQL Server to perform complex calculations for banking applications like ROA/ROE, Loan to Deposit Ratio (LDR), Net-Interest Margins, Cost of Funds.
- Improved ETL solutions in SSIS, reducing the execution time for over 30 data products by 17% with optimized data retrieval & processing using transformations such as merge/anti joins, Derived Columns.
- Plunged dashboard refresh time by 12% through customer data unification by revamping older Tableau dashboards and optimizing SQL queries using CTE, indexing, ingestion scripts, and user-defined functions to track KPIs, ensuring uptime to stakeholders.
- Streamlined repetitive data collection, processing, & analysis through data pipelines to intermediate S3 buckets using AWS Lambda & Python scripts, resulting in a 30% reduction in processing time.
- Leveraged PySpark's distributed computing and DBFS for efficient file storage, effectively managing scalable Spark clusters with Databricks to drive significant, large-scale data transformations.
- Implemented self-service EMR clusters for trigger ETL jobs in a test environment using git hooks, saving 5 hours of manual efforts for testing data products.
- Maintained Snowflake databases, schemas, tables, views, and stored procedures, ensuring data consistency and accuracy throughout the data pipeline.

MyVerkoper

Data Analyst (Jun 2016- Apr 2017)

- Produced monthly performance reports using sales and marketing data, with automated email delivery for authorized users. Monitored KPIs: sales revenue, customer satisfaction, acquisition costs.
- Spearheaded efforts to enhance data accuracy and accessibility, resulting in a notable 30% reduction in manual data processing. Achieved this by streamlining the ETL process into Power BI dashboards.

- Pioneered the development of a real-time Power BI dashboard, that saved 10 hours per week in manual reporting tasks. This empowered effortless identification and analysis of company KPIs.
- Improved data retrieval efficiency by 25% through optimized procedures, mastering complex SQL queries (CTE, indexing, no locks, joins) for historical data analysis from data warehouses.
- Executed the creation of SSRS reports derived from SQL Server Analysis Services (SSAS) cubes. Leveraged DAX scripts to effectively query data from OLAP cubes.
- Collaborated with cross-functional teams to design data integration solutions that met the organization's needs by understanding business requirements and identifying data sources.

Projects:

Capstone Project | Python, Power-BI, SAS JMP: (Payout annuity study : LIMRA)

- Developed a predictive model in Python to predict annuity holder longevity which achieved an accuracy of 77% and developed Power-BI dashboards for insights on retirement and life insurance to provide recommendations for effective business strategies.

Product Return Process | Oracle SQL, ERD & Swim Lane:

- Created Microsoft Visio-based business process and data model to detect e-commerce platform return fraud, with an analysis query-enabled database to flag damaged orders with insights to identify gaps in business processes, enabling informed decision-making.

Humana Housing Insecurity Prediction | Python- Plotly, NumPy, Pandas, Scikit-Learn, Matplotlib

- Developed a data model and defined KPIs using 800+ variables for identifying behavior & housing insecurity risk in members, & achieved an AUC of 0.73, to propose health-improving solutions.

Netflix Movie Recommendation System| Python- NumPy, Pandas, Scikit-Learn, Matplotlib

- Designed a movie recommendation system within Netflix through correlations in Python by dealing with over 40M+ rows accounting for over 17K movies and 500K customers data to recommend best shows to users based on self & others behavior

Data science and Time Series Forecasting | Walmart | SAS JMP, NLP, Tableau:

- Utilized ARIMA & ARIMAX models in SAS Studio to forecast weekly sales, interpret trends, and seasonality for 45 Walmart stores in the US, and offered department-specific product availability recommendations based on 10 years of data.