Hemanth.k

Data Engineer
Mobile: +1 (928) 255-7216
Mail: hk485@nau.edu
LinkedIn: linkedin.com/in/hemanth-kurapati1021/
GitHub: github.com/Kurapati1309

PROFESSIONAL SUMMARY

- Data Engineer with around 5 years of experience in designing, developing, and deploying end-to-end data engineering solutions across Various domains.
- Experience in data pipelines, cloud platforms, and BI dashboards, with strong expertise in **financial reporting, cost forecasting, and variance analysis** to support decision-making and ensure **data accuracy and compliance**.
- Experienced in **A/B testing, customer segmentation, personalization modeling, and LTV prediction** using Python, SQL, and Databricks to support data-driven product optimization.
- Skilled in building scalable **ETL** pipelines with **Python, SQL,** and **Apache Spark** to process high-volume structured and unstructured data.
- Hands-on experience with **AI-powered analytics for call center operations**, tracking KPIs such as AHT, CSAT, containment, and FCR.
- Designed and executed **A/B testing experiments** to evaluate model-driven recommendations and feature enhancements, providing actionable insights to product teams.
- Built **customer segmentation and personalization models** using Python (scikit-learn, pandas) to optimize engagement and retention.
- Implemented **customer lifetime value (LTV) prediction models** in Databricks, improving targeting accuracy for marketing campaigns.
- Proficient in designing and managing cloud-based data infrastructure using AWS (S3, Redshift, Glue) and Azure, enabling scalable, production-ready deployments.
- Experienced in data lake and warehouse architecture, building schemas, marts, and optimized pipelines to reduce query latency by 30–35%.
- Applied data governance, compliance, and security frameworks (GDPR, CCPA) with validation, encryption, and monitoring to ensure reliability and regulatory adherence.
- Automated infrastructure and workflows using Terraform and dbt, improving scalability and reducing manual setup time.
- Proficient in building **Power BI dashboards** to monitor AI-driven initiatives (virtual agents, intelligent routing, automation) and deliver actionable insights.
- Applied **NLP, predictive modeling, and machine learning** in Databricks and Azure to optimize customer experience and automate workflows.
- Strong ability to translate **complex datasets (CRM, IVR, chatbots, call logs)** into business insights that support AI transformation strategies.
- Strong background with big data ecosystems (Databricks, Delta Lake, Hadoop) for distributed data processing and machine learning integration.
- Built monitoring and alerting systems for batch and streaming jobs with AWS CloudWatch, ensuring SLA adherence and proactive issue detection.
- Optimized relational (PostgreSQL, MySQL) and NoSQL (MongoDB, Cassandra) databases, tuning queries and indexes for performance improvements.
- Delivered predictive and forecasting models (ARIMA, Prophet, Random Forest) integrated into pipelines to enhance planning and decision-making.
- Developed interactive dashboards with Tableau and Power BI, providing executives real-time visibility into KPIs and operational metrics.

- Collaborated closely with data scientists, analysts, and cross-functional teams, translating business needs into scalable engineering solutions.

- Recognized for mentoring peers in Spark optimization, ETL best practices, and cloud data workflows, enhancing overall team capability.

TECHNICAL SKILLS

Programming Languages	Python (Pandas, NumPy, SciPy, Scikit-learn, PySpark), R, SQL, Java, C++, JavaScript, HTML/CSS
Statistical/ Analytical Skills	Predictive Modeling, Time Series Forecasting (ARIMA, Prophet), Feature Engineering, Hypothesis Testing, Regression Models, Random Forest, Logistic Regression, Financial Reporting, Variance Analysis, Budget Forecasting, Data Accuracy & Validation, Compliance (GDPR, CCPA, HIPAA basics).
Big Data & Distributed Systems	Apache Spark, Databricks (Predictive analytics, NLP pipeline integration) , Hadoop (HDFS, MapReduce), Hive, Kafka, Delta Lake
Databases	Relational (PostgreSQL, MySQL, SQL Server, Oracle), NoSQL (MongoDB, Cassandra), Data Warehousing (AWS Redshift, Snowflake)
Cloud Platforms	AWS (S3, Glue, Redshift, EC2, SageMaker, CloudFormation, CDK), Microsoft Azure, GCP (basic exposure)
ETL & Workflow Management	Apache Airflow, dbt, AWS Kinesis (Data Streams & Firehose), Control-M, ServiceNow ITSM
Data Visualization & BI	Tableau, Power BI (Advanced DAX, Call Center KPIs visualization), Matplotlib, Seaborn, Plotly
MLOps & Deployment	MLflow, Terraform, Streamlit, REST APIs, Docker, CI/CD with GitHub Actions
Version Control & Collaboration	Git, GitHub, Bitbucket, Agile/Scrum methodologies, Jira
Development Tools & Environments	VS Code, PyCharm, Jupyter, Anaconda

WORK EXPERIENCE

Client: Minisoft Technologies LLC, Houston, TX **July 2025 - Present**
Role: Data Engineer

- Designed and deployed real-time ETL pipelines using Apache Spark and AWS Glue to process structured and unstructured data from APIs, IoT devices, and relational databases, reducing data latency by 30% and enabling near real-time business insights.
- Migrated relational and NoSQL datasets into AWS Redshift and S3, optimizing schemas and SQL queries to reduce query execution time by 35%, improving performance for analytics dashboards and executive reporting.
- Implemented a data governance and compliance framework with automated validation rules, encryption protocols, and GDPR/CCPA controls, ensuring regulatory compliance and data integrity across all systems.
- Built and maintained data lakes and warehouses to consolidate enterprise datasets, improving scalability for analytics workloads and providing analysts with a unified source of truth.
- Developed SQL-based data marts tailored to business reporting needs, streamlining data access for analytics teams and reducing redundant data pulls by 25%.
- Automated infrastructure provisioning and orchestration using Terraform and dbt, improving environment consistency and reducing manual setup time for data pipelines.

- Partnered with data scientists and analysts to design custom analytics-ready datasets for modeling, accelerating time-to-insight for predictive analytics initiatives.
- Established monitoring and alerting systems for batch and streaming jobs using AWS CloudWatch, ensuring SLA adherence and proactive issue detection.
- Conducted performance tuning and index optimization in PostgreSQL and MongoDB, improving database query speeds by 20–25% for critical business applications.
- Delivered knowledge-sharing sessions for junior engineers on ETL best practices, Spark optimization, and cloud data engineering workflows, enhancing overall team capability.
- Designed **call center analytics dashboards in Power BI**, monitoring KPIs such as AHT, CSAT, containment, and agent productivity, enabling leadership to track AI-driven initiatives.
- Integrated multi-source datasets (CRM, IVR, chatbots, and call logs) into centralized models, validating AI outcomes and improving virtual agent and intelligent routing performance.
- Validated financial and operational datasets, resolving discrepancies and supporting **variance analysis** in executive dashboards and reports.

Environment: Python, SQL, Apache Spark, **Power BI**, Databricks, AWS Glue, Redshift, MongoDB, Terraform, dbt, Tableau, **NLP, Call Center KPIs (AHT, CSAT, Containment, FCR)**

Client: Raven Software Solutions Inc., Irving, TX **March 2025 – June 2025**
Role: **Software Developer**

- Designed and implemented enterprise-grade applications using Python and SQL to support multiple client business domains, automating workflows and reducing manual effort across departments.
- Developed automated reporting modules with SQL and Python, reducing report generation time by 40% and improving accuracy of insights provided to senior leadership.
- Collaborated in Agile sprint planning to design, build, and deploy new system features, ensuring timely delivery of enhancements aligned with client expectations.
- Created stored procedures, triggers, and scripts to manage large-scale client databases, improving reliability and streamlining recurring data operations.
- Conducted unit and integration testing in coordination with QA teams, identifying defects early in the cycle and improving release quality by 25%.
- Strengthened system security protocols by implementing access controls and encryption techniques, ensuring adherence to client confidentiality requirements.
- Authored technical documentation and user guides, improving knowledge transfer across teams and reducing onboarding time for new developers by 20%.
- Diagnosed and resolved production-level application issues, reducing downtime and improving SLA compliance across client systems.
- Partnered with project managers and stakeholders to prioritize feature requests, ensuring delivery of high-value enhancements on time.
- Conducted end-user training and workshops, supporting smooth adoption of newly deployed applications and improving client satisfaction.

Environment: Python, SQL, AWS, Apache Spark, Tableau, Power BI (Automated Reporting), Git/GitHub, Agile/Scrum, Security & Access Controls.

Client: BulkMagic (StableCoupons Inc.), New York, NY **Feb 2025 – March 2025**
Role: Data Engineer

- Built supply chain analytics pipelines integrating procurement, sales, and inventory data, enabling leadership to identify inefficiencies and improve order fulfillment rates by 15%.

- Designed and deployed predictive pricing models using Python and SQL, helping optimize bulk retail strategies and increasing pricing accuracy across product lines.
- Developed and validated **recommendation models** using customer purchase behavior and segmentation data to optimize cross-sell and upsell strategies.
- Conducted **A/B testing for pricing and promotional campaigns**, identifying data-driven strategies that increased user retention and engagement rates.
- Created interactive Tableau dashboards displaying KPIs, supply chain costs, and trends, providing executives with real-time insights for strategic planning.
- Leveraged predictive modeling and **AI-enabled automation** to optimize customer interaction strategies, improving response accuracy and reducing service handling times.
- Consolidated structured and semi-structured datasets into centralized data models, improving reliability of analytics and eliminating reporting inconsistencies.
- Applied time-series forecasting techniques (ARIMA, Prophet) to predict demand fluctuations, helping procurement teams plan inventory with greater accuracy.
- Automated data extraction, transformation, and validation scripts in Python, saving 15+ hours weekly while ensuring higher data consistency.
- Performed scenario and what-if analysis on procurement strategies, identifying opportunities for cost savings and risk mitigation.
- Partnered with leadership to embed analytics-driven practices into pricing and procurement decisions, directly impacting financial performance.
- Delivered executive-level reporting and presentations, improving transparency of business operations and influencing board-level decisions.
- Standardized reporting workflows and data documentation, ensuring reproducibility and consistency of insights across the organization.
- Improved **budget forecasting and cost analysis** accuracy by embedding predictive models into reporting pipelines, reducing financial variances and supporting leadership decision-making.

Environment: Python, SQL, AWS (S3, Redshift, EC2), Apache Spark, Tableau, Prophet, Databricks, Predictive Modeling, A/B Testing, Recommendation Systems, Customer Segmentation, AI-enabled Automation, Linux, Jira, GitHub.

Client: Northern Arizona University, Flagstaff, AZ **July 2023 – Dec 2024**
Role: Teaching Assistant

- Assisted faculty in delivering graduate-level coursework in Data Analytics, Databases, and Cloud Computing, ensuring students gained practical, hands-on knowledge.
- Mentored students in **Python, SQL,** and **R** programming, supporting projects on data analysis, visualization, and predictive modeling.
- Guided lab sessions on cloud platforms (**AWS & Azure**), teaching students how to build scalable data pipelines and deploy analytics solutions.
- Supported development of **Tableau** dashboards and data visualization assignments, improving student proficiency in BI tools.
- Designed and graded assignments, quizzes, and projects, ensuring alignment with course objectives and academic rigor.
- Guided lab sessions on **AI/ML use cases in cloud platforms (AWS & Azure)**, including deploying models for customer analytics and service optimization.
- Conducted review workshops and one-on-one sessions, helping students strengthen problem-solving and technical skills.
- Collaborated with professors to refine course content, integrating modern data engineering and machine learning practices.

- Provided technical troubleshooting during lab sessions, resolving issues in programming environments and cloud setups.
- Assisted in research projects, contributing to data cleaning, preparation, and visualization tasks.
- Promoted academic integrity and supported a collaborative learning environment across classrooms and labs.
- Mentored students in data analysis and visualization, including **financial/operational reporting projects** using Tableau and SQL.

Environment: Python, SQL, R, AWS, Azure, Tableau, Jupyter, Git, AI/ML labs and customer analytics use cases.

Client: Center For Human Security & Studies, India **Feb 2021 – Dec 2022**
Role: Data Analyst

- Designed and deployed Tableau dashboards to visualize engagement metrics and operational KPIs, enabling leadership to make evidence-based policy decisions.
- Integrated CRM data (Salesforce, HubSpot) with Google Analytics and Hotjar, creating a unified dataset that improved understanding of stakeholder engagement trends.
- Conducted predictive analytics and forecasting with statistical models, supporting policy interventions and improving planning accuracy by 20%.
- Implemented automated KPI tracking systems in ServiceNow, reducing downtime incidents by 25% and improving system reliability.
- Applied **NLP and text analytics** on stakeholder feedback datasets, generating insights that supported chatbot optimization and enhanced customer experience strategies.
- Performed data quality audits and validation checks across multiple systems, significantly improving accuracy and consistency of datasets used in reporting.
- Streamlined data integration pipelines for multi-source datasets, reducing manual consolidation efforts and ensuring faster reporting cycles.
- Authored and delivered analytical reports and insights to executive leadership, influencing operational strategy and long-term planning.
- Mentored junior analysts on data methodologies, visualization techniques, and statistical best practices, enhancing team productivity and knowledge.
- Researched and refined data collection methodologies, improving accuracy and efficiency in measuring KPIs across various programs.
- Collaborated with vendors and cross-functional teams to implement secure, compliant, and scalable data-driven solutions for organizational needs.
- Delivered **financial and operational insights** through KPI dashboards and variance reporting, enabling evidence-based decisions and resource allocation.

Environment: Tableau, Python, SQL, HubSpot, Microsoft Azure, ServiceNow, Google Analytics, Hotjar, NLP, Text Analytics, Chatbot Optimization.

Client: Indian Servers – Software Development, India **Jun 2021 – Sep 2021**
Role: Data Analyst

- Assisted in developing Python scripts for preprocessing and analyzing datasets, supporting client reporting and small-scale analytics projects.
- Designed SQL queries and stored procedures for ETL processes, improving data extraction and transformation efficiency.

- Supported deployment of applications on AWS EC2 and S3, gaining exposure to cloud-based infrastructure.
- Applied basic machine learning algorithms (Logistic Regression, Random Forest) on sample datasets to explore predictive modeling techniques.
- Collaborated with senior engineers to understand end-to-end project workflows, building foundational skills in data engineering and software development.

Environment: Python, SQL, AWS (EC2, S3), Machine Learning, C++, Predictive Modeling, Data Analysis

EDUCATION

Northern Arizona University, Flagstaff, AZ **Jan 2023 – Dec 2024**

Masters of Science in Information Technology

- Focused coursework in Data Science, Cloud Computing, and Machine Learning.
- Completed projects in data pipeline automation using AWS and Azure.
- Assisted in teaching labs on Python, Databricks, and Data Visualization (Tableau, Power BI).