**Kavya Guddeti**

**Professional Summary:**

* **Data Architect** with over **8.8 years** of strong experience in Data analysis, Data pipeline workflows, Data transformation, Data quality, Data assessment, Machine Learning, Data Modeling, Data warehousing, Data visualizations, Automation, and Multi-tier distributed applications.
* Proficient in utilizing cutting-edge analytical programs such as **Python** and **R**, along with the latest software tools, to drive efficient data processes.
* Developed **modular data pipelines** using cloud-native services (e.g., AWS Lambda, Azure Functions) to create flexible, scalable, and efficient data processing workflows.
* Designed and implemented **cloud-based data models**, ensuring alignment with business requirements and leveraging cloud services for scalability and flexibility.
* Extensive experience working with leading Datawarehouses and Databases, including **Snowflake**, **AWS** **Redshift**, Azure SQL, PostgreSQL, Mongo, MSSQL, Sybase and Oracle.
* Created data models compatible with multi-cloud environments, ensuring portability and flexibility in data management across various cloud platforms.
* Adept in leveraging **AWS Cloud services**, including Lambda, S3, CloudWatch, Glue, IAM, Databrew, Step Functions and CloudFormation, to design and implement robust data solutions.
* **Led successful data migration initiatives, managing data transfers between legacy systems and modern cloud platforms using AWS and Azure.**
* Experience in creating Datasets, Data manipulation, Preparation of reports, tables, listings & stylish graphs and also experienced in Forecasting, Predicting and analyzing any type of large datasets.
* Experience in developing **End to End** BI Echo systems using background in **RDBMS systems like SQL server, Oracle, Azure SQL** and ETL tools **like SSIS, DOMO, Alteryx** and BI tools **like Power BI.**
* **Optimized data architecture by applying modular principles, resulting in improved data flow, reduced redundancy, and enhanced system performance.**
* Experienced in using **Power BI Desktop, Power Query and Power BI service and writing complex DAX queries.**
* Experienced **in Performance Tuning, Debugging and Error Handling of ETL workflows** and implemented best practices while loading **data into downstream data feeds.**
* Experienced with SQL Server and **T-SQL in constructing Tables, Views, Indexes, Stored Procedures, Triggers and User Defined Functions.**
* Designed and implemented complex **data workflows and pipelines using Azure Data Factory**, orchestrating data extraction, transformation, and loading processes efficiently.
* Experienced in using **agile** methodologies including extreme programming, **SCRUM and Test-Driven Development** (TDD).
* Possess good interpersonal, analytical presentation skills, ability to work in Self-managed and Team environments, fast learner and an organized self-starter.

**Technical Skills:**

|  |  |
| --- | --- |
| **Databases** | MS SQL Server, Snowflake, Oracle, MySQL, Azure SQL, Postgre SQL, Sybase, Redshift |
| **NoSQL Databases** | MongoDB |
| **Programming Languages** | Python, SQL, Pyspark |
| **Cloud Technologies** | AWS- S3, Lambda, Glue, DataZone, DataBrew, IAM; Snowflake; Azure Data bricks,Athena |
| **ETL Tools** | SSIS, Alteryx, Domo, Glue, Informatica |
| **Reporting & Visualization** | Power BI, Tableau. |
| **Certifications** | Azure Data Engineer; Alteryx Core Designer, Power BI 300 |

**EDUCATION AND INTERNSHIP**

Rajiv Gandhi University of Knowledge Technologies (**IIIT**)-RK Valley, India Bachelor of Technology, Electronics and Communication Engineering; Minor in Managerial Studies

**Indian Space Research Organisation (ISRO)** - Sriharikota, India; **Project Engineering Intern on Space Science data**

**Professional Experience:**

**Client: Consolidated Food Co. (FMG – Chicago) Oct 2023 – Current**

**Role: Lead Data Engineer/Data Architect**

**Responsibilities:**

* Led end-to-end data engineering, including legacy file automation to AWS Cloud, with a focus on data governance, security, and quality using Glue, Lambda, Athena, S3, Redshift, Cloudwatch, lake formation etc.
* Designed a scalable, efficient, and secure AWS Cloud data processing architecture, emphasizing data governance principles.
* Engineered a comprehensive business catalog using Data Zone and Lake Formation, ensuring data governance compliance and security.
* Developed AWS-based pipelines automating API, JSON, and database data extraction etc, with a strong emphasis on data quality and security measures.
* Collaborated daily with stakeholders from CTOs to end-report users, prioritizing data governance, security, and quality standards.
* Trained business users on Amazon QuickSight for independent report creation, highlighting data governance practices and security considerations.
* Implemented IAM controls and other robust security measures to ensure compliance standards and enhance data governance.
* Proactively identified and resolved data pipeline issues, ensuring adherence to data governance principles, high security standards, and data quality.
* Successfully migrated legacy files to AWS, optimizing data storage and access with Spark capabilities, maintaining high data quality, and adhering to security, CI/CD and data governance standards.
* Review PySpark programs that are used to ingest historical and incremental data.
* Created comprehensive documentation for efficient team knowledge transfer, covering data governance, security, quality procedures, and Spark integration strategies.
* Implemented ongoing improvements based on user feedback and business needs, focusing on data governance, security, quality, and Spark optimization.

**Environment**: Power BI, SQL, Snowflake, Pyspark, Python, AWS Glue, Lambda, S3, Data Zone, Redshift Erwin, SQL Server Management Studio, Visual Studio 2017, EC2, StepFunctions, Athena, Quicksight

**Client: PepsiCo, IN May 2021 – Sep 2023**

**Role: Lead Data Science Engineer/Data Architect**

**Responsibilities:**

* **Led end-to-end data ETL migration­­**, overseeing cloud-native Data Mesh architecture design through development to create a decentralized and scalable data ecosystem.
* Applied domain-driven design principles to create **AWS - based data models**, ensuring alignment with business domains and fostering decentralized data ownership.
* Implemented and managed a comprehensive **business catalog on AWS**, utilizing services such as AWS **Glue, DataZone** and AWS Lake Formation.
* Expertise in interacting with multiple data storage types and products, such as S3,Postgres, Mariadb, MongoDB, **Redshift**, Snowflake, Azure Databricks.
* Familiarity with mlops concerns, including monitoring and alerting tools such as datadog,Logstash, CloudWatch, Prometheus, Grafana and similar.
* Implemented **modular data security** measures in cloud environments, ensuring that each module adhered to best practices for securing sensitive information using **IAM and DataZone**.
* Engineered business scripts for diverse computing platforms, developed multiprocessing data analytics pipeline tools in **Python** on **Azure/AWS Cloud**, handling extensive structured and unstructured data from diverse sources.
* Implemented **CI/CD pipelines** for cloud-based data applications, facilitating the seamless integration and deployment of modular components.
* Involved in Designing **Star Schema (identification of facts, measures, and dimensions), Snowflake Schema** for Data Warehouse, ODS architecture by using tools like **Data Model, ERwin.**
* Analyzed the **SQL scripts** and designed the solution to implement using **PySpark.**
* Developed **Python, PySpark** to Transform, and Load data across on premise and AWS cloud platform.
* Analyzed the **SQL scripts** and designed the solution to implement using **PySpark.**
* Used **SQL** queries and other tools to perform data analysis and profiling.

**Environment**: Power BI, SQL, Snowflake, Alteryx, Pyspark, Python, Azure Synapse Analytics, Azure Data Lake Storage:, Azure Event Grid, Azure Data Factory, AWS Glue, Lambda, S3, Data Zone, Redshift Erwin, SQL Server Management Studio, Visual Studio 2017

**Client: Optum, IN Jun 2019 – May 2021**

**Role: Data Engineer**

**Responsibilities:**

* Played a key role in formulating a roadmap for migrating on-premises Data Science infrastructure to the cloud, offering optimal solutions for executing data science tasks across multiple cloud platforms.
* Leveraged Azure services for data discovery and cataloging, enhancing the discoverability and accessibility of data assets across the organization.
* Engaged in **AWS tasks, including S3, Databrew, Glue, Lambda, Redshift, Lake Formation, IAM, DataZone and CloudWatch** as needed, contributing to the seamless integration of cloud services.
* **Migrated data from on-premises Oracle database to Snowflake using Python.**
* Orchestrated the implementation of **Azure Data Lake Storage** as part of the Data Mesh strategy, optimizing storage and retrieval of diverse data types.
* Implemented automated data pipelines within Azure Data Factory, ensuring the efficient flow of data across domains while maintaining autonomy.
* Modelled Alteryx Spatial Analytical apps which help in visualizing the geographic alignment and report on summary statistics (e.g., # of pharmacies, sales, and forecasting) for each territory and span of control for pharmacists
* Developed **Power BI/Tableau** data visualization using Cross tabs, Heat maps, Box and Whisker charts, Scatter Plots, Geographic Map, Pie Charts and Bar Charts and Density Chart.
* Prepared **ETL** design document which consists of the database structure, change data capture, Error handling, restart and refresh strategies.
* Created workflows, mappings using **Domo Magic ETL** and worked with different transformations such as lookup, source qualifier, update strategy, router, sequence generator, aggregator, rank, stored procedure, filter, joiner, sorter.
* Designed and implemented configurable data delivery pipeline for scheduled updates to customer facing data stores built with **Python.**

**Environment:** Alteryx, AWS, Power BI, Python, Snowflake, Domo, MS-SQL, Agile, AWS, Tableau, Azure

**Client: Accenture, IN Apr 2015 –May 2019**

**Role: Sr. Data Analyst**

**Responsibilities:**

* Worked as the Primary Point of Contact between the client and a team of 4 developers thus, helping the team understand the business requirements and project goals.
* Designed automatic performance dashboards in Tableau and revamped performance metrics of top 500 sales managers; implemented corrective criteria (controlled attrition, hiring, training) to increase the revenue of sales cohort by $0.5 M
* Created ETL data models and reports for supply chain team to help find the decrease in demand and to forecast the demand resulting in planned operations; improved data extraction by 15% and reduced inventory costs by 7%
* Developed 30+ complex SQL procedures to identify data mismatches between two databases and visualized data alignment
* progress in Tableau dashboard with enhanced faster decision-making. • Developed an Automated ETL jobs monitoring engine using SQL, Python and windows batch scripts that checks job failures and re-executes if it’s a known error else triggers email alert to team, saved 8hrs/day average team manual work.

• Supported ad-hoc client analytics requirements for sales and finance team using Cognos reporting. • Performed health check of complex SQL queries and automated operations for filtering and transformation of the raw data

**Environment: MS SQL Server, SSIS, SSRS, Power BI, Tableau, Python**

**PROJECTS**

**Millennial Leadership Project (MLP)** – Associated with Telangana state government (T-hub) and PepsiCo

• Data coordinator of start-ups and PepsiCo collaboration program for different problem statements and designed a Tableau dashboard for top executives to shortlist top 10 start-ups by domains, fundings and other metrics

**Data Rationalisation App** – Associated with Accenture Innovation Program

• Designed Alteryx analytical app which reads all data sources and identifies the repetitive metrics match%.

• This app helped to analyse in eliminating duplicate Tableau dashboards/software applications used across the team and saved total cost of maintenance and storage by 28%.