JANAKI RAMUDU DADI

Data Engineer

Location: AZ | Email: janakiramudu584@gmail.com | Phone: 984 223 7995

SUMMARY

- Experienced and results-driven Data Engineer with a proven track record of over 4 years of experience software development experience with expertise in Big Data, Hadoop Ecosystem, Cloud Engineering, Data Warehousing.
- Sound Experience with AWS services like Amazon EC2, S3, EMR, Amazon RDS, VPC, Amazon Elastic Load Balancing, IAM, Auto Scaling, Cloud Front, CloudWatch, and Lambda to trigger resources.
- Experience in building data pipelines using Azure Data Factory, Azure Databricks, and loading data to Azure Data Lake, Azure SQL Database, Azure SQL Data Warehouse to control and grant database access.
- Good experience with Azure services like HDInsight, Stream Analytics, Active Directory, Blob Storage, Cosmos DB, Storage Explorer.
- Strong Hadoop and platform support experience with all the entire suite of tools and services in major Hadoop Distributions – Cloudera, Amazon EMR, Azure HDInsight, and Hortonworks.
- Strong familiarity with GCP components, Google Container Builders, and client libraries, leveraging cloud SDKs for development efficiency.
- Led Proof of Concept (POC) initiatives to assess cloud offerings, including Google Cloud Platform.
- Hands-on expertise in GCP tools such as Cloud Shell SDK for configuring services like Data Proc, Storage, and BigQuery, ensuring optimal performance and scalabilit
- Proficient in handling and ingesting terabytes of Streaming data (Kafka, Spark streaming, Strom), Batch Data, Automation and Scheduling (Oozie, Airflow).
- Profound knowledge in developing production-ready Spark applications using Spark Components like Spark SQL, DataFrames, Datasets, Spark-ML and Spark Streaming.
- Expertise in developing multiple confluent Kafka Producers and Consumers to meet business requirements. Store the stream data to HDFS and process it using Spark.
- Strong working experience with SQL and NoSQL databases (Cosmos DB, MongoDB, HBase, Cassandra), data modeling, tuning, disaster recovery, backup and creating data pipelines.
- Experienced in scripting with Python (PySpark), Scala and Spark-SQL for development, aggregation from various file formats such as XML, JSON, CSV, Parquet.
- Great experience in data analysis using HiveQL, Hive-ACID tables, Pig Latin queries, custom MapReduce programs and achieved improved performance.
- Experience in monitoring document growth and estimating storage size for large MongoDB clusters as part of the data life cycle management.
- Hands-on experience on Ad-hoc queries, Indexing, Replication, Load balancing, Aggregation in MongoDB.
- Expertise in creating Kubernetes cluster with cloud formation templates and PowerShell scripting to automate deployment in a cloud environment.
- Sound knowledge in developing highly scalable and resilient Restful APIs, ETL solutions, and third-party integrations as part of Enterprise Site platform using Informatica.
- Highly involved in all facets of SDLC using Waterfall and Agile Scrum methodologies.
- Involved in migration of the legacy applications to cloud platform using DevOps tools like GitHub, Jenkins, JIRA, Docker, and Slack

WORK EXPERINCE

Data Engineer, *JPMC, AZ* ***Aug 2023-Current***

- Mastered a wide array of Azure technologies, including **HDInsight, Databricks, Data Lake, Blob Storage, Data Factory, Synapse Analytics, Azure SQL Database, and SQL Data Warehouse**, to streamline cloud operations and data management.
- **Designed and automated** custom input adapters using **Spark, Sqoop, and Oozie**, significantly enhancing data ingestion from RDBMS to Azure Data Lake, showcasing a blend of innovation and efficiency in data handling.
- Played a pivotal role in the creation of **automated workflows** for daily incremental data loads, facilitating seamless data transition from traditional RDBMS systems to more scalable data lake solutions, thus optimizing data availability and reliability.
- Executed comprehensive **ETL (Extract, Transform, Load) processes**, utilizing Azure Data Factory, T-SQL, Spark SQL, and U-SQL for Azure Data Lake Analytics, ensuring efficient and scalable data movement and transformation strategies.
- Engineered and deployed **data pipelines** in Azure Data Factory, targeting enhanced ETL processes from diverse sources such as Azure SQL, Blob storage, and Azure SQL Data Warehouse, thereby ensuring data integrity and accessibility.
- Crafted and deployed **JSON scripts** in Azure Data Factory, leveraging SQL Activity for efficient data processing and integration, highlighting technical proficiency and advanced deployment capabilities.
- Demonstrated deep understanding of data ingestion and orchestration using **Airflow Operators** and relevant Python libraries, enhancing data flow and operational efficiency.
- Analyzed SQL scripts and implemented solutions using **PySpark**, focusing on optimizing data processing and analytics workflows.
- Extensively used **Databricks notebooks** for interactive analytics, employing Spark APIs to facilitate advanced data analysis and insights.
- Integrated data across various sources including **MongoDB, MS SQL, and cloud platforms**, using Azure Data Factory, SQL API, and Mongo API, to ensure a unified data ecosystem.
- Enabled efficient data visualization and reporting through **interactive Power BI dashboards and reports**, providing actionable insights to decision-makers.
- Spearheaded the **Continuous Integration and Continuous Deployment (CI/CD)** of applications into Azure Cloud, enhancing operational agility and system reliability.

Data Engineer / Software Engineer, *Capgemini Technology Services Ltd., India* ***Jun 2019 - Aug 2022***

- Contributed to diverse projects in Data Engineering and Mainframes technologies at Capgemini Technology Services Ltd.
- Utilized key components within the Hadoop ecosystem such as Spark, HDFS, Hive, Sqoop, HBase, Zookeeper, and Oozie to develop various data loading strategies and performed transformations for analyzing large datasets.
- Implemented data warehousing with Hive and seamless data load and transform from Relational Database Systems to HDFS using Sqoop.
- Fine-tuned Spark applications to improve overall processing time for data pipelines.
- Applied advanced SQL queries including Joins, Correlated Subqueries, Views, Stored Procedures, and Triggers for data analysis.

- Demonstrated good knowledge of Hadoop architecture and various components, including HDFS, Job Tracker, Task Tracker, Name Node, Data Node, and internal workings of MapReduce, Spark Batch processing, Spark Streaming, and PySpark frameworks.
- Initiated as a Mainframes Operator, showcasing expertise in Mainframes operations, including JCL, COBOL, CICS, IBM DB2, and CA7.
- Experience in consuming and processing streaming data from Kafka using Spark Streaming API and flatten the data then store the data in Data Lake in parquet format.
- Designed and implemented data processing workflows using Azure Data Factory, orchestrating the movement and transformation of data between various sources and targets.
- Leveraged Azure Databricks for large-scale data processing and analytics, utilizing its collaborative notebook environment for interactive data exploration and model development.
- Developed and optimized complex SQL queries for data analysis and transformation, ensuring efficient data retrieval and processing.
- Designed and implemented data models in Azure SQL Database, ensuring scalability, performance, and data integrity.
- Utilized Azure Blob Storage for scalable and cost-effective storage of unstructured data, implementing data lifecycle management strategies to optimize storage costs.
- Developed and deployed RESTful APIs in Azure using Azure Functions, enabling seamless integration with other applications and services.
- Engaged in the Software Development Life Cycle (SDLC) using Agile scrum methodology for Analysis, Design, Development, Implementation, Maintenance, and Support.

Jr. Data Engineer, ***Ouranos Technologies Pvt Ltd, India*** ***Dec 2018 -May 2019***

- Engineered and managed the creation of 10 automated, scalable, code-based data pipelines, leveraging Amazon Redshift, Data Grip, Amazon S3, and Glue. These pipelines efficiently processed millions of data points, demonstrating a high level of expertise in cloud-based data architecture and automation.
- Successfully executed query optimizations on AWS, utilizing Amazon RDS and Amazon Redshift, to improve database and data warehouse performance by 32%.
- Utilized AWS Athena to manipulate CSV data files stored in AWS S3, applying Scala queries for proficient data extraction and transformation, showcasing versatility in data handling methodologies.
- Formulated and implemented Python solutions to extract data from AWS S3 and populate it into SQL Server, directly supporting business team requirements through effective data integration strategies.
- Played a key role in contributing to a Databricks Delta Lake environment on AWS, employing Spark for sophisticated data processing tasks, enhancing data lake utility and performance.
- Conducted ETL operations using Python, Spark SQL, S3, and Redshift, handling substantial volumes of data to derive actionable customer insights, thus driving business intelligence initiatives.
- Orchestrated automated CI/CD pipelines utilizing AWS CodePipeline, Jenkins, and AWS CodeDeploy, significantly enhancing deployment efficiency and operational agility.
- Authored PySpark scripts to streamline ETL processes, extracting data from S3 with a crawler and generating a data catalog to consolidate metadata.

Data Analyst ***June 2018 - Nov2018***

- Utilized NumPy and Pandas for data preprocessing and analysis, improving data processing time and accuracy in data modeling, supporting decision-making processes.
- Executed SQL techniques to streamline data queries, enhancing overall data analysis efficiency and generating actionable insights for informed decision-making.
- Created interactive data visualizations using Tableau, effectively supporting strategic decision-making processes within the organization.
- Adopted Agile methodologies, particularly SCRUM, in analytical projects, fostering rapid iterations and continuous feedback loops, enhancing project delivery speed and adaptability to evolving requirements in data analysis processes.
- Used Python scripting to automate data collection, transformation, and analysis tasks, enabling efficient handling of large datasets and complex calculations.

SKILLS AND CERTIFICATION

- **Programming Languages:** SQL, Pig Latin, HiveQL, Scala, Python, Unix Shell Scripting.
- **Databases:** MS-SQL SERVER, Oracle, MS-Access, MySQL, Teradata, PostgreSQL, DB2.
- **Big Data Technologies:** Yarn, MapReduce, Pig, Hive, HBase, Cassandra, Oozie, Apache Spark, Scala, Impala, Kafka.
- **Hadoop Distributions:** Apache Hadoop 2.x/1.x, Cloudera CDP, Hortonworks HDP, Amazon EMR (EMR, EC2, EBS, RDS, S3, Glue, Elasticsearch, Lambda, Kinesis, SQS, DynamoDB, Redshift, ECS) Azure HDInsight (Databricks, Data Lake, Blob Storage, Data Factory, SQL DB, SQL DWH, Cosmos DB, Azure DevOps, Active Directory).
- **NoSQL Database:** Cassandra, MongoDB.
- **Reporting Tools/ETL Tools:** Informatica, Talend, SSIS, SSRS, SSAS, ER Studio, Tableau, Power BI.
- **Methodologies:** Agile/Scrum, Waterfall.
- **Development Tools:** Eclipse, NetBeans, IntelliJ, Hue, Microsoft Office Suite (Word, Excel, PowerPoint, Access)
- **Operating Systems:** Windows, Macintosh, Linux, Ubuntu, Unix.
- **Others:** Machine learning, NLP, Stream Sets, Spring Boot, Jupyter Notebook, Docker, Kubernetes, Jenkins, Jira.

CERTIFICATION

- Microsoft Certified: AZURE Fundamentals (AZ-900)
- Microsoft Technology Associate (MTA)
- Recognized as a Microsoft Technology Associate for Programming using Python.
- Microsoft Technology Associate (MTA)
- Recognized as a Microsoft Technology Associate for Web Development using HTML, CSS and JavaScript

EDUCATION

Master of Science in Computer Science ***Aug 2022 - Dec 2023***
Northern Arizona University, Flagstaff, Arizona

Bachelor of Technology in Computer Science and Engineerin ***Jun 2015 - May 2019***
Jawaharlal Nehru Technological University, Kakinada, India