|  |  |  |
| --- | --- | --- |
| **Revanth Reddy Kanumuri**  **Sr Data Engineer** |  |  |
| [revanthreddy2709@gmail.com](mailto:revanthreddy2709@gmail.com) |  |
| 216-394-5011 |  |

**Professional Summary**

* Over 7+ years of comprehensive IT experience as a AWS Solutions Architect Associate, specializing in ETL processes, database management, and leveraging Hadoop/Big Data technologies to manage diverse data types including structured, semi-structured, and unstructured data.
* Proficient across all phases of the Software Development Life Cycle (SDLC), ensuring efficient project execution from requirements analysis to integration.
* Skilled in utilizing a broad range of AWS services, including EC2, S3, RDS, Elastic Load Balancing, SQS, IAM, and more, for effective data processing and storage solutions.
* Experienced in advanced data modeling techniques such as Dimensional, Relational, Star Schema, and Snowflake Modeling, aimed at optimizing database structures through FACT & Dimension tables.
* Proficient with Hadoop ecosystem components like HDFS, MapReduce, and Hive, enhancing data ingestion, analytics, and visualization.
* Demonstrated expertise in developing robust ETL processes using SQL, Python, and tools like Informatica to ensure data integrity.
* Highly skilled in both SQL and NoSQL database management, utilizing Oracle, Teradata, and other databases; adept at creating complex queries and developing insightful reports and dashboards with tools like Tableau and Power BI.
* Experienced in designing and developing comprehensive data pipelines for data ingestion and transformation, utilizing modern CI/CD practices with Docker and Jenkins for streamlined deployment and testing.
* Utilized Java and Apache Spark to perform Extract, Transform, and Load (ETL) operations on diverse datasets, ensuring real-time data integration across heterogeneous sources.
* Effective in source control management with tools like Git and Bitbucket, facilitating team collaboration and maintaining version control.
* Experienced in Agile/Scrum and Waterfall methodologies, adapting project approaches to meet diverse requirements.
* Plays a supportive role in all developmental, testing, and operational phases of system deployments, ensuring seamless transitions and robust system functionality.
* Knowledgeable in deploying and managing containerized applications using Kubernetes, and proficient in high-performance ETL processes using Spark, enhancing system efficiency through in-memory processing.

**Technical Skills**

| **Category** | **Technologies/Tools** |
| --- | --- |
| **AWS Services** | EC2, S3, Glacier, Redshift, RDS, EMR, Lambda, Glue, CloudWatch, Recognition, Kinesis, CloudFront, Route53, DynamoDB, Code Pipeline, EKS, Athena, QuickSight |
| **Data Warehouse Technologies** | AWS RedShift, Google Cloud Storage, Snowflake, Teradata |
| **ETL Tools** | AWS Glue, Airflow, Spark, Sqoop, Flume, Apache Kafka, Spark Streaming |
| **Hadoop Distribution** | Cloudera, Hortonworks, MapR, AWS EMR |
| **Hadoop Tools** | HDFS, HBase, Hive, YARN, MapReduce, Pig, Hive, Apache Storm, Sqoop, Oozie, Zookeeper, Spark |
| **Methodologies** | Agile/Scrum, Waterfall |
| **Monitoring Tools** | Splunk, Chef, Nagios, ELK |
| **NoSQL Databases** | MongoDB, Cassandra, Amazon DynamoDB, HBase |
| **Programming and Scripting** | Spark Scala, Python, Java, MySQL, PostgreSQL, Shell Scripting, Pig, HiveQL |
| **Build & Development Tools** | Jenkins, Maven, CI/CD |
| **Containerization** | Docker, Kubernetes, OpenShift |
| **Source Code Management/Version Control** | JFrog Artifactory, Nexus, GitHub, Code Commit |
| **SQL Databases** | Oracle DB, Microsoft SQL Server, IBM DB2, PostgreSQL, Teradata, Amazon RDS |
| **Data Visualization Tools** | Power BI, Tableau, Hadoop |
| **Web Technologies** | HTML5, CSS3, JavaScript, ReactJS, JSON, Bootstrap, jQuery, AngularJS, NodeJS, Express JS |
| **J2EE Technologies & Backend Frameworks** | Servlets, JSP, JDBC, SOAP, REST, XML, Spring, Hibernate, Spring MVC, Spring Boot, Spring Batch, Spring Cloud |

**Professional Experience**

**O’Reilly Auto Parts, Springfield, MO** Jan’2025 - Present

**AWS Data Engineer**

* Proficient with AWS platforms including EC2, S3, EMR, Redshift, Lambda, and Glue, optimizing cloud-based data processing and storage solutions.
* Developed and optimized Spark applications using Python to handle diverse data sources, including RDBMS and real-time streams. Skilled in using Spark RDD, DataFrame, Dataset, and Data Source APIs, as well as Spark SQL and Spark Streaming for efficient data processing.
* Enhanced algorithm performance within Hadoop ecosystems using Spark tools such as SparkContext, Spark-SQL, and Spark MLib.
* Implemented real-time data processing jobs using Spark Streaming integrated with Kafka to manage live data feeds effectively.
* Successfully migrated on-premises applications to AWS, utilizing EC2 and S3 for seamless data handling and storage. Managed Hadoop clusters on AWS EMR to ensure robust data availability.
* Leveraged AWS Glue and PySpark for data ingestion into S3 buckets, used Elasticsearch for data filtering, and integrated with Hive for external table management.
* Contributed to data warehousing efforts by configuring Snow Pipe for data transfers into Snowflake and optimized data storage in Snowflake's staging area. Also managed Amazon Redshift to consolidate data warehouses.
* Designed and maintained Cassandra database systems, understanding and applying complex architectural concepts like replication strategies and snitches. Developed and managed columnar family data models for efficient query performance.
* Utilized Java and Spark APIs to orchestrate ETL workflows on AWS Glue, facilitating seamless data ingestion from various sources into Amazon S3 buckets. Integrated with AWS Lambda for serverless data processing and transformations.
* Managed and optimized Java-based Spark jobs on AWS EMR, fine-tuning configurations and resource allocation to enhance efficiency while minimizing costs.
* Configured Kafka brokers and managers from scratch, ensuring robust data streaming services. Strengthened Kafka security settings and enhanced system performance to support high-volume data ingestion.
* Created custom UDFs in Python and Hive for specialized data transformation and querying tasks. Employed advanced data formats such as Avro, Parquet, and JSON for optimized data storage and retrieval.
* Developed and managed Oozie coordinators to automate Hive script scheduling, creating efficient data pipelines. Executed MapReduce jobs using PySpark and Numpy, integrating with Jenkins for continuous integration.
* Continuously monitored and managed Hadoop clusters using Cloudera Manager, ensuring optimal performance and reliability.

**Visa, Foster City, California**  **Nov’ 2022 – Dec’2024**

**AWS Data Engineer**

* Spearheaded projects utilizing Spark, SQL, and the AWS cloud environment, leading to enhanced data operations and system integrations.
* Designed, implemented, and maintained AWS infrastructure and services within a managed service setup, ensuring robust, scalable, and secure applications.
* Developed and executed data governance strategies to effectively structure and manage previously ungoverned data environments.
* Engaged in data ingestion, transformation, manipulation, and computation using tools like Kinesis, SQL, AWS Glue, and Spark to optimize data workflows.
* Worked closely with system analysts in requirement gathering sessions to understand upstream source data formats and patterns, ensuring accurate data solutions.
* Executed seamless data migrations from RDBMS to NoSQL databases and designed end-to-end data solutions encompassing storage, integration, processing, and visualization within AWS.
* Developed PySpark pipelines and established an Enterprise Data Lake using EMR and AWS S3, supporting diverse analytics and reporting needs.
* Collaborated with clients and architects to ensure data quality, engaging in data cleansing, transformation, and maintaining relational integrity.
* Implemented a Security Framework using AWS Lambda for secure data access and automated data pipeline operations with AWS Step Functions and Python.
* Enhanced algorithm performance in Hadoop ecosystems by optimizing Spark jobs with Java-based SparkContext, Spark SQL, and MLib libraries, achieving faster query execution and data processing.
* Utilized Power BI and developed Kibana Dashboards to create insightful visuals and dashboards for real-time data analysis and decision-making support.
* Integrated Apache Airflow with AWS to manage multi-stage machine learning processes, enhancing the capabilities of Amazon SageMaker.
* Developed custom user-defined functions (UDFs) with Scala and PySpark, provided comprehensive training sessions on Big Data concepts, and crafted test scripts to support continuous integration and development.
* Collaborated with enterprise data support teams for Hadoop system updates and patches, and actively participated in an Agile environment using tools like Rally for efficient project management.

## **Tata Consultancy Services, Hyderabad, India.** **Jun’ 2017 – July’2022**

**Jr. Data Engineer**

* Engaged in projects that integrate machine learning, big data, and data visualization using R and Python, along with foundational SQL and Unix skills.
* Performed exploratory data analysis utilizing Python libraries such as NumPy, matplotlib, and pandas to uncover trends and insights.
* Assisted in quantitative analysis and data mining, presenting data effectively to highlight key patterns and insights.
* Analyzed data using Python tools including Pandas, NumPy, SciPy, and Matplotlib. Configured AWS IAM for enhanced data security and user authentication.
* Contributed to system design and feasibility studies, recommending cost-effective solutions on the AWS platform.
* Developed and maintained complex SQL queries and scripts to accurately extract and aggregate data according to business requirements.
* Evaluated data quality, documenting limitations and conducting standard SQL queries for data validation. Prepared summary reports and pivot tables in Excel to support business analysis.
* Supported backend development in Python, integrating with Django to manage business logic and create dynamic web page templates.
* Worked with Django REST framework to manage API endpoints and performed unit testing using PyUnit to ensure code reliability.
* Created high-level analysis reports and visualizations using tools such as Excel and Tableau, employing various techniques like Maps, Heat Maps, and Tree Maps for deeper data insights.
* Utilized advanced pandas functions and Python libraries to perform time-series analysis, data loading, and visualization scripting for interactive web applications.
* Used Bitbucket for version control and collaborated effectively within a team environment to meet project deadlines and requirements.

**Education**

* Master of Science in Computer science, Concordia St. Paul, Minnesota.
* Bachelor of Technology in Mechanical Engineering in JNTUK, India.