**Manish Reddy**

**Professional Summary:**

* Around 10 years of IT experience in Architecture, Analysis, design, development, implementation, maintenance and support with experience in developing strategic methods for deploying big data technologies to efficiently solve **Big Data processing** requirement.
* A very good experience on Cloud Engineer using HADOOP framework **BIG DATA** using **HADOOP** framework and related technologies such as **HDFS**, HIVE, **HBASE**, **Map Reduce**, **HIVE**, **SPARK, SCALA, KAFKA, PIG**, **FLUME**, **OOZIE**, **SQOOP**, and **ZOOKEEPER**.
* Good Experience in writing custom **UDFs** in java for **Hive** and **Pig to extend the functionality**.
* Designed and implemented scalable ETL pipelines using Databricks to ingest and process large datasets from various sources.
* Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems (RDBMS) and from RDBMS to HDFS.
* Develop Databricks Python notebooks to Join, filter, pre-aggregate, and process the files stored in Azure data lake storage.
* Implemented efficient algorithms using Scala collections, taking advantage of immutable and mutable data structures to balance performance and safety.
* Utilized deep learning techniques like Gen AI to enhance applications, improving object detection accuracy by 25% and enabling real-time image recognition in diverse environments.
* Knowledge on Google Cloud Platform (GCP) services like compute engine, cloud load balancing, cloud storage, cloud Data proc, Cloud Pub/Sub, cloud SQL, **Big Query**, stack driver monitoring, Cloud spanner, looker, Terraform GCP Foundation modules and deployment manager.
* Extensive experience in using Microsoft BI studio products like **SSIS, SSAS, SSRS** for implementation of ETL methodology in data extraction, transformation and loading.
* Experience in Microsoft Azure cloud platform and merge with Python to store the data into cloud with High security.
* Expertise in design and development of various web and enterprise applications using Type safe technologies like Scala, Akka, Play framework, Slick.
* Extensive Development experience with Mark Logic Data Ingestion, transformation to XML/JSON, and Data Curation.
* Utilized Scala collections Lists, Maps, Sets, to efficiently manage and manipulate data within applications, ensuring optimal performance.
* Designed reusable and modular code using Scala traits, promoting code reuse and enhancing maintainability.
* Ensured compliance with data security standards and regulations by implementing robust access controls and encryption mechanisms in Netezza environments.
* Developed and optimized data models and schemas in Cassandra to support high-volume, low-latency data access patterns.
* Ingested the data to **Mark Logic**, transformed to expected XML or JSON formats using XSLT and ML Content Pump.
* Participates in the development improvement and maintenance of snowflake database applications
* Consulting on Snowflake Data Platform Solution Architecture, Design, Development and deployment focused to bring the data driven culture across the enterprises
* Expertise in creating mappings in TALEND using tMap, tJoin, tReplicate, tParallelize, tConvertType,, tflowtoIterate, tAggregate, tSortRow, tFlowMeter, tLogCatcher, tRowGenerator, tNormalize, tDenormalize, tSetGlobalVar, tHashInput, tHashOutput, tJava, tJavarow, tAggregateRow, tWarn, tLogCatcher, tMysqlScd, tFilter, tGlobalmap, tDie etc.
* Deployed AES encryption functions within ETL data flows using coding/scripting for column-level transformations.
* Experience in AWS platform and its features **including IAM, EC2, EBS, VPC, RDS, Cloud Watch, Cloud Trai**l, Cloud Formation AWS Configuration, **Autoscaling**, **EMR**, CloudFront, **S3**, SQS, RedShift, Terraform SNS, Lambda and Route53.
* Automated **Docker** Image building and deployment using Continuous Delivery processes, streamlined deployment by 40%.
* Specialized in provisioning the Global HTTPS **load balancer routes the traffic** to GCP GKE cluster via Terraform modules.
* Implemented data quality checks and validation processes within Databricks pipelines.
* Designed and managed Snowflake data warehouses, including schema design, table creation, and performance tuning.
* Involved the MarkLogic admin activities like Backup the whole Database and restore it, re-index periodically and maintain the configuration setup in all the environment Developed and Bug fix after post launch for search module using Field search.
* Provide Business Intelligence support using **Tableau and Microsoft** **Power BI** for implementing effective Business dashboards & visualizations of data.
* Used **Kubernetes** to orchestrate the deployment, scaling and management of Docker containers.
* Worked extensively with Dimensional modeling, Data migration, Data cleansing, Data profiling, and ETL Processes features for data warehouses.
* Use Kafka a publish-subscribe messaging system by creating topics using consumers and producers to ingest data into the application for Spark to process data and create Kafka topics for application logs.
* Proficient with **Azure Data Lake** (ADLS), Databricks &amp; python **Notebook** format, Databricks
* Data visualization expertise for Enterprise Reporting Applications with **Cognos**, **Looker**, **Tableau**
* Hands On experience **Graph Database language Relational AI technology. (GQL).**
* Good experience working with ON-Prem Clusters like Horton works and **Cloudera** Distribution.
* Experience in writing **TDD** development test cases, Junit and **MRUnit** Test cases.
* Experience working with **JAVA**, J2EE, JDBC, in **Object Oriented Analysis Design** (OOAD) ODBC, JSP, Java Eclipse, Java Beans, EJB, **Servlets, MS SQL Server**.
* Experience in **GCP** **Data** **proc**, **GCS**, Cloud functions, **Cloud SQL & Big Query.**
* Implemented robust backup and disaster recovery strategies for **Cassandra databases**, ensuring data integrity and minimal downtime
* Experience in all stages of SDLC (Agile, Waterfall), writing Technical Design document, Development, Testing and Implementation of Enterprise level Data mart and **Data warehouses**.
* Engaged with stakeholders to identify and implement best practices in **data management**, and effectively resolved complex technical issues to enhance system performance and reliability.
* Communicated technical issues, solutions, and project statuses clearly to both **technical** and **non-technical stakeholders,** ensuring transparency and understanding through detailed documentation and regular updates

**Technical Skills**:

|  |  |
| --- | --- |
| **Programming languages:** | Java, Python, R, Spark, Rust, Hadoop, OOP’s, Kafka, Scala, Alteryx, Kubernetes, Airflow, Data Bricks |
| **Database:** | MYSQL, MongoDB, Oracle, DynamoDB, Postgres, Snowflake, PostgreSQL, Redis. |
| **Cloud Services:** | GCP, Azure, AWS cloud, S3, Step Functions, Neo4j, REL AI, Redshift, Glue. |
| **Web Technologies:** | HTML, CSS, JavaScript, TypeScript, Node.js. React, Groovy, Django. |
| **Bigdata Services:** | HDFS, Spark, Scala, Spark, Map Reduce, HDFS, Oozie, Hive, Pig, Sqoop, Flume, Zookeeper, Kafka and HBase, Databricks, data lake, Mark logic |
| **Methodologies** | Agile, Waterfall. |
| **ETL Tools** | Talend, NIFI, Hyper flow’s, AWS Glue, Power BI, Talend, Tableau, SSIS/SSRS |
| **Scripting Technologies** | PowerShell, NodeJS, JavaScript, HTML, Jinga, CSS, Shell Scripting, XQuery, XSLT |
| **Networking protocols** | DNS, VPC, TCP/IP, HTTP. |
| **Version Tools:** | GitHub, SVN, Visual Studio, MS Excel. |

**Work experience**

|  |  |  |  |
| --- | --- | --- | --- |
| **Role** | **Sr. Data Engineer** | **Start Date** | Jan 2020 |
| **Client** | **Optum**, **Eden Prairie MN** | **End Date** | Till Date |

**Responsibilities:**

* Managed Multiple Aws Accounts with multiple VPC’s for both Prod and Non-Prod environments.
* Experience in defining, designing, and developing Python applications, specially using Hadoop by leveraging frameworks such as Hdfs, Python, **PySpark** and **AWS** Cloud.
* Developed and maintained scalable AI infrastructure using cloud services AWS and containerization technologies like Docker and Kubernetes for deployment.
* Conducted data analysis and generated insights using Databricks SQL and Spark.
* Wrote SQL queries and optimized the queries in **Redshift**, **Snowflake**.
* Design/Implement large scale pub-sub message queues using Apache Kafka.
* Developed a **Sqoop script** and developed scripts to extract structured data from **AWS Redshift**, **MySQL** and other **RDS databases** data into Hive tables directly on a daily incremental basis.
* Design and Develop ETL Processes in **AWS Glue** to migrate Campaign data from external sources

like S3, ORC/Parquet/Text Files into AWS Redshift.

Extensive Experienced in design **EC2** instances in all the environments to meet high availability and complete security. Setting up the **Cloud Watch** alerts for **EC2** instances and using in Auto scaling

* Experience with **Snowflake** cloud data warehouse and AWS S3 bucket for integrating data from multiple source system which include loading nested **JSON** formatted data into **snowflake** **table**.
* Developed **Snowflake** **views** to load and unload data from and to an AWS S3 bucket, as well as transferring the code to production
* Developed Spark Programs using Scala and Java API's and performed transformations and actions on RDD's.
* Developed REST APIs using Scala, Play framework and Akka.
* Used Data frame API in Scala for converting the distributed collection of data organized into named columns, developing predictive analytic using Apache Spark Scala APIs
* Developed REST APIs using Scala and Play framework to retrieve processed data from Cassandra database.
* Collaborated with cross-functional teams using **Databricks** **Notebooks** for interactive data exploration, analysis, and sharing of insights.
* Configured Chef to build up services and applications on the instances once they have been configured using **CloudFormation**.
* Used Scala collection framework to store and process the complex consumer information.
* Developed real time data streaming solutions using Spark Structured Streaming with Scala applications to consume the JSON messages from Kafka topics
* Designed and Implemented the ETL process using **Talend** Enterprise Big Data Edition to load the data from Source to Target Database.
* Automated data workflows and scheduling using Databricks notebooks and Databricks Workflows.
* Implemented a 'server less' architecture using **API Gateway**, **Lambda**, and **DynamoDB** and deployed AWS Lambda code from **Amazon S3 buckets, Redshift, EMR**
* Responsible for maintaining our on-premises environment Database Management Systems on MySQL, **PostgreSQL** server and **Teradata**.
* Designed and implemented **MDM** solutions using Talend MDM to ensure data consistency and accuracy across various business units.
* Used Scala functional programming concepts to develop business logic.
* Implemented applications with Scala along with Akka and Play framework. Expert in implementing advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala.
* Developed a Restful API using & Scala for tracking open-source projects in GitHub and computing the in-process metrics information for those projects
* Worked on to Fetch secrets like Postgres user and passwords from Vault and to pass the values to Project code through **lockbox** migration.
* Experience on log parsing, complex **Splunk searches**, including external table lookups, **Splunk data** flow, components, features and product capability.
* Loaded historical transactional data into **Neo4j** graph database by writing cypher queries.
* Designed and implemented complex **graph database models using Neo4j**, optimizing data structures for efficient traversal and relationship representation.
* Worked closely with data scientists to build and deploy machine learning models using **Databricks** **MLflow**.
* Participated in the design and deployment of data warehousing solutions on AWS, ensuring seamless integration with MDM systems.
* Implemented robust security measures, including role-based access control, data masking, and **encryption** to protect sensitive data.
* Worked on installing new relic **Kafka** plugin implementation for monitoring of Kafka cluster
* Developed real time data streaming solutions using Spark Structured Streaming with **Scala** applications to consume the JSON messages from **Kafka** topics.
* Developing Spark applications using Spark - SQL in Databricks for data extraction, Participating inmigration Scala code into Microservices.
* Created Notebooks in Azure Data Bricks and integrated it with AWS to automate the same
* Experience in defining, designing, and developing Java applications, specially by leveraging frameworks such as Spark, **Kafka**.
* Design **MarkLogic** Data Ingestion workflows for Full Load Data Ingestion, and Incremental Data ingestion on daily basis
* Implemented Snowflake's time travel and cloning features to maintain historical data and streamline development processes.
* Developed a **Talend** Code for S3 Tagging in the Process of Moving data from source to S3.
* Created and optimized ETL processes to ingest, transform, and load large datasets into Netezza, ensuring data accuracy and consistency.
* Conducted troubleshooting and debugging of issues related to Snowflake data warehouse, identifying root causes and implementing solutions to minimize downtime and optimize system performance.
* Improved the performance of the **Kafka** cluster by fine tuning the Kafka Configurations at producer, consumer and broker level.
* Develop MarkLogic REST API Services for rendering the data on Customer Portal from **MarkLogic** **Database.**
* Enabled secure data sharing and collaboration between different teams and external partners using Snowflake’s data sharing capabilities
* Worked with web deployment technology specially Apache/**Tomcat**/Java.
* Utilized AI techniques such as natural language processing (**NLP**) to extract insights from unstructured data sources, enabling more comprehensive data analysis and decision-making.
* Configuring and managing **AWS** Simple Notification Service (**SNS**) and Simple Queue Service (SQS).
* Experience in writing Sub Queries, Stored Procedures, Triggers, Cursors, and Functions on SQL and PostgreSQL database.
* Developed Databricks ETL pipelines using notebooks, PySpark data frames, Spark-SQL, and python scripting
* Experience in handling the applications authentications by **using Oauth2/SSO, SAML.**
* Experienced in Agile Methodologies, Scrum stories and Sprints, PI Meetings.

**Environment**: AWS Cloud, Scala,Data Bricks, Docker, Kubernetes, Mark Logic, Jenkins, Bogie File, Hadoop Framework, Python, MDM, Scala, Hive, Erato code, Snowflake, No Sql, XQuery, XSLT SQL, Java, Grails, UNIX, Taledn,Shell Scripting, Oracle 11g/12g Jenkin.

|  |  |  |  |
| --- | --- | --- | --- |
| **Role** | **Data Engineer /Scala Engineer** | **Start Date** | Oct 2018 |
| **Client** | **SunTrust Bank, Atlanta, GA** | **End Date** | Jan 2020 |

**Responsibilities:**

* Exploring with Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark context, Spark-SQL**, Data Frame**, pair RDD’s, **Spark** **YARN**.
* Recreating existing application logic and functionality in the Azure Data Lake, Data Factory, SQL Database and SQL data warehouse environment.
* Designed and implemented by configuring Topics in the new Kafka cluster in all environments.
* Strong experience working with Spark Data frames, **Spark SQL** and Spark Structured Streaming APIs using Scala. Implemented batch processing of jobs using Spark Scala API’s.
* Proficient in utilizing Snowflake features such as **SnowSQL** and **SNOWPIPE** for seamless and continuous data ingestion for analytical purposes.
* Strong experience working with Spark Data frames, Spark SQL and Spark Structured Streaming APIs using Scala. Implemented batch processing of jobs using Spark Scala API’s.
* Developed real time data streaming solutions using Spark Structured Streaming with Scala applications to consume the JSON messages from Kafka topics.
* Worked on migrating the old java stack to Type safe stack using Scala for backend programming.
* Used Scala collection framework to store and process the complex consumer information. Based on the offer’s setup for each client, the requests were post processed and given offers.
* Developed interactive dashboards and reports for business stakeholders.
* Conducted performance tuning and optimization activities on **Snowflake** to enhance query performance, minimize latency, and improve overall system efficiency.
* Developed real time data streaming solutions using Spark Structured Streaming with Scala applications to consume the JSON messages from Kafka topics.
* Extensively worked with automation tools like **Jenkins**, **Artifactory**, **SonarQube** for continuous

integration and continuous delivery **(CI/CD)** and to implement the End-to-End Automation.

* Involved in creating Hive Tables, loading the data from the **cornerstone** tool.
* Developed and managed data integration workflows to load data from various sources into **Snowflake** using tools like **Snow** **pipe**, **ETL/ELT** frameworks, and third-party connectors
* Conducted Spark ETL pipeline delivery on **Azure Databricks**, orchestrating data transformation, achieved a 25% improvement in data processing
* Developed data pipeline using EventHub’s, PySpark, and Azure SQL database to ingest.

customer events data and financial histories into Azure cluster for analysis.

* Involved in **converting Hive/SQ**L queries into **Spark** **transformations** using **Spark RDDs**, by using **Python**.
* Optimized data processing workflows for performance and cost-efficiency using **Databricks**.
* Ensured compliance with **data governance** policies and industry standards.
* Wrote Automation Script to auto start and stop the services in **Azure** **cloud** for cost saving.
* Automated resulting scripts and workflow using **Apache Airflow** and shell scripting to ensure

daily execution in production.

* Have knowledge on partition of **Kafka** **messages** and setting up the replication factors in **Kafka Cluster.**
* Developing Spark applications using **Spark - SQL** in **Databricks** for data extraction, Participating in migration **Scala code** into **Microservices**.
* Responsible to develop the Code and Unit Test and move the code to **UAT** and **PROD**.
* Used Amex Internal Framework **Event engine** to trigger the jobs and monitor the jobs.
* Optimized complex **SQL** queries and materialized views for improved performance and reduced compute costs.
* Automated data workflows using **Databricks** Jobs and integrated with orchestration tools like Apache Airflow for scheduling and monitoring.
* Used **Snowflake's query** profiling and performance monitoring tools to identify and resolve bottlenecks.
* Worked Extensively on Talend Admin Console and Schedule Jobs in Job Conductor.
* Conducted Spark ETL pipeline delivery on Azure Databricks, orchestrating data transformation via Azure Data Factory, achieving a 25% improvement in data processing
* Data sources are extracted, transformed, and loaded to generate CSV data files with **Python programming** and **SQL queries**.
* Used Rally for project tracking and project status.

**Environment**: Hadoop, Snowflake,Scala, DataBricks,data bricks, HDFS, Talend, Azure, Azure Data bricks, Pig, Sqoop, HBase, Shell Scripting, Maven, Jenkins, Ubuntu, MarkLogic, MDM, Linux Red Hat, Junit, Hive, Java (JDK 1.6), Hadoop Distribution Cloudera, Azure, SQL, UNIX Shell Scripting.

|  |  |  |  |
| --- | --- | --- | --- |
| **Role** | **Big Data/Hadoop Developer** | **Start Date** | Oct 2016 |
| **Client** | Chemical abstract Services, Columbus OH | **End Date** | Oct 2018 |

**Responsibilities**:

* CAS deals with chemical data in order to build their products like SciFinder or STN Next or Patent Pak by using different types of domains with their respective schema.
* Experiences in handling the **Chemical data** and responsible for managing data coming from different sources.
* Experienced with CAS's different domains data which were build using the priority order of **XSLT, XQuery**.
* Leveraged a good knowledge while dealing with **XSLT** and **XQuery** using **Altova Xmlspy**.
* Performing data analysis and data profiling using complex SQL on various sources systems including Oracle and Hadoop Data Validation, Data Cleansing, Data Verification and identifying data mismatch, Data quality Data Migration using ETL tools.
* Writing and **Testing SQL and PL/SQL statements** - Stored Procedures, Functions, Triggers and packages.
* Implemented best practices for data partitioning, caching, and parallel processing.
* Integrated Hive and HBase for effective usage and performed MR Unit testing for the Map Reduce jobs.
* Worked on managing the **Spark** **Databricks** by proper troubleshooting, estimation, and monitoring of the cluster
* Developed **Databricks** **ETL** pipelines using notebooks, PySpark data frames, Spark-SQL, and python scripting.
* Extensive experience in **MarkLogic** in establishing the Clustering, Bulk Data loads, XML Transformation, Faceted Searches, Triggers, Alerting, **Xquery** Code, Forest rebalancing
* Worked on creating data ingestion processes to maintain Data Lake on the GCP cloud and Big Query.
* Involved in transforming data from Mainframe tables to HDFS, and HBase tables using Sqoop and Pentaho.
* Involved in creating a **job scheduling** using **scripting**, data **ingestion** to data processing storage on HDFS.
* Managed and optimized data exchange processes to facilitate real-time analytics and business intelligence
* Develop ETL mappings for various Sources ( .TXT, .CSV, XML ) and load the data from these sources into relational tables with Talend Enterprise EditionBAch
* Used ETL methodologies and best practices to create Talend ETL jobs.
* Participated in supporting Data Governance, and Risk Compliance platform utilizing **MarkLogic**
* **Implemented Partitioning** (both dynamic Partitions and Static Partitions) and **Bucketing** in HIVE.
* Created Talend custom components for the various use cases and worked on XML components, Data quality, Processing and Log & Error components.
* Managing messages on Kafka topics using Talend Studio Jobs.
* Experience in developing at-scale machine learning systems.
* Experience in Automation Testing, **Software Development Life Cycle (SDLC)** using the **Waterfall Model** and good understanding of **Agile** **Methodology.**
* On a day-to-day basis, experiencing Agile Scrum methodology and Jira ticket system for project development.

**Environment**: Hadoop Framework, Talend, Marklogic, Xquerry, Xslt, MapReduce, GCP, Hive, Hive Meta store, Sqoop, Pig, HBase, Flume, MDM, Data Bricks, Ozie, Java (JDK1.6), UNIX Shell Scripting, Oracle 11g/12g,

|  |  |  |  |
| --- | --- | --- | --- |
| Role | Software Developer | Start Date | Apr 14 |
| Client | Toll Plus Inc, Hyderabad, India | End Date | Dec 15 |

Responsibilities

* Designed and developed Web Services using Java/J2EE in WebLogic environment. Developed web pages using Java Servlet, JSP, CSS, Java Script, DHTML, and HTML. Added extensive Struts validation. Wrote Ant scripts to build and deploy the application.
* Involve in the Analysis, Design, and Development and Unit testing of business requirements.
* Developed business logic in JAVA/J2EE technology.
* Implemented business logic and generated WSDL for those web services using SOAP.
* Worked on Developing JSP pages, Implemented Struts Framework.
* Modified Stored Procedures in Oracle Database.
* Developed the application using Spring Web MVC framework.
* Worked on the Spring DAO module and ORM using Hibernate. Used Hibernate Template and Hibernate Dao Support for Spring-Hibernate Communication.
* Configured Association Mappings such as one-one and one-many in Hibernate
* Worked with JavaScript calls as the Search is triggered through JS calls when a Search key is entered in the Search window
* Worked on XML, XSL and XHTML files.
* As part of the team to develop and maintain an advanced search engine, would be able to attain.

**Environment**: Java 1.6, J2EE, Eclipse SDK 3.3.2, Java Spring 3.x, jQuery, Oracle 10i, Hibernate, JPA, Json, Apache Ivy, SQL, stored procedures, Shell Scripting, XML, HTML and JUnit, TFS, Ant, Visual Studio.