Rajdeep Charla email :kumar@tricubic.ai Phone:9364104666



**PROFESSIONAL SUMMARY:**

* Around 10 **years** of experience in IT industry which includes **8 years** of experience in development using **Big Data and Hadoop ecosystem** tools
* Experience in designing & developing applications using Big Data core technologies **Map Reduce, Hive, Spark & Spark SQL**
* Excellent in learning and **adapting to new technologies**
* Experience with configuration and development on multiple Hadoop distribution platforms **Cloudera and Hortonworks (**on Premise**)** including cloud platforms **Amazon AWS**
* Experience on cloud base services like **AWS EMR, EC2, S3, Redshift, Athena** to work in distributed data model
* Experience in configuring and maintaining long-running Amazon EMR clusters manually as well as through Cloud Formation scripts on Amazon AWS
* Experience in implementing **Hive simple, generic custom UDF’s** and developed multiple hive views for accessing **underlying Tables** data
* Experience in doing performance tuning for map reduce, Spark jobs & hive complex queries
* Experience in importing and exporting data using **SQOOP** from **HDFS** to Relational Database Systems and vice-versa
* Have good knowledge in **Oozie, Airflow, Zookeeper, Kafka, Spark structured streaming & PostgreSQL**
* Strong understanding in Object-Oriented Programming concepts and implementation
* Generated a script in AWS Glue to transfer the data and utilized **AWS Glue** to run ETL jobs and run aggregation on PySpark code
* Strong understanding and working experience on injecting and accessing data to and from a Cassandra cluster using different java and spark API’s
* Strong Knowledge of spark with Cassandra and Hadoop
* Experience in working **Agile** based applications in different phases like design, development, testing and deployment
* Ability to work effectively with associates at all levels within the organization
* Drive microservices style of architecture when applicable. Apply Domain Driven Design concepts to model microservices
* Strong background in mathematics and have very good **analytical and problem-solving skills**
* Capable of processing large sets of **structured, semi-structured and unstructured data** and supporting systems application architecture
* Strong knowledge in real time and near real time applications using apache NiFi, amazon kinesis.
* Able to assess business rules, collaborate with stakeholders and perform source-to-target data mapping, design, and review
* Familiarity with a broad mix of technologies including a strong subset of Microservices, PaaS
* Knowledge in migrating to microservices cloud solutions to build “cloud native” applications

**TECHNICAL SKILLS:**

| **Hadoop/Big Data** | Hadoop (Yarn), HDFS, **MapReduce**, **Spark**, **Hive**, **Sqoop**, Flume, Zookeeper, Oozie, Tez |
| --- | --- |
| **Programming Languages** | **SQL**, **Python** |
| **Distributed computing environments** | **Amazon EMR** **(Elastic MapReduce)**, **Horton Works (Ambari),** **Cloudera (CM)** |
| **Data flow Tools** | **Apache NiFi, Kafka**, Amazon Kinesis |
| **Relational Databases** | Oracle**,** MySQL**, SQL Server** 2005,2008 |
| **NoSQL Databases**  | **HBase**, Cassandra |
| **Cloud Environments** | AWS (EC2, EMR, S3, Kinesis, DynamoDB), Azure, Docker container (Kubernetes), GCP (Google Cloud Platform) |
| **Data File Types** | **JSON**, **CSV**, **PARQUET**, **AVRO**, **TEXTFILE, ORC** |
| **Version Control Tools** | **Git, Subversion** |
| **Methodologies** | **Agile/ Scrum**, Rational Unified Process and Waterfall |
| **Operating Systems** | MacOS, Windows, **Unix,** and Solaris |

**Equality Health, Arizona**

Role: Data Engineer Mar 2024 - Present

**Responsibilities:**

* Run ETL Pipelines in AWS services S3, DynamoDB, Athena, Redshift, Glue.
* Automate and schedule AWS Glue jobs using Apache Airflow.
* Developed and implemented an ETL pipeline to fetch files from SFTP Server to AWS Environment.
* Implemented AWS Glue Job to process the data and create data extracts.
* Created a system to place the Extracts in SFTP for stakeholders.
* Extract, Transform and Load data using file formats like EXCEL, CSV and TXT.
* Manage code deployment processes through Azure Devops and establish robust CI/CD pipelines.
* Manage end-to-end data responsibilities including data modeling, ad-hoc analysis, and implementation of algorithms for extracting Quality and Membership metrics in the PowerBI.
* Integrate PostgreSQL with Airflow for managing workflows and triggering jobs dynamically.
* Perform data analysis and validate data of US-based Health Care/Insurance companies to deliver actionable insights and improve operational efficiency.
* Create, develop, and maintain SQL processes for optimal performance data warehousing.
* Monitor System health and logs and respond accordingly to any warning or failure conditions
* Write Spark applications using Python to interact with the PostgreSQL database using Spark JDBC

**Nike, Inc., Oregon**

Role: Data Engineer Jan 2021 - Feb 2024

**Responsibilities:**

* Worked on Airflow Version Upgrade from 1.1.10 to 2.2.5
* Successfully developed Nike Global Sales Pipelines and onboarded stakeholders to consume the data to support multiple use cases across Nike.
* Extracted product data feeds from Kafka to Spark processing system and stores the order details in Hive databases.
* Architected robust and scalable data integration pipelines utilizing Pyspark, enhanced data processing by using EMR and EC2 instances and orchestrated the pipelines using Apache Airflow.
* Executed end-to-end ETL pipelines encompassing data modeling, ad-hoc analysis, and the formulation and implementation of algorithms to extract Sales and Revenue metrics within the AWS environment.
* Designed and developed Tableau dashboards to generate sales and revenue reports, providing actionable insights for strategic decision making. Orchestrated this process using Airflow.
* Elevated data integrity and quality assurance through advanced test design and documentation, enabled self-service customer support.
* Created Delta tables in Databricks to enhance data reliability, streamline data processing workflows and utilized Delta Lake features for ACID transactions and time travel capabilities.
* Developed materialized views in Databricks for Stakeholders addressing specific analytical needs and improved query performance.
* Resolved complex data integration challenges through optimized ETL strategies and influenced data-driven opportunities across cross-functional teams.
* Created an e-mail notification service upon completion of job using Pandas for the stakeholders who requested the data in CSV format.
* Managed code deployment processes through GitHub and Jenkins, establishing robust CICD pipelines

**Carelon Global Solutions**

Role: Sr. Data Engineer May 2019-Dec 2019

**Responsibilities:**

* Experience in AWS services like **EC2, EMR, S3, DynamoDB, Athena, Redshift, Glue**
* In-depth understanding of **MapReduce** Framework and **Spark** execution model
* Capable of processing large sets of **structured, semi-structured and unstructured** data implements and leverages open source technologies
* Migrating servers, databases, and applications from on premise to **AWS**
* Mastered in using different columnar file formats **(ORC, Parquet**)
* Worked on implementing cost and time efficient solutions in AWS for the end clients.
* Monitored the MR, Tez, Sqoop, Spark jobs using UI interface (Namenode Manager, Resource Manager ETC) in AWS
* implemented Partitions, bucketing concepts in Hive for query optimization and designed both Managed and External tables in Hive to optimize performance
* Exported final output data to NoSQL database Cassandra using spark connectors.
* Developed **ETL parsing and analytics data pipelines using Python/Spark** to build a structured data model
* Implemented Spark **using Python and Spark SQL API** for faster testing and processing of data
* Developed Spark scripts using **Python on AWS EMR** for Data Aggregation, Validation and adhoc querying.
* Extracted the needed data from the server into **S3** and Bulk Loaded the cleaned data into Cassandra using Spark
* Rigorously used **PySpark (RDD’s, Data frames, Spark Sql**) and Spark - Cassandra -Connector API's for various tasks (Data migration, Business report generation etc.)
* Moved Relational Database data to **S3** using Sqoop into s3 data lake and created external hive tables on top of the raw data.
* Created an e-mail notification service upon completion of job for the team which requested for the data
* Played a key role in productionizing the application after testing by BI analysts
* Worked in Agile methodology actively participating in Scrum Planning, Review call, Retrospective and daily scrum in addressing problem/issues.
* Scheduled jobs using airflow orchestrating tool.

**Capgemini Technology Services**

Role: Hadoop Developer Aug 2017 – Apr 2019

**Responsibilities:**

* Played a major role in gathering requirements, analysis of entire system and providing estimation on development, testing efforts
* Worked in the process of configuring, installing **Cloudera Hadoop Environment**.
* Written Spark applications using python to interact with the PostgreSQL database using Spark jdbc and accessed Hive tables using Hive Context
* Created multiple **Hive tables, implemented Dynamic Partitioning** and Buckets in Hive for efficient data access
* Involved in creating **Hive External tables**, also used **custom Serdes’s** based on the structure of input file so that schema on read is implemented in Hive tables
* Involved in converting **Hive/SQL** queries into Spark transformations (map, filter and many more) and action commands (reduceByKey, groupByKey with Spark **RDDs** using **Python**
* Implemented Spark structured Streaming and Spark **SQL using Data Frames**
* Integrated product data feeds from **Kafka to Spark** processing system and store the order details in PostgreSQL database
* Worked on **scheduling** all jobs using **Oozie.**
* Migrated the existing data to Hadoop Environment (**HDFS clusters**) from RDBMS using Sqoop for processing the data and developed ETL pipeline to efficiently process data using python.
* Implemented common use cases like data integrity checks, data sourcing using Spark.
* **Used Jira for ticketing and tracking** issues and Jenkins for continuous integration and continuous deployment
* **Monitor System health and logs** and respond accordingly to any warning or failure conditions

**Cognizant Technology Solutions**

Role: SQL/ETL Developer Aug 2014 – July 2017

* Understood the requirements by interacting with business users and mapping them to design and implement in the **AGILE Development methodology**.
* Worked as an SQL Developer and establish strong relationships with **end users and technical resources** throughout established projects.
* Work closely with the development team to provide support for database objects (**stored procedures/triggers/views**) and performing code review to ensure proper design and implementation of **database systems**.
* Experience with **database administration** including performance monitoring, access management and storage management.
* **Create, develop, and maintain SQL** processes for channel performance data warehousing.
* Participated in Designing databases **(schemas)** to ensure that the relationship between data is guided by tightly bound key constraints.
* Extensive experience in **Data Definition, Data Manipulation, Data Query, Data Transaction and Control Language**.
* Loaded Data into **Oracle Tables using SQL Loader**.
* Debug the procedure, function, package using DBMS\_UTILITY.
* Experience in designing and **creating Tables, Views,** **Indexes, Stored Procedures**, Cursors, Triggers and Transactions.
* Configured the data mapping between sources and destination in **Oracle environment** and **tested performance accuracy** related queries under SQL Server.
* Closely working with development manager and onsite team to deliver solutions.