Kartik Rothe

Kartikrothe@gmail.com

9665157681 | Infomania Technology Solutions Pvt Ltd |Bangalore|

Objective

Dedicated software professional with history of meeting company goals utilizing consistent and organized practices. Skilled in working under pressure and adapting to new situations and challenges to best enhance the organizational brand.

Experience

- **Infomania Technology Solutions Pvt Ltd**   June2020 - Till Date
  Data Engineer
  - 2.6 years of Hands-on Experience in Big Data Ecosystem components.
  - Have Good knowledge about Spark ,Pyspark,Hive,Sql,AWS.
  - Uploaded and processed terabytes of data from various structured and unstructured sources into HDFS.
  - Collecting large amount of structured data using Apache Sqoop and aggregating using HIVE.
  - Designed Internal and External tables in hive.
  - Good understanding of Partitioning, Bucketing concepts in Hive for performance.
  - optimization and Experience in rebalancing the data in the cluster.
  - Manage and review Hadoop log files.
  - Strong problem solving & technical skills coupled with clear decision-making.
  - Hands-on Experience in Hadoop eco-system. Good exposures in Requirement.
  - analysis, Customization and user acceptance.
  - Experienced in working with Amazon Web Services (AWS) using EC2 for computing and S3 as storage mechanism.
- **Bharuka Construction Pvt Ltd.**   June 2019 - May 2020.
  HR Executive
  - Prepare or maintain employment records related to events such as hiring,termination,
  - leaves, transfers, or promotions, using human resources management system software.
  - Statutory Compliances knowledge Employee Database Management, Issuance of
  - Various Letters, Leaves & Attendance Management, Payroll, EPF-ESIC & Legal Compliance, Performance Review & Management, Trainings & Development activities.
  - Follow up for timely data updation for all employee records, Appointment Letters
  - Confirmations letters, Transfers, Probation, and Exits etc

Education

- **MGM Institute of Management**
  MBA (Post Graduation in Human Resource)   2019
- **MGM Dr.GY Pathrikar CS and IT, Aurangabad**
  Bachelor's of Computer Science   2017
- **Maharashtra State Board**
  HSC   2014
- **Maharashtra State Board**
  SSC   2012

Skills

- Platforms: Windows, Linux.
- Programming Language: SQL, Python
- Scripting Language: Shell Scripting
- Big Data Ecosystem: HDFS, Sqoop, Hive
- Management Tool: Zookeeper.
- IDE: Eclipse, PyCharm.
- Cloud Technologies: AWS (S3, EC2, EMR)
- Relational database: MySQL, Oracle.

Projects

- **Healthcare System Data Pipeline**
  Duration: November 2021 to Till Date
  Role : Data Engineer
  Team Size : 11

  This project we are fetching the data from various Internal and External sources and processing the data in pyspark like data cleaning, data integration, data transformation, data Modeling After processing the data in pyspark, the output result we are storing on AWS S3 as a Business View for further generating the reports as per client's requirements.

  Roles &Responsibilities :
  - Data cleaning, Data Integration, Data Transformation for Data compatibility.
  - Monitor Spark jobs.
  - Importing the data into Pyspark from various sources.
  - Processing the data in Pyspark as per the client requirements.
  - Perform optimization techniques in order to enhance performance of Spark jobs.
  - Storing the final processed data in to AWS S3.
  - Monitor Data Validation.
  - Data Extract from Edge Node.
  - Process RAW data .
  - Process for Data Cleaning for Null Values, Corrupt Record, Schema Record,
  - Duplicate record .
  - Written Hive Queries for Data Analysis to meet the Client requirements.
  - Manage Internal and External table for Processing data .
  - Experience with data pipeline and workflow management tools.
  - Experience with object function scripting languages: Python, Scala
  - Extensive knowledge of data engineering tools, technologies, and approaches Hands-on experience with ETL tools.
  - Experience in Warehouse: SQL/NoSQL, Amazon, Oracle,, Apache, Hive etc.
  - Knowledge of multiple data technologies and concepts. .

- **Data Injection Framework**
  Duration: June 2020 to October 2021
  Role : Data Engineer
  Team Size : 11

In this framework, we fetching the data from different sources and ingest the data into data lake.For this we are directly connecting to their source system & we ingest the data into our Hive table.For some of the sources, we are using File formats for ingestion. For testing, we are taking some sample data & we ingest that data in to our Hive table with Data Migration application. Then we are loading historical data batchwise into the cluster.

Roles and Responsibilities:
- Import Data into the Hive from various Relational Databases (Oracle, MYSQL Server) into HDFS using Write Hive DDL and DML queries in Hive.
- Ingest Flat Files like Delimited, Fixed Length, etc. into Hive Warehouse.
- Engage with BA to understand the requirements clearly.
- Perform optimization techniques in order to enhance performance of Sqoop job.
- Worked with various File Formats like Text File, Sequence File, ORC
- Files, Parquet Files and various Compression Formats like Snappy, bzip2.
- Designed both Managed and External Hive Tables and Defined static and dynamic partitions as per requirement for optimized performance on production datasets.
- Written Hive Queries for Data Analysis to meet the business requirements.
- Monitoring Data Validation.
- Environment: HDFS, Yarn, MR, Hive, Sqoop, Oracle, Shell Scripting.

Declaration

- hereby declare that the above information given is true and correct to the best of my knowledge and belief.

Date:
Place:Pune   (Kartik D. Rothe)