

Apache Hadoop with 5 Years of Experience

Professional Overview:

- 5 years of IT experience currently working with RBS(Royal Bank Of Scotland) as Data Engineer.
- working with Big data technologies in Automobile and Banking domains, capable of processing large sets of structured, semi- structured and supporting systems application architecture.
- Solutions focused, Meticulous and result-oriented professional
- Conversant with design, development, maintenance and support of BigData Analytics using Hadoop Ecosystem components like HDFS, Hive, Sqoop, Spark, Scala, Kafka and Oozie.
- Worked on multiple POC to evaluate different Hadoop architectures, tools and products in Hadoop.
- Experience in Data Ingestion, Data Filtering, Data Analysis, Data Modelling (Hive and Teradata), Data storage, data processing and data analysis.
- Good Communication, Interpersonal and Leadership skills.
- Proactive, Dedicated and Enjoy Learning New Technologies
- Working Experience in Agile, Scrum, Kanban project methodologies.
- Proficiency in interacting with clients to ensure delivery of a seamless solution that meets the client requirements.
- Instrumental in Teradata, COBOL, JCL, VSAM, CICS, DB2 and Mainframe tools (like Endeavor, File-Aid, CA7, SPUFI etc.).

Technical Expertise:

- Big Data Technologies: Hadoop, HDFS, Hive, Sqoop, Spark, Spark SQL, Oozie, Apache Kafka, Apache Flume, Git.
- Programming Language: Scala
- Tools: Eclipse IDE, GitHub, Hue, Cloudera Manager, Maven, Jira, Rally
- Also Knows: Java, Spark Streaming, Pig, MapReduce, Pivotal Cloud Foundry Hbase, SQL Server, Maria DB.
- SDLC: vAgile, Kanban, Waterfall
- Operating Systems: Unix/Linux, Z/OS
- Languages: COBOL, JCL and SQL
- Databases: Teradata, Maria DB, DB2, SQL Server and VSAM
- OLTP: CICS
- H/W Platform: IBM Mainframes
- Tools: XPEDITOR, ISPF, SPUFI, ENDEVOR, FILE-AID

Qualification :

B.Tech (Computer Engineering)

Certifications & Achievements

- RBS Star Award for my outstanding effort in various PoCs
- Asia Pacific Excellence Award for my outstanding contribution and effort in migrating GUDB project from Mainframe/Teradata to Hadoop.
- Awarded with Infosys Debutant Of The Year - 2015 Award for the contribution in my project – A Unit Specific Award
- Agile and Scrum – Internal Certification from RBS
- IBM Certified in DB2 10.1 Fundamentals database associate.

Projects :

Project: Core Outputs, RBS

Technologies Used: HDFS, Hive, Scala, Spark, Oozie, Kafka, Git, Sqoop, Flume, Maria DB

Project Brief: The Royal Bank of Scotland Group plc (also known as RBS Group) is a private, but partly state-owned, British banking and insurance holding company, based in Edinburgh, Scotland. The group operates a wide variety of banking brands offering personal and business banking, private banking, insurance and corporate finance through its offices located in Europe, North America and Asia.

The purpose of the project is to store terabytes of data related to customer accounts from input feed into hive tables. The data is stored in Hadoop file system and processed using Spark and Hive jobs. The required data is then exported into the Maria db. Downstream applications connect to our database for generating reports through tableau.

Responsibilities:

- Involved in all phases of the Big Data Implementation including requirement analysis, design, development, build, testing etc
- Write Sqoop scripts to import and export data between HDFS/Hive to Maria DB.
- Use Spark API over Hadoop YARN to perform analytics on data in Hive and HDFS
- Use Kafka and Flume to process near real time data from reference systems.
- Design solutions for various data extraction and data analysis requests.
- Perform Data enrichments such as Filtering, Sorting, Aggregation etc.
- Develop workflow in Oozie to automate the tasks of loading the data into HDFS.
- Use Agile scrum methodology as SDLC to achieve the deliverables.

Project: Global Unified Database, Ford

Technologies Used: Spark, Scala, Hive, Oozie, Sqoop, Git, Teradata, SQL Server, Kafka

Project Brief: Global Unified Database is used for providing data to FCSD (ford customer service division) corporate dashboard. Business uses this dashboard to look at various metrics and make business strategies using this. GUDB sends data to downstream interfaces like MRS to create these metrics.

Responsibilities:

- Responsible for delivering the project using AGILE Methodologies.
- Create Spark API over Hadoop YARN to perform analytics on data in Hive using HiveContext, finally saving it to HDFS and creating Hive tables on top of them generating the Business Metrics.
- To develop Hive scripts to cleanse and parse data in HDFS obtained from various data sources and load it in Hive tables.
- Creation of tables, data distribution by implementing partitioning and bucketing, writing and optimizing the HQL queries on Hive data warehouse tool.
- Using HIVE ACID properties and compactions to update data in HIVE tables.
- Setting up workflows and coordinators using Apache Oozie workflow engine for managing and scheduling Hadoop jobs.
- Automate sqoop for data landing from Teradata/SQL server to HDFS using Oozie and loading data in Hive tables.
- Monitor Hadoop cluster job performance and capacity planning.
- Support/Troubleshoot hive & spark scripts running on the cluster
- To export the business required information to RDBMS using Sqoop for BI team to generate reports.
- Requirement gathering and generating process layout.
- Perform the POC; develop the code and liaise with QC for testing, UAT and promote and maintain the code in Production.
- Propose best practices/standards.

Environment: Microfocus Net Express

Technologies Used: COBOL, JCL, DB2, Microfocus Net Express, AQT

Project Brief: Global Atlantic Financial Group, through its subsidiaries, offers a broad range of retirement, life and reinsurance products. PerformancePlus is primary compensation system is a comprehensive, flexible producer and office support system that is fully capable of computing and distributing commissions through various channels. PerformancePlus computes primary compensation, based on a wide range of formulas. It stores, maintains, and manages detailed information about distributors and their relationships. It also measures and manages distributor performance and profitability.

