Senthilkumar P
BigData Hadoop Developer

Professional Summary

- Having Around **8+** years of overall Professional IT experience in which around **3+** years of experience as **Big Data Developer** and around 5+ years of experience in **Oracle PL SQL**.
- Worked on **Real Time Streaming** using **Spark Streaming Context** as micro batches to process it
- Consumed Data from **NIFI** to **Kafka**, from **Kafka** using Spark Streaming Context as micro batches to process it at the real **time/In-memory**
- Worked on data migration from databases such as **Oracle** and **MySql** to Hadoop platform Worked on improvising the tuning options using **HIVE** functions such as **Partitioning**, **Bucketing**, **Index**, **CBO** etc.,..
- Experienced on big data software stacks such as **HDFS, MapReduce, YARN, Sqoop, Hive, Scala, Spark - Core, Sql, Data frame, Dataset** and **Spark Streaming**
- Developed **Hive Queries** to parse the raw data and store the refined data in partitioned tables and developed **UDF** in **HIVE** for a particular requirement. Handled variety of data by using collection items like **Struct**, **Map** and querying from hive tables.
- Handled **Hive performance optimization** by using **ORC/Parquet** storage, **Compression**, **Partitioning, Bucketing, Predicate Pushdown Data, CBO** query analytics and compute analytics on top of tables, queries, and partitions
- Worked on slowly changing dimensions **SCD** type **1** & **2**
- **RDDs**, **Data Frame, DataSet** and **Spark Sql** creation, **cache/persist**, **checkpoint RDD** for the source data chunks.
- Understanding to converts **RDD's** into data frame and store the data into **hive** table using data frame **API** in **Spark programing**
- Worked on **performance tuning**, benchmarking, followed best practices in the development process such as implementing **memory management**, **partitioning**, **broadcast variables**, etc.,.
- Understanding about data sets, read data different source files like **csv**, **Parquet**, **Avro**, **Json**, **Sequence files** and store it in **HDFS**/**Hive** tables using **Spark**
- Experience in working with data load from **HDFS** and **Linux** file system to **hive tables**
- Worked as an **individual contributor** for an end to end application delivery.
- Experience in using IDEs like **Intellij** and **Maven.**
- Hands on experience on writing **Queries, Packages, Stored procedures, Functions, Views, Materialized Views and Triggers** by using **SQL** and **PL SQL**.
- Good knowledge in **Oracle RDBMS**.
- Support development, testing, and operations teams during new system deployments.
- Explored an option to stored data into **AWS S3**

Professional Experience

- **TATA Consultancy Servies Ltd**, Chennai from Oct 2020 to till date.
- **Congnizant Technologies Services**, Chennai from Jan 2019 to Oct 2020.
- **Excellon software Pvt ltd**, Nagpur from Oct 2012 to Oct 2018.

Selected Accomplishments

- **Assistant Consultant** for **TATA Consultancy Servies Ltd Bigdata Platform**.
- **Senior Associate** for **Congnizant Technologies Services Bigdata Platform**.
- **Senior Consultant** for **Excellon software Pvt ltd Bigadata and Oracle Platform**.

Technical Skills

Batch Tools	**Sqoop, Hive and Spark**
In Memory Tools	**Spark Core, Spark SQL, Spark Streaming, NIFI, Kafka**
Control languages	**Oozie, Autosys, GitHub**
Programming	**Scala,Python, PL/SQL, .Net**
Operating System	**Windows, Unix, Centos**
RDBMS	**Oracle 10g, Oracle 11g**
Development Tools	**Intellij 2021.1, Toad 8.6.2, ,PL/SQL Developer**
Scripting	**Linux/Unix Shell Scripting**
Cloud Platform	**AWS**

Project Experience

Project # 1:
Client: VISA **Oct 2020 – Till Date**
Porject : UMF - Auth
Technical Environment: Spark Streaming, Spark SQL, Kafka, Hive, Oozie, Shell Scripting, Scala, Intellij

Project Scope: This project au-producer which consumes the data from UB and do minimal tranformation then write the data into kafka raw topic. Au -streaming consuming the data from kafka raw topic and it will do the lookup and entrichment then write the data into entriched kafka topic and K2H consuming the data from entriched kafka topic and split based on auth, fm, advice, reject then writing into HDFS. parallely audit process will run which consumes the data from au-streaming and write the topic of audit writer which will store the data in rocks DB and will do deduplication and audit reader will read the data from rocks db and K2H hdfs path and do the comparision. once audit proess done the data moved to down stream for further entrichment. Analyst can analyze this data to generate insights about individual consumer behaviors and preferences, and offer personalized recommendations such as

flexible offers and Loyalty and Reward Programs. Business need the ability to analyze dated and live data, as well as predict the future, to distill through and find out what are valuable, sight trends and share insights they may not even have imagined.

Roles and Responsibilities

- Consumed Data from **Kafka** using **Spark Streaming Context** as micro batches to process it at the **real time/In-memory**.
- Worked on **performance tuning, benchmarking**, followed best practices in the development process such as implementing **memory management, partitioning, broadcast variables**, etc.,.
- Creating **Hive** tables with appropriate **partitions**
- Working in **Lambda architecture**
- Worked on scheduling tools such as **Oozie**
- Explored an option to stored data into **AWS S3**
- Took care of **reviewing peers code** before moving to high region.
- Performing testing in development environment and sending it for **production**
- Performing/involving in all the activities of the **release process**
- Involved production support and fix the issues on time.

Project # 2:
Client: JP Morgan Chase **Jan 2019 – Oct 2020**
Project: Realtime Vehicle Controller and Analytics
Technical Environment: Sqoop, NIFI, Kafka, Spark – SQL, Stream, Core, Elastic Search, Kibana & Hive

Project Scope: This project handling streams such as Batch, Inmemory, Realtime Streaming and provides the realtime information of vehicle navigation, leverage real time traffic and driving pattern of the vehicles driven by the different drivers across the world. Data ingestion and acquisition is done through Sqoop and NIFI, Queued into distributed Kafka queue, Enriched using Spark, pushed into Elasticsearch, analyzed in realtime using Kibana and batch analytics will be done using Hive. We will calculate the wages on monthly, quarterly or annually basis, distance driven etc in realtime joining with the batch data and this system helps in real time tracking of vehicle through its GPS incorporated systems with help of which entire activity of the Fleet can be monitored. Real time tracking ensures safe and on time delivery. It also helps in navigation and leverage real time traffic to avoid delay. All the information gathered and provide the best business decision.

Roles and Responsibilities

- Importing and exporting data between **HDFS** (Hadoop Data Lake) RDBMS sources like **Oracle** using **Sqoop**.
- Using **HDFS** for persisting the data for primary staging.
- Creating **linux shell** script to simulate the realtime vehicle movement by reading data from events file and create smaller files.
- Creating **NIFI** data flow to read the data from the files created and push the data to **Kafka**
- Wrinting **Spark Program** to read the data from HDFS location where sqoop imported and read from **Kafka**
- create **dataframes**, **temporary views** in spark and join the sqoop data with the **kafka** data and persist into **Elastic search indices**.
- **Kibana** visualizations and dashboards created for reporting.

- **Tuning** import/export jobs to support faster data transfer between Hadoop & source systems.
- Creating **Hive queries** for analysis and involved in performance tuning of Hive.
- Processing Hive tables using **Spark** and making the final data available for visualization.
- Responsible to handle support issues posted by customer and issues posted by **QA** team
- Debugging using **spark shell** and understand the issues and provide solution for the issue and responsible for monitoring Autosys job scheduler
- Developing new features based on the feedback from customer and integrate with **API** for data processing
- Gathered the **business requirements** from the Business Partners and Subject Matter Experts
- Weekly meeting with business partners on product improvements
- Co-ordinate with team to get the **features/bugs** to be solved, **prioritization**, **planning**, task deviation analysis.

Project # 3:
Client: Ashok Leyland (Chennai) **Oct 2017–Oct 2018**
Project: CRE
Technical Environment: Hadoop, Sqoop, Hive

Project Scope: It is Data lake project that helps the oraganization to acquire, cure, store, process and visualize the enriched data. The data includes customer data, vehicles service data, inventory data, products, purchases, sales, orders, employees, stock and other reference data of the product. Retailers and business can analyze this data to generate insights about the product and customer behaviours, preferences, Recommendations and Loyalty programs. This tool assists our business teams in their Analytics & Information ecosystem journey right from strategy definitions to large scale global implementations & Support using the bigdata ecosystems. By Batch process, this system creates detailes view of sales,purchases,services,sale order and purchase order which perform processing and aggregation with the data provided across all branches to identify each and every customers needs and provide them on time. Structured and semi structured Data from different source systems are injected using sqoop into HDFS and HIVE managed tables as a staging and apply complex HQLs to join, aggregate, filter, convert and load the final data into final external flat and serde tables. Outbound feeds will be exported from hive and sent as feeds using scp or sqoop exported directly to the RDBMS systems.

Roles and Responsibilities

- Worked closely with business customers for **Requirement** gatherings.
- Developing Sqoop jobs with incremental load from **RDBMS** (**Oracle**) using native **dB** connectors
- Designed Hive repository with **external tables**, **managed tables**, **buckets**, **partitions** , and **ORC** compressions for incremental data load of parsed data for analytical & operational dashboards
- Experienced in developing **Hive Queries** on different data and partitioning to yields improvement in performance using **HiveQL**
- Tuning import/export jobs to support faster data transfer between **Hadoop** & **source systems**.
- Preparing **shell scripts**
- Developed Oozie workflow for scheduling and orchestrating the **ETL** process
- Responsible of Importing data from Oracle to HDFS and provided the query capabilities using **HIVE**
- Coordinate with onsite business analyst and application **developers**, attend **client meetings**.

Education Details

Completed **B.E** (Electrical and Electronics Engineering) in **KLN College of Information Technology**, **Madurai, Tamilnadu, India**.

Personal Info:

Date of birth	30-05-1986
Sex	Male
Nationality	Indian
Marital Status	Married
Launguage Known	English, Tamil [Read,write & speak]

Date:
Place: Chennai **(P. SENTHILKUMAR)**