Siva Kumar (Big data Developer)

Email Id : [sivaeee1992@gmail.com](mailto:sivaeee1992@gmail.com)

Contact Number : +919942120547

**JOB PROFILE:**

* Current Role : Associate Consultant
* Company Name : Capgemini
* Experience : Around 4 years
* Client : A C Nielsen(Retail Domain), SCOPE(Banking)

**TECHNICAL SKILLS:**

* Big Data Ecosystems : Spark-core, Spark-SQL, Spark-Streaming, Nifi, MapReduce, HDFS, Hive, Sqoop, Impala, Oozie
* Programming Languages : Java, Scala, SQL, HTML, Tibco BW
* Databases : Oracle, Netezza 10g, MySQL 5.5
* Tools : Eclipse IDE, Putty, FileZilla, DbVisualizer
* VersionControl : SVN
* SDLC : Agile, Waterfall
* Platforms : Windows and Linux

**WORK EXPERIENCE:**

**Big data Developer:**

* Strong Knowledge in File Distributed System in Big Data Arena.
* Good Understanding of the complex processing needs of the big data and familiar with the Hadoop Ecosystems.
* Expertise in developing Spark, Mapreduce Programs and Querying using Hive Query Language.
* Good Experience in using Sqoop commands for importing data into HDFS from RDBMS.
* Experience in working with Oozie workflows.
* Good Experience in Optimization of Spark Jobs.
* Worked with Avro, Json, Parquet and ORC type of files.
* Exceptional ability to learn new concepts.

**Experience Summary:**

|  |  |
| --- | --- |
| **Tata Consultancy Services** | February 2014 – September 2016 |
| **Capgemini** | September 2016 - Till date |

**Project Details:**

**Project 1:**

|  |  |
| --- | --- |
| **Project Name** | **SCUDEE** |
| **Clients** | **SCOPE (Standard Chartered Bank)** |
| **Period** | Sep 2016 to Till date |
| **Technologies & Tools** | Spark, Java, Scala, NIFI(HDF), Hive, AWS |
| **Role** | Developer |

**Project Description:**

SCOPE sends offline batch and online data to various downstream applications and warehouses. Till date the source systems are sending redundant information to same or various downstream which increases longer batch run for the source system impacting the end of day process.

The main objective of the project is to build the data ingestion platform over the Data Lake. We are using NIFI for the file validations, Job orchestration and Spark for the processing layer.

**Roles & Responsibilities:**

* + Participated in design analysis, gathering requirements from BA Team and developed the project from the scratch .
  + Used Nifi to validate the incoming files, Job orchestration and for Job scheduling..
  + Developed Spark programs in Scala and Java for the processing layer where we process records depending on various type of files using Dataframes.
  + Worked with Json, Avro and ORC files.
  + Optimized the Spark programs by providing effective Spark Submit parameters to run on Yarn-Cluster mode, reducing shuffles and repartitioning the data by seeing the DAG flow.
  + Fine tuned various application's Spark job's to effectively use the cluster resources and attained better performance benchmarks.
  + Good in using Dataframe API's and RDD operation's.

**Project 2:**

|  |  |
| --- | --- |
| **Project Name** | **New\_Foundations** |
| **Clients** | **Ac Nielsen** |
| **Period** | March 2015 to Sep 2016 |
| **Technologies & Tools** | HDFS, Mapreduce, Hive, Sqoop, Oozie |
| **Role** | Developer |

**Project Description:**

The Objective of the project is to handle large amount of incoming data and processing , We are migrating the data from Netezza to HDFS and using Hive to process the data.

The entire job is configured in oozie workflow.

**Roles & Responsibilities:**

* To migrate from Java and SQL to Hadoop MapReduce and Hive.
* Developed Map Reduce application using Hadoop, Map Reduce programming.
* Developed Parsers for the Map reduce programs using Customized Record Readers.
* Tuning Map Reduce jobs for efficient and effective processing.
* Load Data to External/Managed tables by using Hive scripts and performing aggregations and transformation using HIVE queries.
* Creating Hive Schema and Partitioning.
* Coding programs for transferring huge files from client terminals to HDFS to enable data processing and analysis.

**Project 3:**

|  |  |
| --- | --- |
| **Project Name** | **Foundations** |
| **Clients** | **Ac Nielsen** |
| **Period** | May 2014 to March 2015 |
| **Technologies & Tools** | Java, SQL, LINUX, Oracle, Netezza and MySQL. |
| **Role** | Developer |

**Project Description:**

A C Nielsen is market research company. We are providing ratings for US televisions based on the data provided to us.

**Roles & Responsibilities:**

* Understanding the Business Requirement as provided by the client.
* Developed Java Codes and SQl Queries.
* Fixing the bugs as reported by the clients.

**POC**:

|  |  |
| --- | --- |
| **Project Name** | **SparkSQL POC** |
| **Clients** | **Ac Nielsen** |
| **Period** | March 2016 to May 2016 |
| **Technologies & Tools** | SparkSQL, Impala |
| **Role** | Developer |

* The objective of the POC is to migrate the Impala queries into SparkSQL.
* To improve the performance of queries, We have used warmCache and Temptable techniques in SparkSQL.
* In Impala the intermediate results are stored as temp tables but In SparkSQL We have used DataFrame’s to store those temp table results which will make the great difference in performance.
* And loaded the final results as parquet files.

**Education:**

* **B.E(Electrical&ElectronicsEngineering),**M.Kumarasamy EngineeringCollege,Dec2013**(80%)**
* **12thStd,**Jaycees MatriculationSchool,2009**(84%)**
* **10thStd,**Jaycees MatriculationSchool,2007**(75%)**

**PersonalDetails:**

DateofBirth :09-12-1991

Nationality :Indian

Sex :Male

MaritalStatus :Single