BESKI ALVA EDISON I Email:[Beskimca@gmail.com](mailto:Beskimca@gmail.com)

Mobile: +91-9566353234

Experience summary

Software professional with 2 years of experience in the IT industry and extensive experience in analysis & design, coding, development, implementation and maintenance of business technology solutions for various business and customer Applications.

* Having experience in Hortonworks and in analysis, Design, Development, in Software application which includes Big Data using Hadoop **HDFS, Hive, Sqoop, Spark SQL ,PIG, HBase, Phoenix, Kafka, NiFi, YARN** and **MapReduce** Program.
* Experience in importing and exporting data using Sqoop from Relational Database Systems to HDFS.
* Experience in coding Hive queries & Pig scripts.
* Extensive experience in analyzing data using Hive, Pig Latin.
* Experience with the Hadoop stack (HBase, Oozie, Sqoop, Hive, Flume, Pig) .
* Integration from Hive to HBase & Pig to HBase.
* Improvising the tuning options using Hive functions such as Partitioning, Bucketing ,Index
* Deep understanding and implementations of various methods to load Hive tables from HDFS and Local File System.
* Worked end to end in optimized Sqoop data extraction from heterogeneous databases (MySQL, Oracle 11g) and real time capture with Kafka, Spark Streaming and data refinery with Pig & Hive and data repo with Hbase, Hive & HDFS and fast computation implementation using Spark over YARN.
* Worked extensively with Sqoop for importing and exporting data from Oracle database to HDFS and Hive.
* Partitioning & Buckets have been used to improvise the tuning options using Hive.
* Handled various data sources like RDBMS, Web Server Logs and Flat file formats.
* Worked extensively on Message queuing tools like **Kafka** and **No sql**databases like **HBASE.**
* Capable of processing large sets of **Structured, Semi-structured** and **Streaming data.**
* Worked on various File formats like **Avro**, **Parquet**, **Text** and **ORC** format.
* Possess knowledge and having diversified experience in Banking domain.
* Comprehensive technical, oral, written and communicational skills.
* Strong data base experience in **MS SQL Server.**
* Have good problem solving and analytical skills and love to innovate in order to perform better.

Professional Experience

* Junior Analyst Programmer at Inceptez Technologies Pvt Limited, Chennai - Jun 2016 to Till Date

Technical skill set

**Languages** : Scala, SQL, Shell Scripting.

**Databases** :Oracle, mysql.

**Hadoop System** : MR 2.0(YARN), spark, hive, pig, Hadoop, HDFS, Sqoop, Kafka, Hbase, Phoenix

**Databases :** MS SQL Server, MySQL

**Organization: Inceptez**

**Description: Inceptez** is one of the leading IT training, Development and staffing company specializing in Big Data, Data Science, Dev-Ops, Enterprise Mobility, Cloud Computing and Internet of things (IOT).

PROJECT PROFILE

|  |
| --- |
| **Project #1 IPS Data Engine June 2016 to July 2017** |

**Role :** Hadoop developer**.**

**Client :** Hollard Banking,south Africa.

**Technologies :** Hadoop, HDFS, HBase, Sqoop, Oozie, MapReduce, Hive, Pig, Phoenix

**Description:**Client is a leading South Africa banking organization and data involved in Pension Payments of external clients of Key Bank. Moving forward client decides to go with Hadoop instead of Mainframe and other traditional databases. Source Data is ‘n’ number of databases. The target is to make all the data to be present in converged platform, so we planned to go with Hadoop platform for efficient way storage and processing with the cost effective way.

**Roles and Responsibilities:**

* Handled importing of data from ‘n’ number of data source viz oracle, Mysql, DB2 etc. performed transformations using Hive, loaded data into **HDFS** or **Hbase** based on the requirement and Extracted the data into Hive using **Sqoop**.
* Experienced in developing **Hive Queries** on different data formats like Text file, CSV file, Log files and leveraging time based partitioning yields improvement in performance using HiveQLand running **Pig** scripts (Pig Latin) to study customer behavior.
* Designed **Hive** repository with **external tables, internal tables, buckets, partitions** for data load of parsed data for analytical & operational dashboards
* Involved using Apache Sqoop for efficiently transferring bulk data between HDFS and relational databases for product level forecast.
* Experienced on loading and transforming of large sets of structured, semi structured and unstructured data.
* Involved in achieving the performance tuning of the queries by using partitioning, buckets, Index, Vectorization, Hive on tez, ORC format and cost based optimization techniques.
* We were also analyzing the data of customers to find the Frustration scoring, so that we will work on the business pain points to improve the business.
* Also we will receive the variety of files on adhoc basis from client to convert it into structured format, so that it can be uploaded in the respective websites.
* **Phoenix** to create a table on top of **hbase** table to perform low latency queries.

|  |
| --- |
| **Project #2 Truck data Analytics in real time GPS Aug 2017to Jul 2018** |

**Role :** Hadoop/Spark developer

**client :** Orbit Showtime Network, South Africa

**Tools :** Apache Spark, Sqoop, Hive, Hbase,NIFI,Kafka,Kibana,Unix,Elastic search,Shell Scripting.

**Description:** On Board Computing (OBC) technologies are designed to track and improve various elements of fleet performance. Private Fleets are using these technologies to increase operational efficiency. It is being used to monitor a driver’s service hours electronically. Instead of relying on an inefficient paper logging process, this technology keeps better accurate records while being more user-friendly to the driver. It keeps track of date& time details, odometer readings, fuel pumped and usage details, states travelled. These details are helpful in calculating Fuel and Mileage Tax amount for every customer on monthly, quarterly or annually basis and reports are generated as per customer desires

**Project Flow:**

**Batch:**

* **Sqoop** for injecting the drivers data from database.
* **HDFS** for persisting the data for primary staging.

**Realtime:**

* Sensors(IOT device) data will passed to kafka through NiFi.
* **NIFI** data flow to read the data from the files created by vehicle sensors and push the data to **Kafka.**
* **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.

**Roles and Responsibilities:**

* Started Working on NiFi, kafka and spark to develop the code.
* Importing and Exporting data into HDFS using Sqoop
* Used various methods to load Hive tables from HDFS and Local File System
* Written Pig queries to change the format of fields that are stored in HDFS
* Participated in Designing the Architecture of the project
* Analyzed data using Spark engine components
* Responsible for running Streaming Jobs to process terabytes of Streaming data
* Load and Transform large sets of Structured, Semi Structured and Unstructured data using Hadoop/Big Data concepts
* Responsible for Creating Hive tables, Loading data and Writing Hive queries
* Done Enhancements to production jobs in order to meet the customer requirements
* Used Kafka as a message queue and passes the data

Education background

* Master of Computer Application from Sri Manakula Vinayagar Engineering College, Pondicherry, India

personal information

Name : Beski Alva Edison

Place : Chennai