Lakshman Sai

balinenilakshman@gmail.com
Mobile: +91-9848196579

Professional Summary:

- **Having 5.0 + years of experience in IT Industry.**
- Strong knowledge **on Spark RDD, Data Frames** and **Spark optimization techniques**.
- Having good experience and knowledge on **Azure Data Engineering Services** such as **Azure DataFactory**, **Azure Databricks**
- Excellent experience and knowledge in **Apache Hadoop** ecosystem components like **Hadoop** Distributing File System **(HDFS), MapReduce, Hive.**
- Excellent experience and knowledge in **Python, PySpark in Databricks** environment.
- Strong experience on developing **Hive queries, execution plan analysis and performance tuning**.
- Hands on experience in **developing ETL workflows using PySpark**.
- Working experience with **Agile SCRUM** and SDLC.
- Flexible, enthusiastic and project-oriented team player with good communication and solidinterpersonal skills.
- Able to work independently and collaborate proactively & cross functionally within a team.

Educational Summary:

- Bachelor of Engineering in Information Technology from Anna university 2017.

Professional Experience:

- Working as **Azure Data Engineer** in Spruce Infotech client (Infosys) from Feb 2022 to till date.
- Worked as **Azure Data Engineer** in Quantzig from 8th Feb 2021 to 7th Jan 2022.
- Worked as **Data Engineer** in Predifast Technologies, from june 2017 to Feb 2021.

Technical Skills:

Hadoop Eco System : Hive, Sqoop, Impala,
HDFS, Yarn Spark : Spark Core, Spark SQL
Azure : Azure Data Factory, **Azure Databricks**
Programming Language : Python
Scripting : Bash, Python

Project -1 May 2020 To Present

Title : Cloud Analytics Platform (CAP)
Role : Azure Data Engineer
Environment : Azure **DataFactory**, Azure **Databricks**, Azure **KeyVault**, Azure SQLServer, PostgreSQL, **Airflow**.

Description:

DAP is an analytics platform that allows you to bring any source data into Azure cloud to perform analytics on it. It serves as a data lake for business data and gives insights on it. Once data has been imported into data lake, it can be accessed via data consumption layer.

Responsibilities:

- Developed a Notebook using PySpark that will capture the business metadata and store into a PostgreSQL to provide insights as well as user access via Azure Data Catalog.
- Developed an ADF Pipeline that has been imported data from various external data sources into ADLS.
- Involved in creating secrets in Azure KeyVault to provide centralized access/security.
- Written a PySpark code that will capture the metadata of files that were stored on ADLS using Spark and ADLS File API to store information in PostgreSQL.
- Using PySpark captured schema from YAML files and created a table using autogenerated DDL in PostgreSQL to provide metadata access to client/user.
- Written Apache Airflow code to orchestrate the complete flow of ADF and invocation of Notebook from it.

Project -2 May 2019 To Mar 2020

Title : Geni Data Lake
Role : Data Engineer
Environment : Hive, Spark, Hive, Sqoop, MySql, Python, PySpark

Description:

The primary motive of this project is to bring the Trading Data into one platform which is being generated by multiple Trading Systems. Generally the source data is in system specific format. As the part of this project, we are collecting data from all the source systems, Performing data cleansing, unification and Transformation using Spark and storing the data in Data Lake (Hive Tables). The hive tables are exposed to the downstream applications who generates the Reports. The source data is also archived for the future purpose.

Responsibilities:

- Developed Sqoop jobs to pull data from Different RDBMS Systems like Oracle, Teradata and MySql.
- Developed Sqoop jobs to pull data from Different RDBMS Systems like Oracle, Teradata and MySql.
- Analyzed performance issues in spark application, optimized the spark code and resource allocation.
- Analyzed performance issues in spark application, optimized the spark code and resource allocation.

Project -3 October 2018 To April 2019

Title : BMAS
Role : Data Engineer
Environment : PySpark, Hive, Hadoop, Mysql.

Description:

BMAS is a Bharati Airtel Network project, which holds Airtel network data related to tower locations.

Responsibilities:

- Developed a parser using Spark SQL that will prepare a composite primary key using the fields located in a CSV files.
- Developed a Spark SQL parser that will un-compress a zip file having CSV files and loads them into a MapRDB table.
- Involved in production deployment of Spark code.
- Communicated with client located in USA and collected requirements to provide a business solution.
- Prepared necessary documentation required for productionizing the code.

Project -4 June 2017 to Aug 2018

Title : ServiceNow Archival , WIH Archival
Role : Hadoop Developer
Environment : HDFS, MapReduce, Pig, Hive, Sqoop, Mysql

Description:

ServiceNow is a Data Archival project that does a job of importing data from MySQL to MapR DB. This project allows BI folks to get access to MapR DB tables from Tableau tool to generate dashboards and reports.

Responsibilities:

- Written Python code to automate the process of generating Apache Drill View statements.
- Involved in writing Sqoop Scripts for importing data from MySQL To MapR DB.
- Written Python script for running a set of Sqoop scripts together.
- Written Hive scripts to integrate MapR DB Tables with Hive Tables to provide SQL access to data from a NoSQL Store.
- Configured hive metastore in MySQL to provide access to multiple clients.
- Developed Python code to communicate with Drill and create views in Drill Server.
- Written scripts to automate the process of moving of MapR DB tables data from development to Production.
- Developed a Python code that converts an Excel data into a JSON format.

- Executed MapR commands to provide access to MapR DB tables data of Development in Production.
- Rewritten Sqoop scripts for few tables to fix Java Heap Space problem while importing data.
- Coordinated with MapR team in resolving issues while in development as well as in Production.