Kasimani

+91- 9360802365| kasimani.ds01@gmail.com

Experience: Total –7 Years| Spark developer

- Possess over 7years of IT experience including over 3 years of experience working in Big Data technologies providing solutions to Healthcare, Pharma & Life Science, Retail and Digital Marketing industries.
- Extensive experience in analyzing, brainstorming and migrating use-cases into Big Data relevant technologies.
- Experience in architecting solution/product/application development/maintenance
- Extensive experience of implementing end to end Scalable, Distributed, HA, MPP systems on AWS / CTL Cloud, to process large data sets.

Skills

Big Data Stack	Apache Spark 2.x(Spark SQL,Pyspark),Scala, Python,
No SQL& RDBMS	Cassandra, HBase, AWS S3,Redis, Oracle, SQL Server, Sybase
Language & Analytics tool	Scala, Java,Python, Unix Shell Scripting, SAS, Tableauand MISTR
Cloud	**A**WS(EC2, EMR, S3, GLUE, RedShift, Athena, Lambda function, Redis, Gateway API, QuickSight) , AZURE(HDInsight, datalake, Blob services, Data factory, etc)

Experience

IT Experience:

Senior Software Engineer at **ICONIC E-TECH**, Chennai from **19th OCT 2015**.

Project Experience

ETL Processor **Aug 2021 to Oct 2022**

ETL processor developed in Spark to implement extract the data from different sources and does the transformation based on the configuration and writes the output to the target system. All the process will be handled in parallel. Based on the configuration entries in db table, Spark system will connect to the source and target and based on the column mapping config, UDF will be applied on the source tables and final output generated.

Role: Spark development (Module Lead)

- **Development of data related instruments / instances**: Use programming skills to develop, customize and manage integration tools, databases, warehouses, and analytical systems
- **Data pipeline maintenance / testing**: During the development phase, test the reliability and performance of each part of a system
- **Manage data and meta-data**: Accountable for storing data in a warehouse either in a structured or unstructured way; additional storage may contain meta-data (exploratory data about data); in-charge of managing the data stored and structuring it properly via database management systems
- **Provide data-access tools**: Responsible for setting up tools to view data, generate reports and create visuals
- **Track pipeline stability**: Monitoring the overall performance and stability of the system as the warehouse needs to be cleaned from time to time; the automated parts of a pipeline are monitored and modified since data / models / requirements can change

Environment: Spark, Scala, Python, Cassandra, EC2, Python, AWS S3, AWS GLUE, Athena, Redshift, PySpark, GitHub, TeamCity, Udeploy, AutoSys, HBAse.

Inventive Analytics **March 2020 to May 2021**

Inventive Data Analytics is designed and developed using Lambda architecture to handle massive quantities of data for both batch and stream processing. Data collected from IOT devices are stored in Sales force cloud. Data in sales force cloud is pushed using Kafka for stream processing and in the meantime, do the batch process using Spark Sql. Both the current data and batch processed data are stored in Cassandra for doing the analytics.

After the data analytics, the results are displayed in the MISTR dashboard for the report generation.

Role: Spark consultant

- Assembling large, complex sets of data that meet non-functional and functional business requirements
- Identifying, designing and implementing internal process improvements including re-designing infrastructure for greater scalability, optimizing data delivery, and automating manual processes
- Building required infrastructure for optimal extraction, transformation and loading of data from various data sources using AWS and SQL technologies
- Building analytical tools to utilize the data pipeline, providing actionable insight into key business performance metrics including operational efficiency and customer acquisition
- Working with stakeholders including the Executive, Product, Data and Design teams to support their data infrastructure needs while assisting with data-related technical issues

Environment: Spark2.4 (Streaming, Spark SQL),Scala, Kafka, AWS S3, Cassandra,Python, Shell Scripting, Datastax, SBT, Eclipse, JIRA, Github and MISTR Reporting.

Health science Reporting March 2019 to Jan 2020

Health reporting internal BIGDATA pipeline where multiple Business lines can provision their data and perform analytics or reporting on top of the data leveraging the Hadoop. In this project we developed a data pipe line which can seamlessly extract data from structured sources like oracle 11g and Teradata and unstructured sources like web service logs into hdfs and hive by integrating SPARK Cluster on top of YARN. The data extraction jobs can be schedule by a scheduler which was built on top of oozie. Both row level and column level data changes are captured and the effected data is pulled onto hdfs. Tableau Visual Analytics is used on top of hive data for creating dash boards for analytics.

Role: Lead Data Engineer

- Develop ETL jobs to load data from Enterprise data warehouse to Data mart, Used DS
- parallel extender's ability to design the DS objects to run efficiently using optimal resources
- Used Environment Variables, Project parameters, job parameters, apt config file, Stage
- Variables and Routines for developing Parameter Driven Jobs.
- Understanding the existing rules of analyzing risk and develop a strategy (ETL) to reduce
- false positives
- Leading Offshore team
- Preparation of the estimates, time lines of the deliverable and project execution plan.

Environment: Spark(Spark SQL, Spark-Shell), Scala, AWS GLUE, Redshift, Shell Scripting, Linux - Cent OS, Yarn, SBT, Eclipse, JIRA, Github and QuickSight.

Power system Analytics: June 2018 to Jan 2019

The smart electricity grid enables a two-way flow of power and data between suppliers and consumers in order to facilitate the power flow optimization in terms of economic efficiency, reliability and sustainability. The aim of the project is to establish infrastructure that permits the consumers and the micro-energy producers to take a more active role in the electricity market and the dynamic energy management (DEM). The challenge in a smart grid (SG) is how to take advantage of the users' participation in order to reduce the cost of power. We are in the process of exploring intelligent methods and solutions for the real-time exploitation of the large volumes of data generated by a vast amount of smart meters.

- Analyze the Business Requirements and Involved in Analysis, Design, Development, UAT
- and Production phases for new modules and enhancements of the application.
- Develop ETL jobs to load data from Enterprise data warehouse to Data mart, Used DS
- parallel extender's ability to design the DS objects to run efficiently using optimal resources
- Used Environment Variables, Project parameters, job parameters, apt config file, Stage
- Variables and Routines for developing Parameter Driven Jobs.

Role: Lead(Big Data)**Environment:**HDFS, Hive, Flume, Storm, Maven, MapReduce, Java, Linux, Oozie, Cloudera, Yarn.

Kraft RecipiesApril 2017 - May 2018

Kraft recipes is a site to host various recipes, dinner ideas and quick and easy meals from Kraft foods. It mainly focuses on the digital marketing their products of Kraft foods. User can browse the recipes by meal, recipe categories and featured collection. This application is interaction with the social Medias. User can add their favorite recipes to their recipe box. They can provide their feedback.

By analyzing customer ratings, usages, financial information, some useful reports like Fast & slow moving product, profitable products report, product weakness report, Area specific product promotion, product coverage report etc. were created. These reports would help to build a stronger business especially focusing on customer needs and also to run the profitable business.
Role: Hadoop Developer
Environment: ApacheHadoop 2.0.x, Pig, Hive, Sqoop, Flume, MapReduce, HDFS, LINUX, Oozie, Java, Eclipse.

Primary *Activities*:

- Analyze the Business Requirements and Involved in Analysis, Design, Development, UAT
- and Production phases for new modules and enhancements of the application.
- Develop ETL jobs to load data from Enterprise data warehouse to Data mart, Used DS
- parallel extender's ability to design the DS objects to run efficiently using optimal resources
- Used Environment Variables, Project parameters, job parameters, apt config file, Stage
- Variables and Routines for developing Parameter Driven Jobs.
- Understanding the existing rules of analyzing risk and develop a strategy (ETL) to reduce
- false positives
- Leading Offshore team
- Preparation of the estimates, time lines of the deliverable and project execution plan.
- Analysis of the data sources.