

Address: Noida, India **Mobile:** +91-9711686323 **E-mail:** ramsayhaysinghy0@gmail.com

Ramsahay Singh Yadav
Python/Scala/Spark/Hadoop/DevOps/Data Science

Technical Analyst with 11 **years** of overall and 5 **years** of relevant experience.

Strong emphasis on Design, Development, Implementation and Deployment of Software Application in HDFS, Hadoop Ecosystem and RDBMS. Professional in Big Data database design - Cassandra, MongoDB, Redis, Hive. Good Knowledge of Microsoft Azure Data Factory. Worked on Data Migration Application from RDBMS to NoSQL. Exposure to Stored Procedure, Triggers, Cursors and Functions for MySQL. Created Customized reports using Report Builder, Power bi Reports. Hands on Experience with Spark Streaming Applications using Pyspark and Scala.

Education:

- **Master of Computer Applications (Rajasthan Technical University, Kota)**
- **BCA [Bachelor of Computer Application]**

Technical Skills

Languages	Java, Python, Scala, HTML, PHP, Java Script
Big Data Technologies	Hadoop, Sqoop, Hive, pig, Mongo dB, HDFS, Map Reduce, spark, Scala, Pyspark, ETL, python, zookeeper, Cloudera, Hortonworks, Data Science, Power BI, SSIS package, YARN, Oozie, GCP, Microsoft Azure, Data warehousing ,AWS ,Data bricks
Cloud service	AWS, S3, EC2, EMR, ETL, GLUE, Azure, Kinesis, Data Pipeline, Load balancers, Security group. Microsoft Azure Cloud Services - Azure Databricks, Azure Data Factory, Data Lake Storage.
XML Technologies	SOAP, Rest.
Reporting Tools	Tableau, PowerBI
Database	MySQL, PL/SQL, Oracle, Mongo DB, HBase, Snowflake db.
Tools	Eclipse, NetBeans, IntelliJ idea, Eclipse, Edit Plus, Core Ftp, WinSCP, Jira, Microsoft Office,
Operating System	Windows, Linux, Unix
Version Control	SVN, GIT, Bit bucket
Methodologies	Agile, Waterfall

Current Employee: Coforge Ltd.

Technical Analyst

Professional Experience:

1. Working with **Coforge Ltd** as a Developer from 17 Feb 2021 to Till Date.

Role: Technical Analyst

Company Address: **Coforge Ltd.** SEZ Developer Unit Plot No. TZ-2 & 2A, Sector Tech Zone,
Greater Noida, UP 201308, India Ph: +91 (120) 459 2300 Fax: +91 (120) 459 2301

2. Working with GSR Business Services (P) Ltd as a Developer from 23 November 2018 to 16 Feb 2021.

Role: Sr. Big data developer

3. Working with Surmount Softech Solutions Pvt. Ltd. as a Developer from 18 March 2018 to 22 November 2018.

Role: Big data developer

4. Working with Geeta infotech India private limited as a Developer from Dec 2011 to 2018 March 15.

Role: Web Developer

PROJECT DETAILS

PROJECT # 1 (COFORGE LTD)

Title : Retailer product

Client : Walmart

Project Type : Development

Designation: Technical Analyst

**Technology: Hive, spark streaming, SparkSql, YARN, data bricks, Kafka, Scala, HDFS, Azure SQL server
Microsoft Azure (Cloud).**

Roles & Responsibilities:

- Load and transform large sets of structured, semi structured and responsible
- to manage data coming from different sources.
- Involved in creating Hive tables, loading with data and writing hive queries
- which will run internally
- Data processing purpose using Spark (Spark 3.0.1)
- Used JIRA to get the requirement from the customer along with dependent
- metadata information
- Used INTELLIJ for code generation tool to generate a JAR

PROJECT # 2 (GSR BUSINESS SERVICES (P) LTD)

Client : Bechtel

Project Type : Development

Designation: Sr. Big Data Developer

Technology: Hive, Spark, SparkSql, YARN, Oozie, Scala/java, Pyspark, HDFS, Azure,ADF.

- Working to create ETL workflow to migrate data from different sources like Oracle DB, SQL Server, Flat Files into Azure Data Lake, data bricks.

- Design and develop data pipelines in **Azure Data Factory** to migrate data from various sources to Azure SQL server.
- Move pipeline into production environment using templates.
- Customize pipeline in ADF for logging and monitoring.
- Perform enhancements in Data Lake as per client requirement.
- **Scheduling, Monitoring, and fixing bugs in ADF pipeline.**
- Generate ETL workflow documents for Data Lake
- Wrote the script files for processing data and loading on to HDFS.
- Wrote CLI commands using shell script.
- Developed the UNIX shell scripts for creating the reports from **Hive** data.
- Created the **External Hive Table** on top of the parsed data.
- Responsible for writing **Hive** queries for analyzing data in the **Hive** warehouse using **HiveQL**.
- Moved all log/text files generated by various products into the HDFS location.
- build aggregate tables that help the business user in decision making.
- Converted HQLs to the **Spark** code with the help of **SqlContext** and **Data Frame**.
- Responsible for preparing project documents and updating the repository.
- Worked on semi-structure dataset like JSON.

PROJECT # 3 (GSR BUSINESS SERVICES (P) LTD)

Title : Digital Supply Chain

Client : Bechtel

Project Type : Development

Designation: Sr. Big Data Developer

Technology: Hive, Spark, SparkSql, YARN, Oozie, Scala/java, Pyspark, HDFS, Oracle, AWS(Cloud).

Description:

- Bechtel is an engineering, procurement, construction, and project management company. DSC transaction data is currently present in Oracle and it is required to move the data in Azure Cloud for reporting.

Roles & Responsibilities:

- As a part Data acquisition, used Sqoop to inject the data server to
- transformation to create new features.
- In data exploration stage, used hive and impala to get some insights about
- the customer data.
- Used Kafka, Sqoop, Hadoop, spark for building data pipeline.
- Load and transform large sets of structured, semi structured and responsible
- to manage data coming from different sources.
- Involved in creating Hive tables, loading with data and writing hive queries,
- which will run internally in map way and written Hive UDFs.
- Preparing the filtered data ready for business analysis in text format and

- some data back to MySQL database in a separate table for Analysis as per the requirement using Sqoop
- Data cleaning and data preprocessing using spark (Spark and Spark SQL)

Project # 4 (Persistent)

Title : Module N

Client : Novartis Pharma

Project Type : Application Maintenance & Development

Designation: Big Data Developer

Technology: Hadoop, HDFS, Hive, SparkSql, Hue, Oracle, AWS.

Description:

Novartis is Pharmaceutical Company. Module N is doing Revenue Management only. Whatever contract patient has signed with Health insurance companies, such kind of data we are getting and doing analysis of data according client requirement. Whatever rebate insurance companies are providing to patient and Novartis, we are keeping track of all data and doing analysis of it.

Roles & Responsibilities:

- Joined as a big data application developer in this project.
- Worked with technologies like Spark, Python, Azure and RDBMS.
- Created ingestion framework for data at source and used blob bucket for the same.
- Defined layers and transformation strategy to produce the refine data and use it as view.
- Worked with Pyspark/Scala/java code to handle the data.
- Worked with team to deliver at time.
- Jira task board followed to handle task on day-to-day basis.

Project #5 Internet Prahari

project (Analytic using Big Data and Spark)

Project Type : Application Maintenance & Development

Role: Big data Developer

Project Description:

This Project proving web interface dashboard for forecasting analytics of different department (HR, INVENTORY, OPERATION, PERSONAL, FINANCE and ADMIN). Source of the data is an application which run within Client Domain. This project involved migration of Structure Data from MSSQL and put it into HDFS Cluster by using Sqoop tool. Hive is used for the ETL process and power bi is used for the visualization tool.

Roles & Responsibilities:

Involved in gathering the requirements, designing, developing, and testing.

- Developing Spark Programs using Scala API's to compare the performance of Spark with Hive and SQL
- Used Spark API over Cloudera YARN to perform analytics on data in HIVE
- Implemented Spark using Scala and SparkSQL for faster and processing of data.
- Developed Spark job to parse the Json Data.
- Load the data into Spark RDD and do in memory data computation to generate the output response
- Involve in converting Hive/SQL queries into spark transformations using Spark RDDs.
- Developed spark scripts by using Scala shell commands as per the requirement
- Implemented spark using **Scala and utilizing Data Frames and spark SQL** API for faster processing of data
- Used spark-sql to load JSON data and create schema RDD and loaded it into hive tables and handled structured data using **SparkSQL**.
- Created **Hive** tables to store the processed results in a tabular format.
- Developed **Sqoop** scripts in order to make the interaction between MSSQL database and **HDFS clusters**.
- Wrote the script files for processing data and loading on to HDFS.
- Wrote CLI commands using shell script.
- Developed the UNIX shell scripts for creating the reports from **Hive** data.
- Created the **External Hive Table** on top of the parsed data.
- Responsible for writing **Hive** queries for analyzing data in the **Hive** warehouse using **HiveQL**.
- Moved all log/text files generated by various products into the HDFS location.
- Designed **Oozie** workflows to automate the entire process of pulling the data from core tables to build aggregate tables that help the business user in decision making.
- Converted HQLs to the **Spark** code with the help of **SqlContext** and **Data Frame**.
- Responsible for preparing project documents and updating the repository.
- Worked on semi-structure dataset like JSON.
- Used different file formats like ORC, Parquet, Sequential and TextFile to store data in HDFS.

- Involved in working with different compression techniques like Gzip, Codec Snappy

PERSONAL DETAILS

Date of Birth : 10/06/1987

Gender : Male

Languages Know : Hindi, English

Permanent Address : Vill-Parasahi, Post-Husainpur, Distt-Ghazipur (U.P)

Pin code : 233001

Pan Card Number : AGIPY5010G

Passport Number : S6770475

I hereby declare that the information furnishing above is true to the best of my knowledge and belief.

Date:

Signature

Place- Delhi

(RamSahay Singh Yadav)