

Roshan Singh

Email: roshan0404singh@gmail.com

LinkedIn: [linkedin.com/in/roshansingh0404](https://www.linkedin.com/in/roshansingh0404)

Contact No: +91 8130368717

OBJECTIVE

To be able to excel at my work with the best of my efforts and to grow with the organization.

TECHNICAL SKILLS SET

Tools: Python, SQL, Scikit Learn

Databases: Oracle 11g/12c, SQL Server 2014, Redshift

Languages: Python, Pyspark, SQL, Unix, PowerShell

ETL Tool: AWS GLUE, Informatica Power Center 10.x/9.x, Talend (Enterprise Big data edition 5.4.2)

MDM Packages: Informatica MDM Multi Domain Edition 10.0. Informatica Data Director (IDD) 10.0 Informatica Data Quality (IDQ) 10.0

RDBMS: Oracle 12c/11g (SQL/PLSQL), Redshift

Utilities: Amazon S3, TOAD 10.1, SVN, TFS, JIRA

Libraries and Packages: SKLEARN, PREPROCESSING, Model Metrics, Train test split, search cv, Pandas, Numpy, MS Office (MS Excel, MS PowerPoint, MS Word), Visual Studio, Java Eclipse

Profile Summary

- An Aspiring Data Scientist Completed Post Graduate Diploma from IIIT Bangalore in Artificial Intelligence and Machine Learning and working in Accenture solution Pvt Limited.
- Having Total 7.5 years of experience in IT industry currently working as Data Engineer with hands on experience in AWS Cloud and Bigdata tool and technology with Spark, Python, Pyspark, AWS S3, Glue, Redshift and working experience in industry ETL tools like Informatica Power Center and Talend Big data with knowledge of Informatica MDM and Data warehousing project.
- Experience in Python and Experience working with Hadoop and Big Data processing frameworks Spark
- Have used AWS Glue to create ETL Jobs, Workflows by inculcating Python/Pyspark scripts for data cleaning and Transformations.
- Have used Spark with Python (Pyspark, Python Scripting) and data processing with AWS cloud.
- Have used AWS Glue Connection for S3 Data transfer and creating Crawler to connect to Redshift database and tables.
- Experience in doing Exploratory data Analysis using Pandas/Pyspark data frames.
- Have developed and used Regression Models, Classification Models and clustering models.
- Have worked on Confusion Matrix, Accuracy, Sensitivity, Specificity Tuning and Roc Curves.
- Experience in analysis, design and development of ETL using Informatica PowerCenter 10.x/9.x tools like Designer, Workflow Manager, Workflow Monitor, and Repository manager using databases like Oracle
- Experience in analysis, design and development of ETL using Talend tools for Big data Edition

WORK EXPERIENCE AND PROJECTS

1. Accenture Solutions Pvt Ltd.

Period: 22 Oct 2018 – Present

Designation: Application Development Team Lead

Location: Gurgaon

Project: Accenture Customer Insight CRM (ACI-CRM) (22 October 2018 - Present)

Environment and tool used: Spark,Python ,AWS GLUE, Amazon S3 Bucket, Amazon Sagemaker, AWS Redshift,Jira,GITLAB,Informatica PowerCenter 10.2,Talend open studio ETL tool (Enterprise edition Big data), SQL Server, Redshift, Microsoft Azure Blob, Unix

Client: Digital CRM

Domain: Retail

Project Description:

Accenture Customer Insight is an intuitive business intelligence solution that provides new insight into customer data. It is responsible for providing insights and doing analytics related development for it. Processing huge amounts of transaction, customer and marketing response data from Digital CRM Client and its Loyalty program providers, to build a massive database in the Cloud.

Enable enduring ETL load process to load loyalty, point of sale, product, store and customer data to suffice reporting needs.

As a Data Engineer, my roles/responsibilities are as follows:

- Analyze customer use case, available data from different sources to prepare for automated ongoing data ingestion to meet customer requirements Design technical solution with all data ingestion, transformation pipeline and provide meaningful insights for customer data
- Scheduling to review sign-off with customer Closely collaborating with other team members ,engineers, consultants and different offshore teams for delivering project
- Developed a generic CMA outbound framework which could be used for 7 different geography for campaign team using Pyspark/Python that helped project in till date realized savings of 27104\$ as against an estimated 31680\$ till end of FY
- Worked for a petroleum client, Understanding the pain points of the client and work on different use cases to increase the overall customer footprints which improved their overall revenue. Analyzed data to identify patterns and correlations among the various data points.
- Worked on building an AI Solution by implementing a clustering model understanding different method of payments generated revenues and different payment partners and patterns affecting sales for different geography and identified and reported patterns affected by promotional offers .
- Developed an archival approach and fine tuning for the feeds to save ~60% time it was historically taking
- Worked on Optimization of Jobs in by Optimizing the Queries and Physicalization.

2. Tata Consultancy Services

Period: 23 December, 2013 – 12 October, 2018

Designation: ETL Developer

Project 1: Abbvie Research and development (Eclinical CDISC MDR)

[31 Dec 2017-12 Oct 2018]

Project 2: (INTEGRATED INFORMATION PLATFORM)

[23 Dec 2013-31 Dec 2017]

Domain: LifeSciences

Environment and tool used: Informatica PowerCenter 10.x/9.x(Repository Manger, Designer, Workflow Monitor, Workflow Manager),Putty, CA Workload Automation Tool, Oracle SQL, Toad, Informatica MDM, IDD, IDQ ,Flat files, SQL, UNIXScripting, Soap UI Tool, Eclipse ,Putty, Autosys.

Client: Large US based pharmaceutical company Major Fortune 150.

Project Description:

Project is to create a centralized data warehouse for the data coming from different OLTP systems, XML files and Flat Files (through message queues). Informatica PowerCenter is used to Extract, Transform and Load the data into the data warehouse. The system is capable of handling both initialand incremental load of data as well as of maintaining history for the records. For each area of interest specific Data Marts were created (as a part of business requirement).

As an ETL Developer, my roles/responsibilities were as follows

- Worked as ETL developer for an Informatica ETL development team
- Extracted data from various heterogeneous sources like Oracle, SQL Server, Flat Files
- Knowledge of STAR /Snowflake Schemas and Top/Bottom Approach used in data ware housing
- Developed Slowly Changing Dimension Mappings for Type 2 SCD using Date Concept to maintain history

EDUCATIONAL QUALIFICATION

Year	Qualification	School/College/University	Grade/Percentage
2019-2020	PG Diploma In Machine Learning and Artificial Intelligence	IIIT Bangalore	3.41/4
2009-2013	B. Tech ECE	GOGTE INSTITUTE OF TECHNOLOGY, Belgaum	71.27%

Roshan Singh