**GOPAL CHANDRA SARANGI Contact No: +91-**8796860359

**Software Engineer Mail ID: *gopal2sarangi@gmail.com***



**SUMMARY:**

* **Having 4.8 years of IT experience.3 years as spark(pyspark) developer and 1.8 years as ETL datastage developer.**
* **Collaborates with cross-functional team in support of business case development and identifying modeling method (s) to provide business solutions. Determines the appropriate statistical and analytical methodologies to solve business problems within specific areas of expertise.**
* **Integrated Teradata with pyspark for BI platform and also implemented corporate business rules**
* **Participated in Business meetings to understand the business needs & requirements.**
* **Arrange and chair Data Workshops with SME’s and related stake holders for requirement data catalogue understanding.**
* **Guide the full lifecycle of a Hadoop solution, including requirements analysis, platform selection, technical architecture design, application design and development, testing, and deployment**
* **Performed Datamapping between source systems to Target systems, logical data modeling, created and used SQLqueries to filter data**
* **Enabled speedy reviews and first mover advantages by using Oozie to automate data loading into the Hadoop Distributed File System and PIG to pre-process the data.**
* **Used various techniques using Pyspark data structures to get the data in right format to be analyzed which is later used by other internal applications to calculate the thresholds.**
* **Maintaining conceptual, logical and physical data models along with corresponding metadata.**
* **Done data migration using pyspark from an RDBMS to a NoSQL database, and gives the whole picture for data deployed in various data systems.**
* **Developed triggers, stored procedures, functions and packages using concepts associated with the project using PL SQL**
* **Used Meta data tool for importing metadata from repository, new job categories and creating new data elements.**

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Technical Domain Knowledge** | BI/DWH Testing ,BIGDATA Testing |
| **Functional Domain Knowledge** | Telecom ,Pharma |
| **Technology & Tools** | Spark,Python,Hadoop,Hive,Impala,Data Stage 11.3,Informatica Power Center 9.X**,** Aginity Workbench, Oracle Developer, |
| **Database** | Netezza 7.0,Oracle 10g |
| **No SQL Database** | Hive ,Impala |
| **Languages** | Python,SQL |
| **Operating System** | UNIX/LINUX 7.0 , Windows 2K/XP |

**PROFESSIONAL EXPERIENCE:**

* **Worked as Software Engineer with** IBM India Pvt Ltd(from 27th Aug to Feb 17th 19)
* **Worked as Software Engineer with** Zs Associates (from 11th Mar to 26th July 19)
* **Working as Software Engineer with** United Health Group(from 5th Aug to Till Date)

**KEY PROJECTS HANDLE:**

**Project Title: Vodafone BI/DWH (MMFA)**

**Client:** Vodafone India Pvt. Ltd.

**Technologies:** DataStage 11.3, informatica 9.X,Netezza 7.0,Oracle 10g

**Database:** Netezza, Oracle 10g,

**Programming Languages:** UNIX, SQL.

**Description:** Vodafone Provides Telecom Services to millions of customers on a daily basis. Hence it becomes imperative to maintain a concise pool of information based on the customer’s calling (minutes of usage), revenue generation trends and type of services availed and its variation on the basis of demographics etc. in order to be able to come up with new strategies to further boost revenues and customer satisfaction.

Usage, 3G and Non Usage data is extracted from different source systems in the form of flat files and after the various transformations on the raw data as per the business rules and logic it is loaded in the different fact tables, from these fact tables data is extracted for the COGNOS Reports for the Business Users.

The warehouse for this data has been implemented in the Star Schema Model with a number of Dimension Tables containing business keys to be referenced by the FACT Tables. This project has been implemented in the form of Data marts with each data mart processing a particular subset of data.

Therefore MMFA (MSISDN MART FOR ANALYSIS) Project provides business with the critical reports which are an amalgamation of the existing Data Warehouse i.e. 3G, MTR, Recharges, Usage and Non Usage Data etc.

**Project Title: Filed Sales Incentive Program(FSIP)**

**Client: Astazeneca**

**Technologies:** Amazon Web Services

**Programming Languages:** Spark**,**Python,Sql

**Description:** Astrazeneca is a pharmaceutical company, the purpose of FSIP is to analyze sales made by the representatives and provide them the Incentives apart from the salary what they deserve. Here based on the sales and prescribers we set the Goals for the representatives for a quarter in the middle of the previous quarter of that quarter. Then we evaluate the performance of the representatives on weekly basis. And in the start of the next quarter we decide their Payout based on their performance that how much sale have been made because of a representatives promotion .

**Project Title: Clinical Data Service Management(CDSM)**

**Client: United Health Group**

**Programming Languages:** Spark,Python,Sql

**Description:** United Health Group analyze the data about the patients and decide the efficiency of the doctor. Which it receives from different submitters(Hospitals). The data contains the details about the patients ,the tests patient is doing in the hospital, which insurance patient has ,how many times he/she coming to the hospital for the same disease.

**Responsibilities**:-

* Involved in functional study of the application.
* Worked on data cleaning and reshaping, generated segmented subsets using Numpy and Pandas in Python
* Worked on loading data from source system to target using Pyspark.
* Extracted data from various source system eg:Oracle,Mysql,Json,CSV and loading them into Raw zone into HDFS. Further data is cleanse and scrubbed and loaded into refined zone into HDFS. Finally business logic is applied to the refined data and loaded into Enriched Zone in hive table.
* Wrote and optimized complex SQL queries involving multiple joins and advanced analytical functions to perform data extraction and merging from large volumes of historical data stored in Oracle 11g, validating the ETL processed data in target database
* Good understanding of Teradata SQL Assistant, Teradata Administrator and data loadExperience with Data Analytics, Data Reporting, Ad-hoc Reporting, Graphs, Scales, Pivot Tables and OLAP reporting.
* Identified the variables that significantly affect the target
* Continuously collected business requirements during the whole project life cycle.
* Conducted model optimization and comparison using stepwise function based on AIC value
* Applied various machine learning algorithms and statistical modeling like decision tree, logistic regression, Gradient Boosting Machine to build predictive model using scikit-learn package in Python
* Developed Python scripts to automate data sampling process. Ensured the data integrity by checking for completeness, duplication, accuracy, and consistency
* Generated data analysis reports using Matplotlib, Tableau, successfully delivered and presented the results for C-level decision makers
* Generated cost-benefit analysis to quantify the model implementation comparing with the former situation
* Worked on model selection based on confusion matrices, minimized the Type II error

**EDUCATIONAL QUALIFICATION:**

* **B.Tech with 6.9 % from Biju Patnaik University of Technology,Odisha in 2013.**
* 12th with 73% from RCST ,Odisha in 2009.
* 10th with 74% from KPU VIDYAPITHA,Odisha in 2007 .

**PERSONAL SYNOPSIS:**

* Father’s Name:-MR. PRASANTA KU. SARANGI
* Date of Birth:-15-JUN-1991.
* Current Address:- Hyderabad
* Permanent Address:- JAJPUR,Odisha:-755006.