**Satyasobhan Dasmohapatra**

**Mobile**: +44-07448507740| **Email**: satyadm@gmail.com

**Profile Highlights**

|  |  |
| --- | --- |
| **Languages:**  **Scala, Python,** SQL, PL/SQL, Shell Scripting**,Java**  **Tools**:  Informatica Power Center,  Informatica BDM,  Business Objects, PLSQL, Spotfire,QlickView  **Hadoop Technologies:**  HDFS, YARN, Hive, Sqoop, **Spark**, Flume, Ambari, Kafka ,Hbase  **Database Systems:**  Oracle, Teradata,  MySQl, **Cassandra(no-sql)**  **Methodologies:**  Agile Scrum Master  Cloud:  **AWS,Azure,Snowflake,Redshift**  Oracle, | * An established software professional with an experience that spans over 12 years in designing and developing Big data lake and data warehousing systems * Architect, Design and develop Big Data Solutions practice including setting up of Big data road map, building supporting infrastructure. * Expert in Big Data and Hadoop Development using Apache Spark in Python,Scala, Hive and Sqoop, Apache flume. * Expert in Functional programming using Scala and object oriented programming in python. * Core Java Development and Python development experience * Expert in Distributed processing framework such as Spark * Cloudera certified Hadoop and Spark Developer. * Stream Processing using Kafka and Spark Streaming * Experience of working in AWS and Azure technologies * Expert in Informatica BDM(ETL Tool), Teradata and Oracle (RDBMS). * Expert in Informatica Administration. * Excellent technical and analytical skills with clear understanding of design goals of ER modelling for OLTP and dimensional modelling for OLAP. * CI/CD integration knowledge using GIT lab,Github * Experience in all the phases of Data warehouse lifecycle involving Requirement Analysis, Design, Coding, Testing, and Deployment. * Expert in Data mart design, and data modelling. * Expert in analysing business processes and gathering core business requirements. * Delivered projects in Agile Methodologies. * Certified Agile Scrum Master * Worked In various geographies such as UK, India and Mexico * Adept at maintaining focus on achieving bottom-line results while formulating and implementing advanced technology and business solutions to meet a diversity of needs. |

**Career Graph**

* **TATA CONSULTANCY SERVICES, MUMBAI  Assistant Consultant (Oct 2008 till date)**

**Key Projects:**

**CHADS Data solutions**

**Client:** Department of Works and Pensions

**Duration:** Jan 2019 -ongoing

**Project Details**: CHADS aims to replace the existing data warehousing solution with much more advanced Big data lake enabling users to do real time data analysis. The data lake solution is built on cloudera using technologies such as spark scala python and kafka. The streaming data is processed from Siebel systems into Big data using kafka and processed using spark streaming. Informatica BDM is also used as a tool to process data into hdfs from various sources including relational ,flat file and streaming sources such as Kafka

**Responsibilities**:

* Playing lead data engineer role in developing solutions in Spark and scala and python .
* Developed ELT piple lines for both batch and stream processing .
* Developed/implemented ML algorithms using Spark Streaming(Structured streaming) and Kafka
* The solution was based on Microsoft Azure environment.
* Unit testing suite in Scala and python
* CI/CD integration using GIT lab.
* Developing complex algorithms and etl pipelines are written using spark /scala .

.

* Integrating streaming data from kafka topics into big data lake and process using spark and Python
* Making data usable for various end user needs through qliksense and other data analysis tools
* Prepped data sets to be fed into ML algorithms.
* Involved in designing the CI/CD pipeline using Git labs

**Client: Mastercard**

**Duration:** Jan 2017 till Dec 2018

**Project Details:** Develop the fraud detection application using Spark, Spark streaming and cassandra. Sparks ML libraries were used to develop models. History data was loaded to Cassandra using SPark . Sparks ML libraries were used to develop models. Using Spark streaming the real time streaming data was loaded from Kafka and Fraud detection application was developed.

**Responsibilities**:

* Developed Spark jobs to load Historical data from HDFS to Cassandra applying transformations.
* Developed spark streaming jobs to load the Kafka streaming data and the ML models to enable fraud detection
* Developed Industry standard frameworks using Spark and scala and Python.
* Loaded the data from RDBMS server to Hive using Spark framework.
* Developed ML algorithms using Python

**SFDC Data lake Integration.**

**Client:** Wells Fargo

**Duration: Feb -2016 to Jan-2017**

**Project Details:** To integrate and enhance Salesforce data mart for Wells Fargo using data lake. The project was developed in Spark using Scala and Hive. An entire complex data mart was re written in Spark and Hive to load data into data lake. The spark framework was leveraged to load more than 30 sources into Hadoop data lake. Complex queries were rewritten in Spark and scala code to do the ETL processing in Hadoop data lake.

**Responsibilities**:

* Leading a team of 6 people from offshore and developing solutions through Spark and scala code.
* Developed Industry standard frameworks using Spark and scala to load 35 plus sources into data lake.
* Loaded the data from RDBMS server to Hive using Spark framework.
* Used Spark framework to load data into data lake.
* Used Spark and scala code to rewrite complex ETL transformations
* Created managed and external tables in Hive and loaded data from HDFS.
* Explored and optimized the entire processing which is done in Spark and Hive.
* Optimized the Hive tables using optimization techniques such as partitions and bucketing to provide better performance with HiveQl queries.
* Automated data testing between Hive and Teradata using Spark and Scala.

**FRD Development:**

**CCAR Development.**

**Client:** General Electric – Capital Americas

**Duration:** July 2014 till Jan 2016.

**Project Details:** The project aims at integrating the Risk data for entire GE Capital, EMEA, ASIA, ANZ Data to EDW. It’s a great milestone project for GE.

**Responsibilities**:

* Analyzing business process and gathering core business requirements. Interacted with business analysts and end users.
* Leading a team of talented technical professionals through the on-going development and implementation of new BI processes and systems.
* Performing a strategic review of all existing data streams to fully understand the function's capabilities and areas for development.
* Designed ETL process to load data from source to target.
* Extensively used Teradata Utilities like Fast-Load, MultiLoad, and BTEQ & Fast-Export.
* Designed and developed common load framework for 100 plus businesses through concurrent processing in Informatica.
* Dynamic parameter files generation and changed Session parameters, mapping parameters, and variables at run time.
* Suggested feasible solution, to integrate the core and the new property systems
* Implement RI checks and other exception handling.
* Developed and modified the shell/Perl scripts as per the business requirements.
* Implemented parallelism in loads by partitioning workflows using Pipeline, Round-Robin, Hash, Key Range and Pass-through partitions.
* Implemented CDC logic through Informatica mapping and also Teradata bteq.
* Used various Transformations like Expression, Filter, Joiner and Lookups for better Data messaging, migrate clean and consistent Data.
* Testing and debugging of code.
* Troubleshooting of long running sessions and fixing the issues.
* Monitored workflows and collected performance data to maximize the session performance.
* Involved in writing SQL scripts, stored procedures and functions and debugging them.
* Developed shell scripts for Daily and weekly Loads and scheduled using Applications Manager.

**CEF Risk DataMart Development**

**Client:** General Electric – Capital Americas

**Duration:** Dec 2012 till June 2014

**Project Details:** The Project was aims at developing the RISK Data-Mart for Equipment Financial Services segment of GE Capital. It included end to end design of the Risk Data-Mart

**Responsibilities**:

* Analyzed business process and gathered core business requirements. Interacted with business analysts and end users.
* Designed and developed end-to-end ETL process from various source systems to Staging area, from staging to Data Marts.
* Implement Dimensional Modelling.
* Implemented various Data Transformations using Slowly Changing Dimensions.
* Implemented partitioning to load historical data.
* Designed Unix Main Script and List File mechanism to ensure restartability.
* Carried out changes into Architecture and Design of Oracle Schemas for both OLAP and OLTP system.
* Developed simple & complex mappings using Informatica to load Dimension & Fact tables as per STAR Schema techniques.
* Optimized various Mappings, Sessions.
* Debugged and Tested mappings.
* Handled time critical Development requirements.

**Certifications**

* Cloudera certified Spark and hadoop developer(CCA175)
* Received many client appreciation certificates for leading the offshore team in Informatica Upgrade and Hardware upgrade and many process optimization initiatives.
* Rated as Grade A as per TCS performance standards.
* Six Sigma Green belt Certified.

**Education**

* **Bachelor of Technology** (B Tech) in Electrical Engineering from Indira Gandhi Institute of technology, Sarang, Orissa India (2008)

**Personal Particulars**

* **Date of birth**: 12th July 1985
* **Languages Known:** English, Hindi, Oriya