|  |
| --- |
| BISWADEEP PATRA**Email: - biswadeeppatra2012@gmail.com**Contact No: - +91-8240784757/9040656168 Bigdata Developer  |
| “To acquire a challenging career with a solid company utilizing theOpportunity to offer proven and developing skills within the company” |

# Career summary: -

* Proficient in all aspects of Software Development Life Cycle (SDLC) and **agile** methodology.
* **4 years' experience** in development of **Hadoop** and **Spark applications using Scala and Python** with **PySpark**.
* Full stack developer in **Hadoop ecosystem tools** for data ingestion and processing – Using **Sqoop, Flume, Hive, Impala, HDFS,** **SQOOP, Oozie, Kudu**
* Working experience in **Apache Spark & Kafka ,AWS** Building Spark batch using **SparkSql** as well as **Spark Streaming** applications for data processing. Reading messages from Kafka, applying transformation and persistence.
* Familiar with **Hadoop environment with Json, PARQUET** and other file formats
* Expertise in **shell scripting** for connecting into DB and extracting report.

# Technical skill: -

* **Bigdata Stack -**

Hadoop Eco System- HDFS, SQOOP, Hive, Impala, KUDU, Oozie

Spark Frame Work –Spark Core, Spark Sql, Spark Streaming Using Scala/Python

 Distributions-Cloudera, Hortonworks, Messaging Tool- Kafka, flume

* Language: Scala, Python
* Database: PL/SQL, SQL Server, ORACLE, Netezza, MySQL, Impala, Hive, MariaDB
* Scripting language: Unix Shell Scripting, Python
* Operating System: Windows XP,10, 2007, Enterprise Linux
* Version control tool: SVN, Git
* Tools: Eclipse, Super PuTTY, Bitbucket, FileZilla, Jira, PyCharm
* Build tool: Maven, Jenkins
* Github Link=[https://github.com/BiswadeepPatra](https://www.google.com/url?q=https://github.com/BiswadeepPatra&sa=D&source=hangouts&ust=1551133893485000&usg=AFQjCNHkvgegbHkmjFBcvjE7HmLgwG-A4Q)

# eMPLOYMENT HISTORY: -

1. **DBS Bank (July 2019 – Till date)**

# Project EXPERIENCE: -

1. **Enterprise Performance Management System-Fund Transfer Pricing**

 **(July 2019 to present)**

* Role: Data Engineer
* Skillset: Hive, Impala, MariaDB,Spark Core, Spark Sql,Pyspark

#  RESPONSIBILITIES: -

* FTP individually measures how much each source of funding is contributing to overall profitability of firm, which DBS use it as means of outlining the areas of strength and weakness within the funding source based on the operational needs and function of loan and deposits.
* Data will be extracted from different sources like Hive tables, Maria Db Tables, Files using Spark programs and target data is written to Hive or Maria Db tables depending on the requirements, triggering the workflows using Tivoli Work Scheduler (TWS), shell scripts to trigger the spark applications and exposing the required reports to the users.
* Independently worked on spark upgradation process from Spark 1.6.2 to 2.3 and It reduced the time consumption of existing application by 30% in production.
* Currently working on integrating pyspark data frame, pandas data frame and koalas for efficient data processing.
1. **TATA Consultancy Services (July 2016 – July 2019)**

# Project EXPERIENCE: -

# 1. Name: SEM Replication

* Role: Data Engineer
* Skillset: Hive, Impala, Sqoop, KUDU, Scala, Oozie,Spark – Spark Core, Spark Sql, Spark Streaming

# RESPONSIBILITIES: -

* Requirement was to replicate sql server tables in Hadoop system to analyses in real time mostly near real time. So, a system was designed to fulfil that using spark streaming technology. Kafka was used as message broker.
* As soon as there was any DML operation happened on any of the configured table at RDBMS side, the CDC generated and triggered a message to a Kafka producer application using REST API.
* The producer then pushed the message to Kafka topic, and the consumer started to consume the messages and consumed messages were processed and put into Hadoop tables (Hive, Impala, Kudu)

# 2. Name: Impression reconciliation

* Role: Data Engineer and analyst
* Skillset: Hadoop - Hive, Impala, Sqoop, Scala, Oozie,Spark – Spark Core, Spark Sql

# rESPONSIBILITIES: -

* Developed a system in spark using Scala that processed application logs in batch mode and referenced to reconcile with the production data.
* I developed a flow that used to read the backend application logs and a spark job was written to process those application logs in batch mode and store all the impressions in a table. Before storing the data into Hadoop system, a lot of validations happened on those records i.e., IP validation, Country validations etc.

**3. Name: Sync up rdbms and impala**

* Role: Data Engineer and analyst
* Skillset: Hadoop - Hive, Impala, Sqoop, Shell Scripting

Developed around 100 Sqoop import and export jobs for extracting data from Oracle, SQL server, Netezza, MySQL into HDFS and populating in impala tables and vice versa.

Added the features of incremental load while Sqoop import so that only new record will be imported instead of full refresh.

# tRAININGS AND CERTIFICATIONS: -

* Successfully completed 6 months class room training for Hadoop ecosystem and Apache spark using Scala from Nimble Tech Institute, Kolkata
* Successfully completed Oracle 10g training for SQL and PLSQL

# aCADEMICS and aCHIEVEMENTS:

* Bachelor of Technology in Electrical and Electronics Engineering from GIET, BPUT University of Technology with CGPA of **8.51** in 2016.
* Intermediate, Science from CHSE, Orissa with an aggregate of **78.86%** in 2012.
* Matriculation from HSC, Orissa with an aggregate of **89.16%** in 2010.
* Award Received ‘Technical Excellence award’, `Special Initiative’, ‘Star of The Month award’ from TCS and Best Innovator in DBS Bank.

# PERSONAL DETAILS: -

* Date of birth: 28-July-1994
* Language: English, Hindi, Odia, Bengali.

 Address for correspondence: Plot No- 14-1-211/435, Parvath Nagar, Mothi Nagar Hyderabad 500018.

Date: - Biswadeep Patra

Place: - Hyderabad