**Sagnika Chakraborty**

**Spark, Scala,Hive,Sqoop,Hadoop and Sql, Python**

Surya Apartments, Silpa Park,Hanuman Nagar,

Kondapur, Hyderabad,Telengana - 500084

*sgchkty19@gmail.com*

*+91 8248341869*

|  |
| --- |
| PROFESSIONAL EXPERIENCE |

**Cognizant Technology Solutions(4 Years and 1 Months)**

**Data Analyst/Big Data Engineer**

* Experience in technologies like Hadoop, Hive, Sqoop, Scala, Apache Spark & Cassandra
* 1 Year of experience in Spark and Scala
* Involved in creating of HIVE partitioned tables and configuring different metasores.
* Good working experience of using SPARK SQL for faster processing of data.
* Worked in resolving data issues , implementing the fix and designing the data flows.
* Have work in the development of two banking applications and one data analytical application.
* Excellent technical abilities possess strong oral and written communications skills, analytical and interpersonal skills.
* Built 50+ analytics dashboards using Google dashboards
* Designed metrics to help modeling businesses
* Implementation of machine learning algorithms
* Provided analytical insights for both vendors and clients
* Built and productionized robust data warehouses
* Designed and implemented end to end pipeline for extraction and loading using multiple technologies(Python, Google SQL, Dremel SQL etc)

|  |
| --- |
| EDUCATION: |

* *M.A in Economics, December 2020*

*IGNOU*

* *Bachelor of Technology, June 2015, BPUT, Orrissa*

*Major in Electronics & Telecommunication Engineering*

*CGPA: 7.3/10*

* *Grade XII, March 2011, ICSE*

*Marks: 86%*

* *Grade X, March 2009, AISSE,*

*(Discipline: General)*

*Marks: 93%*

|  |
| --- |
| Projects |

## Skills : Data analysis, Data warehousing, Machine Learning, Google SQL, python, Pandas, Google Dashboard, Google sheets, Google docs, SSTable, Google slides

Skills : Data analysis, Data warehousing, Google SQL, python, Google Dashboard, Google sheets, Google docs, Google slide

|  |  |  |
| --- | --- | --- |
| Customer | | **Google Inc(March 2019-Present)** |
| Industry | | Imagery Data Analytics |
| Project Description | | Google LLC is an American multinational technology company that specializes in Internet-related services and products, which include online advertising technologies, search engine, cloud computing, software, and hardware. It is considered one of the Big Four technology companies, alongside Amazon, Apple and Facebook. |
| Role | | Data Analyst |
| Responsibilities/ Deliverables | * Develop short term and long term business strategies through reports, analysis etc * Generate sample data for different evals * Provide training and test datasets for machine learning * Structured the workflow and technical implementation in close collaboration with the Client team * Design data warehouse and perform data collection, scrubbing, transformation and ingestion to maintain a good data hygiene * Develop dashboards to show insights and find patterns * Work with subprojects like Herschel, Traffic lights, **IMRE** * Changed the workflow of few subprojects to improve the productivity and quality * Built dashboards to find patterns in operational activities in order to find the areas for improvisation * Made a dashboard to show peaks and drops in the operations and analyzed the triggers in details * Worked closely with the project management team to analyze and consolidate request requirements to improve the decision planning process   Contributed to build business dashboards to conduct trend analysis on metrics, especially relating to cost saving and reducing manual effort through automoderation | | |

|  |  |  |
| --- | --- | --- |
| Customer | | **WalMart DataAnalytics (June 2018-December 2018)** |
| Industry | | Telecommunications |
| Project Description | | Cicso in one of the famous Telecom cutomer based out of USA  As there is no automated process to restore the solutions for the issues raised. Hence, to improve customer satisfaction, Cisco customer teams wants the support team to operate on the raised issues on a Pro-Active model |
| Role | | Big Data Developer |
| Responsibilities/ Deliverables | * Imported all the streaming data from different regions using flume component of hadoop to HDFS. * Processed the Unstructured data using MapReduce for understandable format to store data into tabular formats. * Involved in creation of Hive External tables and partitioned tables to store the data. * Stored the all network related data into business function tables into Hive External Tables. * Involved in exporting the data from hive tables to MySQL tables using sqoop component for creating the reports. * End-to-end performance tuning of Hadoop clusters routines against very large data sets. | | |

|  |  |  |
| --- | --- | --- |
| Customer | | **XTerra Services(June 2017- May 2018)** |
| Industry | | Retail |
| Project Description | | Redesign Xterra application and Leverage ASUP 5.0 infrastructure in memory processing and streaming capability to speed up parsing of ASUPs. Faster ASUP processing and resultant timely case creation enables problem prevention and improved customer service. Xterra uses Spark, Scala for parsing, and stores intermediate data in NoSql Database Cassandra and notifies Xterra service through Kafka |
| Role | | Big Data Developer |
| Responsibilities/ Deliverables | * Collaborated with the Internal BA’s in understanding the requirement and architect a data flow system. * Developed complete end to end Big–data processing in eco system * Exploring with the Spark 2.1.x, improving the performance and optimization of the existing algorithms using Spark Context, SparkSQL, Data Frames. * Implemented Batch processing of data sources using Kafka 0.9 * Imported data from different sources to Spark RDD * Used Spark 2.1.x over YARN ( architecture based ) to perform analytics on data in Hive. * Loaded and extracted the data comprising CSV, XML and JSON files, read the data using Kafka Streaming and finally inserted into Cassandra 3.10 | | |

**Environment:**

|  |  |
| --- | --- |
| SPARK | 2.1.0 |
| SCALA | 2.11.8 |
| KAFKA | 0.9 |
| CASSANDRA | DSE 5.1 (Cassandra 3.10) |
|  |  |
|  |  |

|  |  |  |
| --- | --- | --- |
| Customer | | Close Brothers Premium Finance **(May 2016- April 2017)** |
| Industry | | Banking |
| Project Description | | CBPF is part of the Close Brothers group and has been helping people pay for their insurance premiums for over 35 years. CBPF provide finance through there broker at a competitive rate of interest so customer don’t have to pay for their insurance premium in one go. |
| Role | | Data Tester |
| Responsibilities/ Deliverables | * Requirement analysis * Test case design * Test case review * Test Case Execution * Defect Management * UAT support * ADPART Design And SoapUI | | |

# SKILLS

* **Data Analysis & Visualization :** Elementary Python, Scala, SQL, Google Dashboard, Google sheets, MS Excel
* **Data Warehousing, ETL and Bigdata :** Google Plex, SQL, Python, Hadoop, Hive, Sqoop, Spark and Scala
* **Scripting & Query Languages : Unix,** Python, SQL
* **Databases :** Hive, Dremel
* **Other tools** : Google Dashboard, Google sheets, Google docs, Google slides, MS Excel, Colossus
* **Functional skills** : Data Analysis(Descriptive and Predictive), EDA, Data warehousing, Machine Learning, ETL, Big Data, Backend Development, Database design, Machine Learning, Web scraping

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

Sagnika Chakraborty