**Amartya Kundu**

**Phone: (+91) 740 034 2373**

**Email:** [**amartyakundu93@gmail.com**](mailto:amartyakundu93@gmail.com)

***Profile***

A result-driven professional with 4.5+ years of experience. Quick learner with expertise in Big Data and modern evolving cloud technologies.

***Summary***

* Having total of 4 years and 9 months of total professional experience, of which 2.6 years working experience in Hadoop Ecosystem, design/developing applications and 2.1 years working in BI Reporting tools.
* Involved in designing and development of solutions for Big Data using Hadoop ecosystem technologies such as **HDFS, Hive, Sqoop, Map Reduce, Apache Spark, HBase and Cloud (AWS)**.
* Worked in NoSQL database like **HBase**.
* Having experience in processing Structured and Semi-Structured data like **Parquet, ORC ,Json, Avro, XML** files
* Basic knowledge of working on AWS Components like **RDS, S3, EC2** and **EMR**.
* Scripting knowledge on Scala, Python, Shell scripting.
* Proficient in all the stages of SDLC as well as STLC.
* Worked in Agile as well as waterfall methodologies.
* Worked in Financial Domain related Investment Banking and Wealth Management.
* Have good problem solving and analytical skills and love to innovate in order to perform better.
* A good communicator with strong interpersonal skills.
* Able to connect with stake holders and understand the business rules.
* A strong leader and team player, self-motivated to innovate and implement new ideas.

***Technical Skills***

Big Data Ecosystem: Spark 2.4.7, Hive, Hadoop 2.6.0, HDFS, Sqoop 1.4.6

Database: Hive, Oracle, HBase, Cassandra, MySQL

Distribution: Cloudera 5.13

Operating Systems: Microsoft Windows, Linux

Languages: Scala 2.11.12, Python, UNIX Shell Scripting

Cloud Platform: Basics AWS (RDS, EMR, S3, EC2)

Tools & IDE: Eclipse, Jenkins, Jira, PyCharm

***Work Experience***

Overall Experience: May 2016 –Present

Experience in Big Data: Jan 2018 -Present

Accenture Solutions Pvt. Ltd., Pune as Application Development Analyst

**Project Name: UBS**

**Designation:** Big Data Developer

**Technologies:**RDS, Spark, HBase, Hive, HDFS

**Description:** Client places the required columns in RDS on daily basis. Data is updated monthly on an Hbase table which has around 3200 columns. Daily, based on the columns available in RDS for the given day, the catalog is created dynamically and data is fetched from Hbase into a Dataframe. The data is then sorted based on the RDS columns and written to respective Hive tables which are created dynamically on a daily basis and are flushed out when the next day’s run starts.

**Responsibilities**:

* Involved in extracting Meta data from AWS RDS using spark.
* Performed dynamic schema creation with spark to read data from HBase.
* Worked extensively on performing necessary ETLs with respect to the business required.
* Involved in optimizing the spark jobs during read, transformations and writes.
* Worked on deploying spark jobs in AWS EMR.
* Loaded and transformed large sets of semi structured data like parquet.
* Code & peer review of assigned task. Unit testing and Bug fixing.

**Project Name: UBS  
Designation:** Big Data Developer

**Technologies:** RDS, Spark, Hive, HBase

**Description:** Incremental data is loaded from Amazon RDS(MySQL) on hourly basis. Scala Spark consumes data as REST API(JSON) at a particular interval. Data Comparison between Incremental data and REST API data takes place. Processed data is delivered in the HDFS on top of which a partitioned external hive table has been built. Data delivered to HBase from Hive through Hive handler.

**Responsibilities**:

* Involved in integrating spark with multiple sources like web services, AWS RDS and Hive.
* Handled incremental data loads from AWS RDS using Apache Spark.
* Was responsible for maintaining the cost based optimizations during transformations using spark.
* Involved in integrating the spark jobs with multiple target systems like HDFS, Hive and Hbase.
* Managed issues on migrating the developed jobs to the production.
* Involved in handling incremental data with Hbase storage handler hive table.
* Maintained necessary partitioning in hive table writes using Spark.
* Code & peer review of assigned task. Unit testing and Bug fixing.

**Project: UBS**

**Designation:** Big Data Developer

**Technologies:** Sqoop, HDFS, Hive, Spark, Cloudera

**Description:** I have worked on incremental data with avro schema evolutions in which the data gets consumed incrementally from sqoop and lands at staging area of hdfs , after crunching the data at staging area, we write the data to target external table which is partitioned and the business queries the data from that table.

**Responsibilities:**

* Involved in creating HDFS directories to store the data and Hive staging and final tables.
* Performed importing of data files from edge node to appropriate HDFS directories based in file names.
* Processing of data and appending the changed data from Hive Staging table to appropriate Hive Target table using shell script jobs.
* Code and peer review of assigned task.
* Unit testing and Bug fixes.

***Education and Certification:***

* B.Tech from Kalinga Institute Of Industrial Technology, Bhubaneswar
* Microsoft Certified: Azure Fundamentals

***Personal Details:***

Father’s Name: Anup Kumar Kundu

Date of birth: 04th Oct 1993

Gender: Male

Marital Status: Single

Address: Flat-C/6, Sundaram Complex, Gopalpur, Asansol-713304, WB