Siddhartha Sadhukhan

Mobile: (+91) 896-115-0106 Bengaluru, Karnataka
E-mail: siddhartha.sadhukhan.2014@gmail.com India, PIN - 560037

Professional Summary

- Self-driven Software professional, having around 4.5+ years of experience in Product Development.
- Currently working at Société Générale as a Specialist Software Engineer - Data.
- Having sound understanding of Big Data Technologies Like Hadoop ecosystem, HDFS, Yarn, Spark, Kafka, Zookeeper, NoSQL.
- Hands-on experience in Azure Cloud.
- Familiar with Windows, Linux, Git, Maven, CI/CD Process, Distributed Systems, Kubernetes, Agile and Scrum methodology.
- Business domain knowledge in Manufacturing, Technology and Financial Services.
- Efficiently and effectively managed critical deliverables at the Software Engineering level associated with communication and client-handling skills.
- Interested in Coding, Data Structure and Algorithm, Problem Solving, Data Model, R&D and Architectural Designing.

Work Experience

- **Trading Open Platform, Société Générale** **March 2021 – Present**
  - Working as a part of Core Services Team, design and build uniform Data Model and provides a Hive view of the golden copy of trade data.
  - Azure Kubernetes services are used to observe the changed data and trigger Oozie workflow accordingly for Spark jobs.
  - Key Responsibilities:
    - Development of new source to model mappings and transformations, defect fixing, documentation, POC.
    - Adapt new changes and defect fixing for Kubernetes services.
  - Achievements:
    - Developed a tool to eliminate Hive dependency, can be imported to a spark project to access trade data without knowing about the storage location, partitions and other implementation details of the upstream system.
    - Pairing cache is an initiative to onboard Market Risk Analyst team, which is dependent on Oracle Exadata.
    - Optimization in the process of transferring Exadata data to Azure Data Lake provided 80% saving in storage and significant improvement in timing.
- **Admin Panel Data Processing Pipeline, Mu Sigma** **March 2020 – February 2021**
  - Spark is used to process the telemetry data of a leading communication and collaboration tool stored in Azure Data Lake Storage Gen2, processed data is stored in Cosmos DB and pseudonymized IDs in Redis.
  - Azure Databricks is used to execute the Spark jobs and pipelines are defined using ADF.
  - Key Responsibilities:
    - Development of new matrices, documentation, defect fixing, changing on the existing matrices, POC.
    - Modification and creation of ADF pipeline to fit new execution flow.
  - Achievements:
    - Optimized the pipeline which has decreased the storage cost by 20% and solved the storage throttling issue.
    - Another optimization reduced the pipeline execution time by 5%.
    - Developed CI/CD pipeline along with unit testing framework.
- **Mosaic Data Transformation, Wipro** **December 2018 - February 2020**
  - Using Spark transformed the banking data stored in the Hive table to produce input for a BI system.
  - Key Responsibilities:
    - User story development, Defect fixing, POC, Deployment, Documentation as well as providing support.
  - Achievements:
    - Optimized the data pipeline which reduced the execution time by 15%.
    - Onboarded PEP Data which has significantly strengthen the regulatory reports.
- **Non-Conformity Management System, Wipro** **January 2018 - November 2018**
  - Web app to manage quality check process of an Aviation manufacturing company, accessible using REST API through JSON and XML format.
  - Liferay is used to develop the application along with Spring and Hibernate framework.
  - Key Responsibilities:
    - Development of user stories, Defect fixing, POC and providing DevOps solution.
  - Achievements:
    - Introduced a file link at storage level so that SAP and Liferay both can access the same data, reduction in 50% storage.

Languages and Technologies

Programming Languages: Java, Scala, SQL, Python, Shell-Script
Databases: Hive, Postgres, Azure Cosmos DB
Frameworks: Apache Hadoop, Apache Spark, Junit, Spark-Testing-base, Spring
Workflow Schedulers: Oozie, ADF

Education

- MCA, Academy of Technology, Maulana Abul Kalam Azad University of Technology **2016**
  8.82

Publications

- Sudipta Roy, **Siddhartha Sadhukhan**, Shayak Sadhu, Samir K Bandyopadhyay, "***A Novel Approach Towards Development Of Hybrid Image Steganography Using DNA Sequences***," Indian Journal of Science and Technology, Vol.8, September Issue, 2015
- Shayak Sadhu, Sudipta Roy, **Siddhartha Sadhukhan**, Samir K Bandyopadhyay "***Automated Segmentation of the Human Corpus Callosum Variability from T1 Weighted MRI Image***," IC3T, *Proc. Springer*, Advs in Intelligent Syst., Computing, Volume 379, Hyderabad, India, 2015. ISBN: 978-81-322-2516-4; Chapter DOI: 10.1007/978-81-322-2517-1_7

Personal Projects

- **PcRemote** **2017**
  - One step in the direction of Man-Machine interaction.
  - Microcontroller receives the IR data from the remote and sends those to the computer.
  - Java is used for receiving the codes and performs operations accordingly.
- **Home Automation System** **2016**
  - Complete solutions of electronics and software allow room automation, working towards IoT.
  - Can be operated by both IR remote and custom Android app operate via Bluetooth and Internet.
  - Ability to control electrical devices like light, setting up alarm and custom color to an RGB LED strip.
  - Display information like date, time and room temperature.
- **Virtual Mouse** **2015**
  - Java based standalone application that simulates the job of a mouse.
  - After processing the images taken by the webcam, the system detects the different objects and, according to objects' coordinates performs intended mouse operation.
- **Features Extraction from MRI** **2014**
  - Java based image processing application which processes MRI images.
  - Returns binary images of extracted White Matter, Gray Matter, CSF, Skin and Fat.
- **Steganography** **2013**
  - Java based standalone application towards data security, consisting of encryption and decryption modules.
  - Encryption module uses image file as cover media and encodes the data to create stego-media.
  - Decryption module decodes the original data from the stego-media.