CURRICULUM VITAE

Saikat Bhattacharjee
Contact No : (+91) 9032000565
Email: bhattacharjee.saikat3@hotmail.com
Address - New Town, P.O.- Inda,
City- Kharagpur, State - West Bengal,
India, Pin -700135

Professional Objective

Seeking a challenging position as a software developer or as a technical lead in a quality-oriented project, that will utilize my skills and experience and contribute to growth by learning and adapting existing practices and achieving excellence at work.

GENARAL SUMMARY

- Solution Architect/Data Engineer with **7.5 years** of success in conceptualizing technical solutions and system development predominantly in **Big Data platforms**.
- Hands on expertise in **Cloud Architecture**, **Big Data processes** and tools.
- Experienced in **Distributed** and **Co-located Agile - Scrum methodology**.
- Proficient in **Pyspark**, **Databricks**, **Python**, **Data Factory**, **cosmos Db**, **Azure SQL**, **Logic App**, **Storage Accounts**, **Synapse, Function App, Azure Devops.**
- Extensive experience as technical consultant in customer facing roles for **Insurance**, **Utilities** and **Telecom** domains at multiple geographies.
- Well conversant with **Software Development Life Cycle** (SDLC) and have carried out its various phases like **Requirements Analysis**, **preparing Systems Design development**, **testing** and **implementation**.
- Good Analytical skill.
- Quick learner with excellent Communication and interpersonal skill.

SKILLS SUMMARY

Technology	**:**	Pyspark, Python, Shell-script, Javascript, Kafka, Powershell.
Cloud Technology	**:**	Azure (Databricks, Azure Datafactory (ADF), Cosmos DB, Function App, Datalake gen2, Synapse, Azure Blob container, HDinsight-Ambari, Logic app, Azure Purview, Azure Devops)
Languages	**:**	Pyspark, Python, SQL
Database	**:**	cosmos Db, Azure Sql, Hive,Sql DataWarehouse, Mysql etc
Version Control	**:**	GIT, Clearcase
IDE	**:**	Databricks, Jupyter, Pycharm, Visual Studio, IntelliJ IDEA
Operating Systems	**:**	Windows, Solaris 10
Project Management tool	**:**	Jira, ServiceNow, Mhweb, confluence, SAP(ITSM)

Saikat Bhattacharjee email : bhattacharjee.saikat3@hotmail.com

EXPERIENCE SUMMARY

Client: Gjensidige (January 2021 – Till date)

Analytics Platform Digital Sweden

Overview:

Gjensidige runs several campaigns and advertisements for their product via several marketing channels such as Social Media Platforms Facebook and search Engines Google and Bing.

Currently all the data such as user impressions and clicks getting generated are categorized and stored into individual columns in BigQuery datawarehouse in GCP.

This project aims to bring all the tables from BigQuery in GCP to Azure and Create a dashboard in PowerBI to display an overview of impressions and clicks from different marketing channels.

Responsibilities:

- Worked as a Data Architect /Data Engineer in the project
- Requirement discussion with client.
- Contributed at the architecture design of the application.
- Built a dynamic pipeline in Datafactory which will fetch all the tables from BigQuery Datawarehouse schema.
- Once the table names are extracted Datafactory will loop through all the tables and load the tables in Azure Datalake(ADLS Gen2) in respective folder as per the table names in BigQuery. Folder and filenames are generated dynamically
- Once the data is available in Datalake inside the same loop Azure Databricks notebook will be called and table name, schema name will be passed.
- In Azure Databricks source and target file path are formulated dynamically from the parameters passed from Datafactory that is table name and schema name.
- Once the paths are formulated parquet files are read using predefined schema to avoid any corrupt records.
- Once data is loaded in databricks it is transformed and stored as delta files in ADLS gen2 and external tables are created in hive metastore.
- Once all the external tables are created the the tables are joined into a single delta table to provide an overview of impressions and clicks from Different marketing channels.
- The final delta table is loaded into PowerBI to create dashboards

Saikat Bhattacharjee email : bhattacharjee.saikat3@hotmail.com

Modern Data platform

Overview:

A transformation project for an Insurance client to develop Enterprise Data platform and migrate on-premises solution into Azure.

The project aims to bring customer sensitive and policy related information from IBM mainframe to Azure Datalake whenever a customer logs into Gjensidige portal using SSN and Bank ID to purchase an Insurance or to register a claim.

The data from mainframe will be streamed continuously using Kafka and consumed by Microservices/Databricks.

Databricks will then filter and transform the data and store it into Deltalake(as delta tables partitioned by Policy No., year and month of policy registration) in Datalake Gen2.

Data residing in Datalake will be consumed by Microservices deployed in AKS to feed data to front-end application.

External tables are created and hive is used to store and retrieve table metadata from Datalake in Azure account. The table metadata will let the end user know how to find, read, and process the data that you want to query.

Responsibilities:

- Worked as a Data Architect /Data Engineer in the project.
- Architecting Modern Data platform considering the different modules of the application.
- Working on several POC's to develop a streaming solution between IBM Mainframe and Azure, developing an Enterprise data platform to store current as well as historical data, converting transformation currently done in SAS in databricks using pyspark and spark-sql, developing a data warehouse and hive metastore.
- Providing feasibility to existing microservices to fetch data from Datalake rather than connecting to mainframe via proxy.
- Collaborating with multiple vendors like confluent cloud(Kafka), FDC(IBM mainframe), Databricks.
- Designing a robust solution to be re-used by teams of Gjensidige Sweden, Gjensidige Norway and Gjensidige Denmark

Saikat Bhattacharjee email : bhattacharjee.saikat3@hotmail.com

Client: Australia Energy Market Operator (14th November, 2017 – 15th January, 2021)

5 Min Settlement

A transformation project for Australian Regulatory body to change the 30 mins spot price settlement to 5 Mins spot price settlement. This is purely a Data Engineering project hosted in Azure Cloud. This will affect the whole Australian Energy market and its world's first project for 5 Mins settlement.

Responsibilities:

- Requirement discussion with client.
- Contributed at the architecture design of the application.
- Developed Enterprise Data Platform on Microsoft Azure that aggregate data from all relevant internal and external applications into a consistent structure and context.
- We receive data from varied sources such as oracle (structured data) and cosmos db (unstructured data). Oracle data are extracted as CSV files and cosmos db data are extracted in JSON format. Structured Streaming as well batch processes has been used in the project to write data into Enterprise data platform as parquet files of format "delta".
- Streaming data from Azure cosmos DB using Azure Databricks and storing data into Azure Datalake as delta tables.
- Hive and sql Datawarehouse tables are maintained for delta tables.
- Reading data from Azure Blob in batches and storing it into Azure datalake as delta tables which in turn is used to generate reports using PowerBI.
- Use of delta lake time travel feature to roll back the data if reconciliation of data between source and target fails.
- Reconciliation of Data is performed by verifying SHA256 fingerprint of Source Data with target Data
- Support for end users to run Adhoc queries to verify data integrity using Databricks Jobs
- Packaging of pyspark utility functions into wheel or egg file.
- Code developed in Pyspark and spark-sql.
- Developed backend services written in Node.js using Typescript, containerized using Docker and deployed to a Kubernetes platform hosted on Microsoft Azure cloud in an Agile delivery environment.
- Deployment in SIT and UAT environment and issue resolution.
- Performed root cause analysis and estimation of development items.
- Done POC in Azure Data Factory, Power BI and kafka for upcoming modules in project.

Saikat Bhattacharjee email : bhattacharjee.saikat3@hotmail.com

Client: Ericsson (3rd June, 2014 – 10th November, 2017)

Operations Support System Radio & Core (Platform Security)

OSS-RC provides certificates, authentication and authorization to all applications and nodes that use OSS-RC platform. Cryptography is critical part of this PKI solution where EJBCA is used for creating and managing the certificates. It additionally provides Single Log On service to applications and establishes IPSEC tunnel between nodes in OSS-RC and security gateway.

Responsibilities:

- Developed Security Products - (Digital Signature, Authentication, Authorization)
- Developed public key Infrastructure using EJBCA to secure communications between all OSS-RC applications and Network components (Nodes).
- Developed in-house IDM to authenticate/authorize any actions performed in OSS-RC
- Provided single sign-on capability for all OSS-RC users from application server.
- OSS-RC supports LDAP, NTP, NETCONF, CORBA, IPSEC protocols
- Developed python scripts to deploy Public Key Infrastructure
- Worked on Geo-Redundancy feature in OSS-RC.
- Was engaged in regression testing of the product.
- Worked on CI/CD using Jenkins.
- Automated a few test cases.
- Worked on Customer Support Tickets.

Educational Qualifications

- **B.Tech** from **West Bengal University of Technology** in 2013.

Declaration

I hereby do solemnly affirm that the details furnished above are true to the best of my knowledge and I shall be held responsible for any sort of discrepancies found.

Date: 05th October, 2021 **(Saikat Bhattacharjee)**