Biswajit Pan
AWS|GCP| Big Data Engineer
https://www.linkedin.com/in/biswajit-pan-258102195/

Mobile: +1 (201)746-5956 **E-Mail:** biswajitpan48@gmail.com

Professional Summary

- Over 12+ years of development experience applications using AWS , GCP Cloud, Java/J2EE, Hadoop and Big Data technologies in Banking, Insurance domains.
- Proficient in Java, J2EE, Spring, Kafka, Spark, Spark Streaming, Scala, Python, Airflow, Hive, Snowflake, Sqoop, REST, PL/SQL, Jenkins, jQuery, GitHub, Snowflake, Shell, Hibernate, Oracle, Ant, Maven, JIRA, Log4J, GIT, HTML, JavaScript, JSON, JMS, XML, JNDI, XSLT.
- Software development experience using Big Data technologies to efficiently solve Big Data processing requirement.
- Application Development using AWS Suite like S3, Athena, Glue, EMR, EC2, RDS, Lambda's, SNS, SQS, CloudFormation, IAM, CloudWatch, Dynamo DB, VPC etc.
- Good understanding in Hadoop eco system components HDFS, Map Reduce, Yarn, Sqoop, Spark, Spark SQL, Spark Streaming, and Hive for scalability, distributed computing, and high-performance computing.
- Experience in using Hive/SparkSql Query Language for data Analytics.
- Active contributor in Apache Spark/Hadoop user forums and proficient in developing custom frameworks.
- Implemented Spark using Scala and utilizing Spark Core, Spark SQL API for faster processing of data.
- Used Spark-SQL to Load various kind of data and create Schema RDD and loaded it into Hive Tables and handled Structured data using Spark SQL.
- Developing Spark programs using Scala and python API's to analysis of data
- Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
- Collected and aggregated large amounts of data from RDBMS and stored the data into HDFS for analysis.
- Generated the datasets and loaded to HADOOP Ecosystem.
- Involved in creating Hive tables, loading with data and writing hive queries that will run internally in MapReduce way.
- Worked with various HDFS file formats like Avro, parquet File and various compression formats like Snappy, gzip.
- Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Scala and Python.
- Worked in all phases of SDLC, DevOps and Agile models - Played Scrum Lead and shadow Scrum Master roles.
- Expertise in technically leading Migrations to Open Source Technologies
- Ability in technically leading Java/J2EE, Spark, Bigdata developers and teams.
- **8+ years of Onsite experience** working directly with clients and handling onsite & offshore teams effectively.
- Capable in initiating POCs, Project estimations, Hardware provisioning, release and configuration management
- Highly motivated, results oriented and leadership skills with great team ethics.
- Good in mentoring novice programmers and evaluating team performance.
- Have the ability to take corrective and preventive actions in delivering desired quality.
- Very strong analytical and reverse engineering skills part of technology migrations.
- Experienced in working with geographically distributed development teams.
- Hands on Experience in developing web applications by using Hibernate and Spring Frameworks with Apache Tomcat.
- Expertise in developing data models using Hibernate Object relational mapping (ORM) framework and used Hibernate Annotations to reduce time at the configuration level.
- Worked with design and development of robust and highly scalable Web Service platforms and API's based on REST and SOAP for enterprise applications using AXIS, WSDL, UDDI, XML, JAX-WS, JAX-RS, JAXB, JAX-RPC technologies and Jersey frameworks.

Technical Skills

Big Data Ecosystems	: Spark streaming, Spark, Sqoop, Hive, Flume, Hadoop, MapReduce, HDFS, Pig, Kafka
Web Services	: RESTful Web Services & SOAP
Languages	: Java/j2ee, Scala, Python, JavaScript, Servlet, PL/SQL, Unix Shell
Web Technologies	: HTML, JSP, JSF, CSS, JavaScript, Python, AJAX
Server-Side Frameworks	: Spring, Hibernate, Struts 2
IDEs	: IntelliJ IDEA, Eclipse, RAD
Build Tools	: Maven, Ant
Web Servers /App Servers	: Apache Tomcat 6.0/7.0, IBM WebSphere 6.0/7.0, JBoss 4.3
Static Code Analysis	: Find Bug/Check Style/PMD
Configuration Tool	: GIT-Bit bucket, SVN, CVS, SOAPUI
Database	: Snowflakes, MongoDB, DynamoDB, Cassandra, HDFS, Oracle 11g,
Defects Triage	: Quality Center 9.2.
Cloud Environment	: AWS,GCP, Cloudera

Academic And Certificate Information

Education	**University**
Degree	: Bachelor of Technology (Computer Science) from WBUT University, India
Certifications	: **- AWS Certified Solutions Architect.** **- Big data expert** **- Sun Certified Java Programmer.**

Employment History

- I am currently working in **Comcast from 04/2022.**
- I am previously worked with **Fidelity Investments from 05/2020 – 05/2022.**
- I previously worked with CapitalOne **as Data Engineer from 2018 - 05/2020.**
- I previously worked with **Liberty Mutual, AAA Insurance**.

Company Projects Undertaken

Project: Comast - Nexus
Details: We migrated our services from the MELD to the AWS but some of the clients are still using it. For that purpose, we are having the workflow to load the data from the MELD to the AWS.Processing data from multiple sources like HIVE, MELD, AND TERADATA. The data is stored in datamart database. From DB we will pull the data to s3 data.in bucket.

We have some lambdas to pull rec data from Mysql (application) to S3 in buckets.
In MELD we have different databases like for Nexus called Nexus database, for datamart we have nexus_datamart, for nexus application we have nexus, for nexus ADM like nexus_ADM. These are all the databases residing in the MELD.

Once Data transfer from meld to AWS S3 and give trigger to ETL workflow in AWS Step Functions. Triggers are based on S3 buckets after the data transfer into AWS

Technology: Databricks, Spark, pyspark, python, Lambda,Athena,Glue,S3,Steps etc
Client: Comast
Tools: AWS,DataBricks
Role: Data Engineer
Team Size: 12
Duration: April, 2022 – till

Project**:** Data Strategy
Details**:** Data Strategy is a framework to build data integration. The tool is a replacement of SAS application, fetching data from various data source, save it to AWS S3 for audit purpose and writes it to Hive database, using Athena (AWS) & AWS Glue. To setup the outline, user can create the configuration file and pass on this configuration file deliver the same during the deployment.

Technology: Spark, Scala, AWS
Client: **Fidelity (West Lake,TX)**
Tools: EMR, AWS
Role: Lead Data Engineer and Developer
Team Size: 12
Duration: May, 2020 – May,2022

Accountabilities:

- Involved in design and implementation of building the data ingestion framework to read data from any RDMS , Row file and load the data into AWS S3 and Athena by Hive and Glue.
- Design and implementation of Batch Data Integration framework for migrating data from on-prem to cloud (AWS).
- Developed a "Data Quality" check framework to validate data integrity across different stages of ETL pipelines using Scala, Spark, and AWS SDK.
- Writing in Hive and using AWS Athena and Glue.
- Historical Data evaluation by Tableau user report.

Project**:** Data Express
Details**:** DataExpress is a framework to build Streaming applications. The tool helps fetching data from Kafka cluster, save it to AWS S3 for audit purpose and later validates and writes it to Snowflake database. To setup the configuration, one can create this configuration file through a self-service portal and pass on this configuration file provide the same during the deployment.

Technology: Kafka, Spark Streaming, Scala, python, AWS, Java, Rest webservice
Client: **CapitalOne (Plano,TX)**

Tools: EMR, AWS
Role: Developer and Team Lead
Team Size: 10
Duration: Jan, 2018 – May, 2020

Accountabilities:

- Developed/Enhanced Kafka framework to pull the live stream data from various source systems.
- Involved in design and implementation of building a Streaming data ingestion framework to read data from various Kafka streams and load the data into AWS S3 and Snowflake data warehouse using Scala, Kafka, Spark, AWS SDK, Livy, and Snowflake.
- Design Exception handling for this frame work
- Design to create auto Incident Request and Amazon SNS
- Design Kafka Failover
- Validation of the Schema for incoming data.
- Design and develop highly scalable APIs and components in the java ecosystem, explore integrating with UIs as necessary
- Design, code and test RESTful APIs & reusable components

Title	Customer first(CF)
Technology:	Spark, Scala, Hive, Sqoop, AWS, Java, Rest webservice
Client	**Liberty Mutual (Seattle,WA)**
Team Size	12
Role	Design, Analysis, Detail Design & Developer
Period	December 2016 – December 2017

Project Description

Customer first(CF) applications comprised of 3 different application which includes Customer first property(CFP), Customer first Auto (CFA) & CF Personal Liability Protection(CP).
While individual modules can be loaded independently of each other, every module shares a set of core code which must present to successfully run CF from a development slice.
These applications collectively referred as CF and these applications are used by Liberty Mutual sales staff to get insurance quotes, policy and Policy Endorsement etc.

All the applications are in production and the business is rolling out new functions to the application state wise.
Team is working in agile methodology and the new functionalities will be described in Epics and assigned to the team.

Role and Responsibilities

- Involved in architecture discussions and requirement analysis.
- Involved in discussions with business line.
- Designed and implemented solutions to the technical problem of migrating the on-premises infrastructure to OpenShift cloud Platform with abiding the guidelines of Enterprise Architecture.
- Migrated 10 Java web applications from on-premises infrastructure to OpenShift cloud.
- Developed the CF related service.
- Manual Code Review

Title	AAA Insurance Home Products
Technology:	Java, J2EE, Spring, Hibernate, Soap, Rest
Client	**AAA** (Phoenix,AZ)
Team Size	55
Role	Design, Analysis, Detail Design & Developer
Period	January 2011 – December 2016

Project Description

This project is a product enhancement and customization based on client specifications and requirements for rolling out a system to support foot print and partner club states. The core base service functionality is provided by Exigen product factory runtime and it is based on policy core common requirements. Creation of New custom products on the top of the policy core system by extending baseline services which includes 3 steps process (1) Product Creation with the help of independent loosely coupled components and respective service configurations, (2) Policy Creation.

Role and Responsibilities

- Interacting with the Product Managers and stakeholders at AAA NCNU to better understand their business needs and provide good solutions to them on the same.
- Understanding the Business requirements and configure, customize the product as per the requirements document.
- Preparing custom component lifecycle diagrams, service design documents for each custom component, UML class diagrams and sequence diagrams as per as the client requirements and specifications
- Development of components/artifacts that implements Component interface contract, versioning of the components, and creation of the new business services by extending core policy core services provided by Exigen product factory
- Responsible for packaging of the components into libraries, adding/removing attributes, changing attribute labels, configuration of component constraints, component deployments

Personal Details

Present City : Plano, Texas

Languages Known : English.

Visa : H1B (I-140 approved)