|  |  |
| --- | --- |
| Somabrata Pani | |
| Profile for “Job Title & Company”   * 9+ Years of experienced Big Data Architect with a demonstrated history of designing, building & deploying production grade scalable & fault-tolerant distributed systems solutions around open source & enterprise big data technologies in on-premise & on Cloud (GCP & AWS), serverless Proficient in designing solutions across various business domains including Banking, Insurance, Energy Utilities & Cyber Security * Hands on experience in full stack software engineering, including core backend systems, Webservices & APIs, Build/Release systems, Containers & Microservices and implementing Data Pipelines, Data Lake Warehouse, Data Migration, Data Flow Automation * Experienced in building batch/real-time data ingestion pipelines using big data/streaming   technologies like Apache NiFi, Kafka, Spark, Hadoop   * Strong Data Architecture & Technical delivery/consulting experience working with various global clients across APAC, EMEA, LATAM & NA regions * Experienced in delivering projects on Agile. Knowledgeable in applications of scrum, waterfall, and agile methodologies. Used JIRA workflows extensively for CI/CD   Functional Skillset   * Data Platform Engineering, Solution Design and Implementation, Data Pipeline Automation, Integration & and Data Services * Cloud and On-Prem implementation of scalable solutions involving Application Design & Analysis, Code Development and Production Deployment * Involved in various project life cycle starting from RFP, requirement gathering from stakeholders, architecture design, technical implementation, till delivery * Implementation of Real-Time Cyber Analytics Platform | |
| Personal details Nationality India Location Singapore Languages English Stanton House Contact Name Laura Taylor | Education & Qualifications BTech Electronics & Computer Engineering 2010: Biju Patnaik University of Technology, IND |

Employment History

Accenture

Data Engineering Manager

Nov 2017 – To Date

Project: Real-Time Cyber Analytics Platform

## Description

* Use-cases to process Netflow-IPFIX data to detect/alert cyber- security threats.
* Build, Tuning & Optimization of real time ingestion of high traffic Netflow-ipfix data from Telcos’ PCAP datagram using Apache NiFi Cluster & Kafka on HDP(hot zone): NiFi->Kafka->Spark Streaming->HDFS->Elastic Search->Kibana->Alerts
* Design and setup of CDP(cold zone) for the compute and storage, which involved real-time data wrangling and enrichments from various source lookups
* Integration of two different clusters (Ingestion and Compute) using Apache NiFi
* Fine tuned the processors and parameters in NiFi, Kafka and spark streaming for optimal performance to achieve the required SLA
* Design, Implementation and deployment of spark-scala jobs for Cyber-Analytics use cases to detect attacks and redirect the alerts to KIBANA dashboards
* Technology: Apache NiFi, Kafka, Spark, Hadoop, Oozie, Hive, HDFS, Scala, Shell, Python, Cloudera Data Platform, Hortonworks Data Platform
* Monitoring/Servicing: Cloudera Manager, Ambari, Kibana, HUE

Project: Commercial Bank – ADA Platform – Global Ingestion Framework – Data Factory

## Description

* Global Ingestion framework is a Java and Spark based framework used to ingest data in parquet format to Alluxio, from various sources. This supports transactional and Master(SCD2) data ingestion. This supports formats like – Delimited, Positional, Mainframe, JSON, Excel file ingestion.
* Led implementation of core ingestion modules using Spark and Java which ingests data & stores in parquet format to Alluxio, from various sources
* Designed to capture CDC of full/incremental refresh type and transactional and Master(SCD2) data ingestion of multiple file formats & data sources
* Developed Jenkins pipelines to automate CI/CD for the ingestion framework
* Facilitated business systems onboarding & job scheduling and monitoring to use the GIFW to ingest their system data into ADA platform
* Software: IntelliJ, Airflow, MySQL workbench
* Technology: Spark, Java, Groovy, Shell, Python, Hadoop, Alluxio, Beeline, AIRFLOW

KOGENTIX Singapore Pte (Acquired by Accenture)

Big Data Technical Lead

Nov 2017 – Oct 2018

Project: General Insurance – Data Lake Warehouse (HK / SG)

## Description

* Creation of Single view of customer (Customer-360), along with Ingestion pipeline for various source like Google Big Query, Google Analytics, IBM AS400, ODS.
* Building a scalable and fault tolerant Analytics Platform using various big data technologies to ingest data for analytical processing(Customer-360)
* Developed the Ingestion framework and workflows which includes SFTP, extraction of CSV files from Excel
* Hands on building the ingestion pipeline to ingest data on from various legacy systems like Oracle and IBM AS400
* Designed and integrated Java Client APIs to get the data from Big Query & Google Analytics
* Implemented client APIs to get metrics & funnel dataset from Google Cloud Storage
* Implemented **UDFs** for Hive to get decrypted data from the encrypted columns from various sources to perform joins over datasets
* Software: Eclipse, MobaxTerm
* Technology: Java, Shell, **Hadoop**, Hive, **Impala**, HUE, Google Analytics, Google Cloud Storage, Google BigQuery, Oozie

Persistent Systems

Senior Technology Lead

May 2017 – Oct 2017

Project: Deutsche Bank – WM Pitchbook

* Developed core & custom integration modules for content ingestion framework
* Worked on code optimization & provide the technical solutions to the team
* Developed Jenkins pipelines to automate CI/CD for the content ingestion framework for production deployment on-premise
* Developed custom plugin to support full text search using Apache SOLR library
* Software: Eclipse IDE, Oracle DB
* Technology: Hadoop, Hive, NiFi, Spark, Java, Restful Webservices, Apache SOLR

Fair Isaac Corporation

Solution Consultant II

Jan 2014 – May 2017

Project: iScore – hOriginations (Aug 2015 – May 2017)

## Description

* hOriginations is a complex bureau integrated scoring system developed for iScore, to be used by various banks and financial institutions to on-board customers for various product lines (Credit Cards, Personal Loans, etc.).
* End-to-end development of new modules. Production implementation of solutions. Managed the team and client expectation for a timely delivery
* Software: Oracle DB, Eclipse IDE
* Technology: Java, SOAP Webservices, Activiti

Project: Optum – Custom Creditor Portal (Jan 2015 – Jul 2015)

## Description

* Custom Creditor Portal is customised version of pre-existing Debt Manager Product solution for Optum. The portal provide additional feature like files, images, reports, statements/invoices upload and download for portal users.
* Involved in complete implementation of functional requirements, Solution customisations with enhanced capabilities
* Software: Eclipse Luna IDE
* Technology: Java, Spring MVC, SOAP Webservices, Restful Webservices

Project: Cred System – New Origination System (Mar 2014 – Dec 2014)

## Description

* Origination System is complete solution for Credsystem credit card and instalment payments product i.e. Credline 2.0.
* Involved in Analysis of requirements, Responsible for end-to-end implementation. Solution customisations and complete implementation including integration with blaze advisor RMA implementation
* Software: FASt 8 (OutSystems), Eclipse Indigo IDE

Project: AK Bank – Multiple Case Marker (Feb 2014 – Mar 2014)

## Description

* Custom Customer Notes to Mark Multiple Cases for same Customer.
* Involved in analysing the Change Requests, providing Functional Specifications & Implementation Specifications. End to end implementation involving coding, testing and deployment in production

Infosys

Senior Systems Engineer

Dec 2010 – Jan 2014

Project: FOE (Apr 2013 – Jan 2014)

## Description

* FOE is the project involving development and maintenance of Integrated Supply & Trading portal for a Major Oil Exchange based in UK. There are applications facilitating the dealing and pricing of commodities.
* Software: Eclipse Galileo IDE, Weblogic 9.1 server console, Toad 9.5, IBM Rational ClearCase.
* Technology: J2EE Struts 1.2 frameworks, Hibernate, Oracle 10g, Weblogic 9.1

Project: Data Ceter Modernization (Jul 2012 – Mar 2013)

## Description

* Data Center Modernization is the project involving in-house application modernization and migration for a Major Gas and Electric Utility based in USA. There are set of applications being upgraded to the versions to satisfy the Infrastructure as a Service (IAAS) standard.
* Software: Eclipse Indigo IDE, Weblogic 11g server console.
* Technology: J2EE Struts 2.0 frameworks, Oracle 11g, Weblogic 11g

Project: C2 Maintenance (Aug 2011 – Jun 2012)

## Description

* The scope of the project was to provide enhancements and maintenance for the Customer Information System application for a Leading Utility in Northeast USA, i.e. Customer Central (C2 application).
* Software: Eclipse Ganymede IDE, IBM Lotus Note, Advanced Query Tool, Extra Personal Client, MochaSoft Client.
* Technology: J2EE Struts 1.2 framework, IBM DB2 (Mainframe).
* Training: Infosys Foundation Program (Dec 2010 – May 2011)
* Training Description: The foundation program carried out in 3 phases: Generic, Intermediate & Stream. Scored 4.93 on the scale of 5.0

Systems Skills

* Big Data Stack
* CDP, CDF, Hadoop, HDFS, MapReduce, Hive, Impala, Spark, Kafka, NiFi, Sqoop, HUE, Ambari
* Google Cloud Platform Stack
* Cloud IAM, Cloud Dataflow, Cloud Dataproc, BigQuery, Cloud BigTable, Cloud Storage, Cloud Pub/Sub, Anthos, Istio, GKE, Cloud Shell
* AWS Stack
* EC2, EMR, S3, IAM, AWS CLI
* Relational & NoSQL Database
* DB2, Oracle, Microsoft SQL Server, MySQL, HBase, MongoDB
* Programming & Scripting Languages
* Java/J2EE, Unix, Shell Scripting, SQL, PLSQL, Groovy, Python, Scala
* Workflow Scheduler
* Apache Oozie, Apache Airflow
* Development Tools & IDE
* Eclipse/IntelliJ, MySQL Workbench, TalenD Studio, OutSystems Service Studio, Mendix Develoer Tool
* Versioning Control & DevOps Toolset
* JIRA, GIT, Confluence, Jenkins, Gitlab CI, JFrog Artifactory, Nexus, BitBucket
* Container Orchestration
* Docker, Kubernetes