|  |  |
| --- | --- |
| **Prosenjit Das** |  |

https://www.linkedin.com/in/prosenjitdataanalyst

https://github.com/prosenjitj

+1-669-264-0843 / das.prosenjit.eie@gmail.com

Address: Sunnyvale, California- 94086, United State

Role: **Lead Data Engineer** *(Expertise in Architecture Design and Development auto-scaled and low latency ETL and Data Platform, Complex BI Aggregated Dashboard, Metrics Automation, Audit)*

**Experience**

8+ years Data Engineering

5+years Backend Development

Total: 14+ years’ experience in Software Development

**Technologies**

**Spark- Scala** (Streaming / ETL), **Kafka**, **Python 3** (ML API), Java, Akka(Load Distribution)

Visualization: Tableau, Splunk, SalesForce, Grafana, SugarCRM

Data Warehouse / Data Model: Oracle, MySQL, Elastic-Search

Big-Data DB: Hive, Hadoop/Parquet, Spark SQL, BigQuery

Expertise: ETL, SQL Optimization / Performance Tuning, Business Intelligence, Google Cloud

AWS: Lambda, S3, Redshift, EMR, DynamoDB, Cloud Watch/Trail

Handle Coding / Infrastructure: Github, Docker, Kubernates, Lambda, Intellij, STS

* Designed End to end robust-distributed **ETL Architecture / Data platform** and Aggregated Data Dashboard to help Business Intelligence Reports.
* Understand the problems, and tie them back to data engineering solutions. Analytical mindset to understand business needs, and come up with engineering solutions that scale to answers broader problems.
* Built robust and scalable big data pipeline/ ETL - Data **Collection**, **Filter, Transform, Merge, Analysis**. Data Ingestion / Migration from Legacy Data Sources (structured/ unstructured) as well as Multiple Platforms. (like Data Stream, Kafka, NoSQL Databases, S3)
* Designed Data Lake / Staging Area (by **Hive-Parquet**) – store huge dataset in raw or aggregated format and re-used same record for different analytical reports.
* Tune **Kafka** and **Spark** configuration to control high throughput-low latency and data loss over data streaming to bring down response time close to real time. Also scheduled Batch processing for audit data recalculation or overwhelming data.
* **Python** helps prepare data for Machine Learning Platform APIs with additional statistic information.
* Improved Performance by: Cached in In-memory, Disk Serialized, Elastic-Search, Redis, Checkpoint; Automation on Splunk / Kubernate API logs / hooks; application **error handling;**
* Convert Dataset to CRM components in Cloud APIs like Salesforce / SugarCRM and apply business logic integrate with multiple descendent components.
* Hands on Migrating architecture from On-premises to AWS Cloud. Transform and Process periodic data from S3 to **DynamoDB/ Redshift** by Lambda function.
* Audit after analyzing Service logs/ **Metrics** (like memory / CPU Usage / Crash) in Grafana, Tableau, Kibana, find all out-performed data , which might cause high impact in calculation.
* Built Data Model on Spark SQL/HSQL/Oracle and wrote **Complex SQL queries** (Grouping, joining, Window function) to tune performance and deep analysis on Database or Dataset.
* Experience providing technical leadership and mentor other engineers for the best practices on the data engineering space – Review Code, Brainstorming.
* Automation on Up-scaling / Downscaling by Kubernate / Docker, Crash, Slow N/W, Overflow, out of memory, Data Skewness, variant datasets Joining Optimization.

**Work Experience**

1. **Business Intelligence Data Platform –**

*Apple (contract through Infosys) (2014 – till date)*

Built a smartest Data Platform to track transaction over multiple platforms and represent through CRM platforms. Migrate Legacy Data Source from old CRM/Application/ Cloud. Base Architecture **Kafka-Scala-Spark-Hadoop-Oracle** to support big-datasets.

* Represent Complex Business Logic in Dashboard Format and Automation on notifying any up/down falls and handle the situation. Optimized architecture to handle complex Calculation and improve latency. Keep on Refactoring to response in near real time.
* Redesigned some features in Graph Frame framework to LPA and related components
* Hands experience on AWS Data Platform Environment
* Enhanced API service by Chat-bot to help non-technical users, human commands are converted to SQL by simple NLP.

1. **Google Cloud e2e IOT Status Monitoring**

*Infosys(India) Pvt Ltd. (2013 – 2014) - https://www.infosys.com/*

* Configured Google Cloud IOT Platform and Publish data from different sensor devices to cloud by Pub/Sub Messaging Interface.
* Subscribe the data to build report and Visualize. Also monitor through service log and BigQuery. Upload ML Model to find Precision and Recall.

Technologies: Python, Google Cloud

1. **Recommendation System by analyzing Content**

*Hewlett-Packard (Software Lab) (2011 – 2013) (http://www.hpl.hp.com/)*

* Worked with NLP researcher on extract user comments from different sources and understand intent of the comment. It helps to build recommendation platform for Advertising Agencies.
* Crowd Cloud: Built platform to analyze extracted Live Footage and content, which was converted in small task by NLP and computer vision and distributed by AmazonCloud **Mechanical Turk**

Technologies**:** Java, AWS, Python– NLP, K-means

1. **Streamed Game Data over millions of users mobile**

*July Systems (http://julysystems.com/, Acquired by Cisco) (till 2011)*

* Parsed live streamed game data and visualize on high performance mobile UI. Handled unstructured data on fly, filter, replace bad data and publish in near reality.
* Scalable server deployment using NodeJs for handle over flow for live match / no match.
* Technologies**:** Java, Spring, Rest API, Android, IOS

**Courses / Certificates:**

* Graph Search, Shortest Paths, and Data Structures from Stanford University
* AWS Training on Application Deployment and Monitor (Lambda, EC2, EBS, Cloud Watch/Trail)
* AWS Training on Authentication Integration Service and Automation (Auto-scaling Group,)
* Machine Learning from Stanford University
* MongoDB for NodeJS from MongoDB University

**Education:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Graduation** | **Year** | **University** | **Grade** |
| B.Tech | 2006 | West Bengal University Of Technology (**W.B.U.T**), India | 78.3% |

**General Information:**

**Passport:** Valid up-to 2025

**Visa Status:** H1-B (extended till March, 2022)

**I140**: Approved on October, 2019

**Date of Birth:** 29th May, 1983

**Born in:** Kolkata, India