Basil Sajid **Shaikh**

⚲ **San Diego, CA**   ✉ **shaikh.basil786@gmail.com**   ☎ **+18587890553**   **URL's** Tableau Profile - LinkedIn - GitHub

---

EDUCATION

Master of Computer Science– *San Diego State University - June 2023*   **GPA: 3.92/4**
Bachelor of Computer Engineering (2017) - *University of Pune*   **GPA: 3.87/4**

SKILLS

Languages:	Python, Java, SQL, IronPython, JavaScript, TypeScript, HTML, CSS,
Tools and Frameworks:	Oracle, Postgres, Spark, Hadoop, Kafka, Spotfire (Certified), Tableau (Certified)
Cloud Technologies:	AWS Athena, Lambda, Glue, Redshift, Hadoop EMR, RDS, AWS S3, Lambda

PROFESSIONAL EXPERIENCE

AMAZON - (Amazon.com Services LLC)   **Seattle, Washington**

Software Development Intern – Detection System Engineering   *May 2022- August 2022*

- Optimized the performance of a web-based root cause analysis tool by improving its data scalability by 140% (from 1 GB to at least 14 GB) and user scalability by 500%.
- Completely re-implemented the architecture, migrating from a monolithic to a cloud microservices based architecture.
- Configured AWS services and infrastructure according to the new architecture.
- Developed a native AWS(NAWS) CDK pipeline from scratch with guard rails for continuous development and integration.
- Redesigned and enhanced the UI/UX by implementing user-input validation and error handling.

TIBCO SOFTWARE   **Pune, India**

Senior Associate Consultant - Analytics and Data Science   *October 2017 – August 2021*

Performed tasks related to Software and Data Engineering, Statistical Modelling, Mining, Predictive Analytics, Business Intelligence

Projects Delivered at TIBCO:

1. DX Monitoring Dashboards (2021)   **Role:** Data Scientist   **Tech Stack:** Spotfire, Python, PostgreSQL, Prometheus

- Developed more than 10 real- time dashboards for monitoring different services running on production servers.
- Built ETL pipelines sourcing meta data for services and formulated KPI's to assess the performance of services.
- Forecasted service performances using time-series models such as SARIMA, Holt-Winters and LSTM.
- The dashboards were used 24/7 in performing load management and fault tolerance for the servers.

2. Nielsen - Data Science (2020-2021)   **Role:** Data Scientist   **Tech Stack:** PySpark, Sklearn, Spotfire, SQL

- Performed Market Basket Analysis and Customer segmentation for a leading supermarket chain.
- Designed models based on FP Growth and Apriori to find frequently bought items and performed RFM Analysis for customer segmentation.
- Developed data pipelines, performed data validation, cleaning and improved the performance of existing scripts by 80%.
- The outputs shared valuable inputs to the Sales & Marketing teams and resulted in an increase in profit by 7%.

3. Electricity Authority of Thailand (2020-2021)   **Role:** Data Scientist   **Tech Stack:** Spotfire, Python, Tibco DV

- Analyzed electricity usage and detected abnormal behaviors across different substations and districts in Thailand.
- Developed data pipelines connecting Hive, Oracle DB, Mongo, Spark and transformed it into unified format for analysis.
- Designed QoS performance KPI's like SAIFI/SAIDI and modelled them using SARIMA/Holt-Winters for time-series analysis.
- Processed unstructured logs and developed dashboard at daily, monthly levels based upon the different message codes.
- Analysis detected underperforming substations and improved the Quality of Service by reducing electricity interruptions.

PERSONAL AND ACADEMICAL PROJECTS

1. An end-to-end Data Engineering/Science solution for Cryptocurrency Data   Python, Spark, Kafka, Sklearn, AWS, SQL

- Developed a directory poller using watchdog to monitor for new files put on a data lake.
- Implemented Kafka topics to write these new files name to the topic and consumers to pass them to spark to process the data.
- Performed ETL in Spark and loaded the data in data warehouse and then into data mart using PL/SQL Functions.
- Developed Tableau dashboards on this data and predicted Bitcoin prices using RNN/ARIMA

2. COVID-19 Dashboard (2020)   Python, Oracle, Spotfire

- Developed a real-time Covid-19 dashboard to visualize the pandemics status across the world to track cases and deaths.
- Wrote scripts to consume John-Hopkins's data API, Calculated R(O) and implemented forecasting models using LSTM.
- Showcased the dashboard to the TIBCO management team who then decided to officially develop the dashboard further.

3. HUMAN- Hadoop Used Medical Analytics(2017)   Hadoop, Hive, Sqoop, Python MySQL

- Developed an application to store Electronic Health Records, medical reports, doctor, and disease information etc. for access as well as medical analysis.
- Designed multi-node Hadoop based data warehouse (OLAP) to store historical data and RDBMS based OLTP for recent data.
- Created pipelines to automate data flow between OLTP & OLAP using Sqoop and performed visual/predictive analytics on this data by implementing KPI's to understand relations between patients, doctors, diseases etc.