YASEEN BANU

Phone: +91 8978907726|| **E-Mail:** ybanu123@gmail.com
LinkedIn: www.linkedin.com/in/yaseen-banu-shaik-31111a126

PROFILE SUMMARY

A computer science and Engineering graduate with a keen aptitude for learning and developing innovative things and interest in Programming, Big data, and Data engineering with Strong grasp on Python, Pyspark, and Linux. Looking to take up challenges and high-performance role in the ever-expanding field of big data and implement the expertise and experience gained in this field to develop complex projects with high quality and efficiency.

Location Preference: Bangalore

ORGANIZATIONAL EXPERIENCE

Since Dec'21:   **Deloitte Touché Tohmatsu Limited USI, Bangalore**

Growth Path /Deputation:
May'23 – Current: **Consultant**
Dec'21 –May'23:   *Analyst*

Technologies: Pyspark/Spark, Python, SQL, AWS (Glue, Lambda, S3, EventBridge, SNS), AZURE (Databricks, Delta Lake), GCP (Bigquery, Dataproc, Cloud Storage, Airflow, DAGS)

Jan'19-Dec'21:   **Cognizant Technology Solutions, Bangalore**

Growth Path /Deputation:
August'20 – Dec'21:   *Programmer Analyst*
July'19 – August'20: *Programmer Analyst Trainee*
Jan'19 - June'19:   *Intern*

Technologies: Pyspark/Spark, Python, SQL, UNIX, Hadoop, shell scripting, Pandas, Java, SQL, Kafka framework.

EDUCATION

2019:   B.Tech. (Computer Science) from Sree Vidyanikethan engineering college – 86.04%
2015:   Class 12th from Sri Gayatri Junior college – 94.7%
2013:   Class 10th from Sree Shanthinikethan high school – 9.5

IT SKILLS

- Python 2.x/3.x, SPARK/Pyspark, SQL, AWS, AZURE, GCP, Hive, Shell scripting, NumPy, Pandas, Shell Scripting.

PERSONAL DETAILS

Date of Birth: 22/05/1998
Languages known: English, Hindi, Urdu, Telugu, and Tamil

AWARDS & RECOGNITIONS:

- **Rising Star Award, July'21**: Cognizant Technology Services

CERTIFICATIONS:

- Certified in **GCP Professional Data Engineer** 2023.
- Certified in **Applied Data Science I: Scientific Computing & Python** (with honors) Issued by WorldQuant University
- Certified in **AWS Academy Graduate - AWS Academy Cloud Architecting**
- Certified in an online course - **Apache Kafka series** by Stephane Maarek through Udemy
- Certified in a Big data online course - **Learn Big Data**: The Hadoop Ecosystem Masterclass by Edward Viaene

EXTRA CIRCULARS

- **Ideathon Achievement –** Secured 3rd place in a Deloitte USI team-wide Ideathon competition for developing a proximity marketing solution.

- **Technology Entrepreneurship Program –Offered by Indian School of Business (May 2016-March 2017)**
  I had been a part of TEP. It is a semi-virtual programme, tailored specifically to motivate and generate interest in entrepreneurship among 2nd and 3rd year engineering students.
  Learning outcomes:
  - Improved leadership skills
  - Effective communications skills
  - Decision making and innovative skills.

- **Salesforce trial head Blazer**
  I am Salesforce Trail Blazer with 60 Badges and 53,375 points and achieved the super badges in the following courses:
  - Process Automation Specialist
  - Apex Specialist
  - Security Specialist

Refer to annexure for projects handled

Deloitte Touché Tohmatsu Limited USI, Bangalore

Project 1:   **RSAC Data Engineering**
Client: Reliance Steel and Aluminium
Period: Dec'21 – July'22
Role: Data Engineer
Technology: Pyspark, Python, SQL, Snowflake, AWS (Glue, Lambda, Step functions, Event Bridge, SNS)
Responsibilities:
- Designed and developed a data pipeline in AWS environment using services like Glue, Lambda, SNS, S3 connecting ERP system data with Snowflake for improved process efficiency and faster report generation in Pyspark/Python.
- Worked on writing the Pyspark code which involves data cleaning, transformations as per the business requirements, egressing.
- Worked on writing the complex SQL queries in snowflake for constructing the tables/views based on business requirements from multiple tables also for preparing the ML training dataset from the existing EDW and Non-EDW tables.
- Worked on creating the Validation scripts in Pyspark/Python to automate the manual rework.

Project 2:   **PBM – Centralized Claims Tagging**
Client: CVS Health
Period: July'22 – May'23
Role: Data Engineer
Technology: Pyspark, Python, Spark SQL, AZURE (Databricks, Delta Lake), GCP (Dataproc, Bigquery, Storage, Airflow)
Responsibilities:
- Designed Data Ingestion and tagging Framework using Pyspark in Databricks.
- Worked with Business on finalizing the Functional requirement, Data Architecture.
- Created Logical framework for data ingestion and tagging logic for claims based on client contract and latest rules using PySpark from Databricks tables and executable driver notebooks.
- Orchestrated the Databricks Workflows and Jobs for the Ingestion, Tagging Framework.
- Worked on Building the CCT Common driver Framework which eliminates the necessary of other modules having unique functionality
- Supported and mentored junior team members by providing guidance on coding techniques and assisting them in selecting the most efficient approaches.
- Worked on Adhoc requirements – Data masking for lower environments, Silver Ingestion framework

Project 3:   **GMS AETNA**
Client: CVS HEALTH
Period: May'23 - JUNE'23
Role: Data Engineer
Technology: Python, SQL, GCP (Bigquery)
Responsibilities:
- Worked on writing the SQL QA test cases and Validations
- Provided understandable requirements documentation to support junior team members in their work.
- Collaborated with clients to gain an in-depth understanding of project requirements and facilitate successful project delivery.

Project 4:   **SCM**
Client: CVS HEALTH
Period: June'23 - Current
Role: Data Engineer
Technology: Pyspark, Python, Spark SQL, Snowflake, Airflow, Azure (Databricks, Delta Lake), Checkmarx, SonarQube
Responsibilities:

- Designed and executed a migration strategy for Hive tables to Snowflake, optimizing data transfer processes and ensuring a seamless transition.
- Successfully set up Snowflake connections and configurations, guaranteeing secure and efficient data transfer while also performing performance tuning and optimization.
- Collaborated closely with clients to understand project requirements and provided clear documentation, facilitating successful project delivery.
- Involved in proactive supervision of junior team members, offering valuable assistance in resolving code-related issues, implementing fixes, and ensuring their clear understanding of project requirements

Cognizant Technology Solutions, Bangalore

Project 1:   **CRM Finance and sales**
Client: Applied Materials Pvt Ltd., ITPL, Bangalore
Period: September'19 - Aug'20
Role: Developer, Big Data
Technology: Pyspark, Python, SQL, Hadoop, Hive, Impala, Shell Scripting, Unix
Responsibilities:
- Writing code in hive/Pyspark consisting of multiple joins, unions satisfying the business logic.
- Performing data validation on populated data.
- Writing python codes to automate certain task to increase overall efficiency.
- Migrating tables from HDFS to Kudu using Pyspark and python.
- Building the logic for the processing tables which comes from the Applications like SAP.
- Developing and scheduling redwood chains
- Creating, validating, and testing the tables that are created in DEV and QA environment.

Project 2:   **TiVo In-SAP & BSM Decoupling**
Client: TiVo tech Pvt Ltd.
Period: Oct'20 – Dec'21
Role: Developer
Technology: MySQL, Unix, Java, Mule, Kafka framework, Shell Scripts
Responsibilities:
- Worked on developing the Java application to update for batch status - Creating a mule scheduler, Processor Implementation, Manufacturing database updating and Unit testing
- Did analysis on the IT tools application (BSM UI, FE UI, MIDAS, DMI) and captured all the functionalities, procedure calls and database tables that are being used for all these applications.
- Worked on updating the legacy Logistics application to converge the MEP and ITP flows under the MEP flow.
- Did analysis on Shell scripts to capture all the associated table and column mappings.
- Presented a kafka topic design demo as per the Client requirement.
- Did a gap analysis using sequence diagrams for the Tivo existing architecture
- Worked on writing the shell scripts for populating the data from different sources to the sql tables.

Project 3:   **Image Recognition App (Intern)**
Trained upon: Python, UNIX, C, Java, MYSQL, Java script, UNIX.
Period: May'19-June'19
Responsibilities:
- Image Recognition App Using, Vision API and Machine Learning This project utilizes the concepts of, Vision API, Machine Learning and REST API.
- This project is developed in Java and enables the users to detect the pictures and labels present in the image submitted by the user through Google Vision API.
- Machine Learning is used to train the data model by our own dataset rather than the dataset used by Google.
- Developed a UI which will upload the sample of files from the front end sends it for processing then receive the results from the back end