Md Nuruddin Pathan

(+60)183995292 | github.io | LinkedIn | Email

Research Interests

Data Visualization, Data Mining, Cloud Computing, Big Data, Machine Learning, Natural Language Processing.

Education

Winter 2015 – Winter 2019 **BSc in Computer Science** | *International Islamic University Malaysia (IIUM)*
Major GPA 3.92/4.00
Overall 3.64/4.00

Experience

May 2020 - Present **Freelancing**

- Developed a **serverless scalable encoding** solution using Azure Batch **parallel computing** that takes 4k-8k resolution raw footage, breaks them into chunks, and separately processes them in Docker with FFMpeg AV1 encoder in 100 different **containers**. There is an algorithm to determine the maximum number of parallel execution. After the processing completes, it merges back the transcoded files in a compressed file format for a more excellent web-streaming experience.
- Build full-fledged **Data Mining** and automation solution for Twitter Followers/Tweets and Brandwatch Rules/Tags creation. Automated Azure Data Factory (ADF) pipelines using Azure automation account and **Powershell runbook** for Twitter, which invokes Python script to retrieve doctor's followers and tweets. For Brandwatch, it creates rules and tags using the API. Getting hundreds of thousands of followers data from Twitter was a challenge as Twitter imposes a strict rate-limit on followers/list endpoints with 200 pages, 30 requests per 15 min timeframe. The solution scrapes doctor's data with specialities on various diseases such as atopic dermatitis, multiple sclerosis, severe asthma, etc. The result is a Power BI dashboard to be sold as a product to pharmaceutical industries.
- Build a custom Hybrid solution to **extract keyphrases** using the unsupervised model in Python and help them by cutting the cost of around $300 per day from using Azure Cognitive Service.
- Build robust ADF pipelines to handle API rate-limit for different API types with varying levels of requirements and difficulties for various clients.
- Helped start-up companies by provisioning services in Azure and building frameworks in ADF to streamline their tasks.
- Consulted a many clients to solve challenges with developing solution in Azure Cloud Environment. E.g., helped organizations get data from **RESTful API** to Blob storage, then parsing from JSON and storing them to Azure SQL DB. For some individuals, converted scripts to run in Azure Functions (Python or C#).

April 2019 – Present **Schlumberger, SSO Data and Analytics** | *Kuala Lumpur, Malaysia*
Data Engineer

- **HR Confidential Data Project** – Initially, data was coming in encrypted format from Azure Blob Storage. To decrypt and process the data, I developed a few **C#** .Net application to process it step by step. It downloads the data from the blob, decrypts it, converts it to CSV format, finally uploads it back to the blob to process further with the ADF pipeline, and the blob is end-to-end secured by VNET(Azure Virtual Network). The decrypted data was later pushed to Azure SQL DB using an ETL pipeline using the ODBC Linked Service connection string. Another layer of protection is imposed on the Azure SQL DB using Azure Key Vault with the Always Encrypted feature for encryption/decryption of data at the row level.

- **Azure Batch service** – Worked on Azure Batch, which calls external scripts from ADF's custom activity with serverless architecture for parallel file processing and high-performance computing (HPC) workloads at runtime.
- **Data Automation –** Depending on the requirement from the Product Owner, either using SSIS or Alteryx workflows**.** Collected the data from the source, developed the ETL flow afterwards validated the data, did the UAT, and deployed to the Production environment. I also worked on **Azure Automation Account** to automatically trigger scripts on **Hybrid Worker VM's** for background bulk task processings.
- **Data Migration –** I have migrated data from the on-premises SQL Server data center to **Azure Data Warehouse (Azure Synapse).** Build the ELT pipeline Framework using Azure Data Factory (ADF) leveraged PolyBase feature for faster performance of bulk load, which is currently being used by the company.
- **Azure WebJob and Functions –** Build API to trigger Alteryx Workflows from Azure Data Factor using WebHook Activity and C# .Net script. 40-50 developers are using the API in the D&A department.
- **Automated** Alteryx workflow migration from QA to PROD environment, which is getting triggered from Azure Function App with CI/CD implemented in Azure DevOps.

June 2018 – January 2019 **Schlumberger, Share Service Organization** | *Kuala Lumpur, Malaysia*
Data Analyst Intern, D&A

Contributed and actively worked on the design architecture and development of Extract, Load and Transformation (ELT) solution to migrate 2TB of SSO D&A on-premise SQL Server data to **Azure Data Warehouse** (ADW) by leveraging ADF for business intelligence through Power BI dashboards. I Created several **SSIS** packages to automate ETL processes. I mostly have written in C# codes with other tools to perform the designated task for the SSIS packages. Also, build a text extraction solution using Python **NLP**, Regular expression and Alteryx for the product owners.

Selected Projects

2018 Student Scoring System using Blockchain Consensus Algorithm with Associate Prof Dr. Amelia Ritahani Ismail
Implemented a blockchain system to improve the in-class marking system by allowing real-time updates on markings and consensus for proof-of-work, enabling secure updates and visibility of work for valid participants in the database. The autonomous access to the system eliminates any third-party involvement apart from the students and teacher and secures student's scores from being altered. The proposed Blockchain-based system brings Decentralization, immutability and transparency.

2017 Missing Value Imputation using Supervised Learning Algorithm with Associate Prof Dr. Amelia Ritahani Ismail
Compared to the effectiveness of Linear Regression, kNN & support vector Machine to predict missing values in a variable-sized dataset. The imputation from kNN gives a near to accurate prediction than the other two. The results are compared in terms of MAE, RMSE, RSE, and Wilcoxon test. Datasets were collected from the Kaggle website, which was of around 4k rows. The coding part was done in both R and Microsoft Azure ML. For the Visualization, we used the Microsoft PowerBI tool.

2016 Movie Recommendation Engine with Associate Prof Dr. Amelia Ritahani Ismail
Developed and evaluated a Used Based Collaborative filtering model (UBCF) system for recommending movies, implemented with R language. The collaborative filtering approach is more potent than what an item-based recommender can provide. Users might not be looking for direct substitutes for a movie they had just viewed or previously watched.

Technical Skills

Languages C#, C/C++, Python, Java, JavaScript, R
Tools & Frameworks NumPy, Seaborn, Pandas, Tableau, PowerBI, SSIS, Alteryx

Deployment	Git, Docker
OS	Linux(Ubuntu), Windows
Cloud Computing	Microsoft Azure

Extra-curricular Activities

2015-2016 **Ummatic Festival** | *Islamic International University, Malaysia*
Winning Country
It is a one-week cultural performance festival. The various country performs its cultural dance, drama, melodrama, and singing to represent their country's culture.

2015-2016 **IIUM Bangladesh Community**
Volunteer, Organizing Committee

Others

Citizenship	Bangladeshi