Md Nuruddin Pathan

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

Research Interests

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

Education

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

Publication

2017   Tajwar, M. M., Pathan, M. N., Hussaini, L., and Abubakar, A. (2017). CPU Scheduling with a Round Robin Algorithm Based on an Effective Time Slice. *Journal of Information Processing System, Vol 13, (2017) No.4.*

Experience

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 ELT solution to migrate 2TB of SSO D&A on-premise SQL Server data to **Azure Data Warehouse** by leveraging ADF. I Created several **SSIS** packages to automate ETL processes. The SSIS packages, I mostly have written in **C#** with other tools to perform the designated task. Also, build a text extraction solution using **NLP**, Regex and Alteryx for the product owners.

April 2019 – Present   **Schlumberger, SSO D&A** | *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 blob, decrypts it, then converts it to CSV format, finally uploads back to blob to process further with ADF pipeline, and the blob was 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 afterward, 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 VM's** for background task processing.

- **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, 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.

May 2020 - Present   **Fiverr 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 compressed file-format for a more excellent web-streaming experience.
- Build two full-fledged **Data Mining** solution for Twitter Followers/Tweets and Brandwatch Mentions/Tags, respectively. Automated ADF solution using Azure automation account and Powershell runbook, for Twitter, 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 scrape doctors' data with specific specialities 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 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.
- I consulted a few clients to solve challenges with developing solution in Azure Cloud Environment. For example – Helped them getting data from RESTful API to Blob storage, then parsing JSON, and storing to Azure SQL DB. Converted script to be able to run in Azure Functions (Python or C#).

Teaching Assistant, Department of Computer Science, IIUM

Fall 2017 – Summer 2018   **CSC 1100 | Elements of Programming**
Conducted two weekly tutorial classes for 2 hours each, demonstrated the implementation of basic coding constructs in C++. Each class consisted of 30-40 students. Each week took one test consisting of 10 marks; in total, six tests were conducted.

Selected Projects

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.

2016   **Food Ordering System** with Associate Prof Zahidah Binti Zulkifli
A Java application to order foods based on the menu set inside the program. The database is built on MySQL to store and retrieve the information. A simple GUI is used to interact with the user. The system can be used for any normal boutique to order any menu.

Awards & Certification

2019   **ITIL**
Foundation Certificate in IT Service Management

2015-2018   **Dean's List** | *Department of Computer Science, IIUM*
Seven Semesters

2017   **IIUM Code Jam** | *International Islamic University, Malaysia*
Champion – Team Tyros

It is an onsite programming competition held every year in IIUM. Around 10-12 teams participate from ICT and Engineering departments. The contest was 5 hours long consisting of 8 problems. The contest follows the same ACM ICPC rules.

2017   **Unicode Programming Competition** | *Monash University & School of UOW in INIT*
Participant

More than 50 teams from various universities participated. The contest had 3 rounds, 3 hours each, with increasing difficulty. This contest had been a different experience because like other contest this contest also had 9 questions but a shorter timespan to think for each problem rather than the usual 5-hour contest.

2016   **ACM ICPC Malaysia** | *Al Khwarizmi National Programming Competition*
Honorable Mention (18th ranking out of 53 teams) – (Team Muharram1438)

A yearly competition between all the universities in Malaysia. Where the best competitive programming teams participate. We managed to solve 4 problems out of 9 problems.

Technical Skills

Languages	C#, C/C++, Python, Java, JavaScript, R
Frameworks & Libraries	NumPy, Seaborn, Pandas, 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

Referees

Dr. Amelia Ritahani Ismail
Associate Professor of Computer Science
Kulliyyah of ICT, IIUM
amelia@iium.edu.my

Dr. Raini Binti Hassan
Associate Professor of Computer Science
Kulliyyah of ICT, IIUM
naa@iium.edu.my

Dijo Davies
Global SSO Data & Analytics Technical Manager
Schlumberger
DDavies@slb.com