# Cloud Computing

## Cloud Computing Project

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Project Details | | | | | |
|  | | | | | |
| Student Name: | Aakash Sukre | | | | |
| Student ID: | 23200858 | |  | |  |
| Project Title: | Cloud-Based Healthcare Data Warehouse Management - Hair Loss Dataset | | CC Project: | | 1 |
| Due Date: | 04 / 12 / 2023 | Submitted Date: | | 04 / 12 / 2023 | |
|  | | | | | |

|  |
| --- |
| 1. Objective |
|  |

Big data is essential to the healthcare industry because it provides advanced analysis of disease trends and treatment effectiveness while highlighting the value of privacy and ethical data handling. In keeping with these ideas, the subsequent project centres on an Azure cloud service-based cloud-based solution for managing healthcare data warehouses.

The dataset under examination explores important aspects of hair health. Hair texture, total protein, and total keratin provide insight into structural and nutritional elements. The vitamins listed in the 'Vitamin' column are necessary for healthy hair. Manganese and iron levels offer more information, calcium supports physiological functions, and body water content indicates general health and hydration of the hair. Included are "Liver Data," which examines how the health of the liver affects different body functions, "Hair Fall," which measures hair loss, and Stress Level, which is known to have an impact on general health.

|  |
| --- |
| 2. Project Overview |
|  |

This application has been modified to make use of Azure cloud services. Azure Data Factory (ADF) connects to Power BI to enable simple data management and provides clear data visualization. By enabling reliable data management and informative visual representations of the intricate factors influencing hair health, this integration improves the project's efficiency.

**A diagram of a factory

Description automatically generated**

**A yellow sign with black text

Description automatically generated**

|  |
| --- |
| 3. Problem Description |
|  |

This project aims to understand and control factors that impact hair health, with a focus on hair loss. Given the importance of stress, poor nutrition, and general health in relation to hair-related issues, the project is set up to perform a detailed examination of a dataset on hair loss. The main objective is to find minor patterns and correlations in the dataset, which will offer priceless information for individualised healthcare interventions and creative solutions to problems with hair health.

|  |
| --- |
| 4. Dataset Collection |
|  |

The datasets chosen for this project include a wide range of parameters that are essential for understanding the specifics of hair health. These consist of body water content, stress levels, keratin levels, total protein, vitamins, essential minerals (calcium, iron, and manganese), hair texture, and a measurement of hair fall. Beginning from reliable sites like Kaggle, which houses a bounty of openly accessible datasets, the selected dataset functions as an inclusive storehouse of knowledge. Through the utilisation of these datasets, the project hopes to provide advanced analytics to healthcare professionals, enabling them to make data driven decisions that will improve patient care and overall well-being.

|  |
| --- |
| 5. Methodology and Implementation |
|  |

### Azure Cloud Setup:

For this project, the Azure Cloud Setup involved an organised process to set up the required infrastructure and resources. An outline of the main actions made is provided below:

#### 1. Azure Account Creation:

Using the university email address, aakash.sukre@ucdconnect.ie, an Azure account was created. A screenshot of a computer

Description automatically generated

#### 2. Acquiring Credits:

94.53 Euro credits were obtained for the Azure account, which gave the project the resources it needed to be implemented.

A screenshot of a computer

Description automatically generated

#### A screenshot of a computer Description automatically generated3. Resource Group Creation:

To logically arrange and manage related Azure resources, a special resource group called "HairLossApp-RG" was created.

#### 4. Resource Provisioning:

Within the resource group, specialised resources such as "hairlossstorage," "HairLossADF V2," and others that were necessary for the project were provisioned.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedHair Loss ADF

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Hair Loss Storage

Here, we have to upload the Excel file (hair\_loss\_dataset.xlsx) which we are going to use further.

A screenshot of a computer

Description automatically generated

### Hair Loss ADF Steps:

**1. Pipelines**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedSet Variable:

Purpose:

This first step in the pipeline is to set a variable. Values that can be applied to different pipeline activities can be stored in variables in ADF.

Implementation:

Used to set and maintain dynamic parameters or values for pipeline activities that come following.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedHairLossP1:

Purpose:

This step pipeline, called "HairLossP1," is to process the hair loss dataset as the primary workflow.

Implementation:

Provides a structured and refined dataset for additional analysis by incorporating tasks like data ingestion, pre-processing, and dataset cleaning.

**2. Datasets**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Input Excel file Output CSV file

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Input Excel File :

Purpose:

This dataset serves as the original data source for information about hair loss.

Implementation:

Set up to read information from the Excel file provided as an input and use that information as a basis for further processing.

Output CSV File:

Purpose:

This dataset is to hold the data that has been cleaned and processed.

Implementation:

Set up to export the converted data as a CSV file, providing a uniform structure for improved readability and analysis.

**3. Data Flows**

A screenshot of a computer

Description automatically generated

IngestData Pipeline:

A screenshot of a computer

Description automatically generated

Purpose:

This data flow activity is to extract raw data from the source Excel file, which is the first step in the data ingestion process.

Implementation:

To extract data effectively and maintain compatibility with downstream processes, the implementation makes use of mapping and transformation.

DataPreprocessing Pipeline:

A screenshot of a computer

Description automatically generated**A screenshot of a computer

Description automatically generated**

Purpose:

Preprocessing the extracted data, such as cleansing and transformation, is the scope of this data flow.

Implementation:

Here I used formatting, data cleansing such as renaming columns, Indexing, duplication removal and Excel to CSV transformations to improve the dataset's quality and organisation.

Cleandataset Pipeline:

**A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated**

Purpose:

The above dataset is further refined by this data flow through additional cleaning and organization procedures.

Implementation:

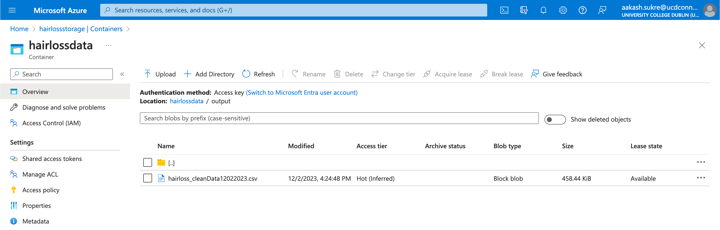
Here we are setting the parameter variable to fetch the data from the DataPreprocessing pipeline.

Summary:

Together, these ADF steps create an extensive data processing pipeline for the dataset on hair loss. They cover everything from the first stages of data intake to the development of a polished and organised dataset, which lays the foundation for further analysis and visualisation.

### Hair Loss Storage Steps:

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Purpose:

These Azure Storage steps collectively establish a secure, organised, and resilient foundation for storing and managing the hair loss dataset. They ensure data integrity, availability, and security, aligning with best practices for efficient data storage on the Azure platform.

Implementation:

The first step was to upload the Excel file containing the raw hair loss dataset to the assigned Azure Storage account.

Here, we can see that two-section input hair\_loss\_dataset.xlsx file and in the output folder hailoss\_cleandata12022023.csv. These files are stored into the storage account.

### Cost Management:

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer

Description automatically generated

Here I have needed cost for this project = 9.24 Euro.

Resource Group Optimisation:

The objective of the above method periodically evaluates the projects’ resource group to find and remove any unused resources. Using this Billing and Cost Management tool to track the resources utilisation inside the assigned resources groups and adjust optimisation on as necessary.

Optimisation of Data Storage:

To prevent excessive supply, regularly examine and improve data storage configuration.

* Description: Review Azure usage reports, cost analyses, and suggestions on a regular basis. As project requirements change, modify resource configurations and strategies accordingly.
* Implementation: Review Azure Cost Management reports on a regular basis and make necessary adjustments to resource configurations and strategies.

### Power BI Account Setup and Integration:

Implementation:

1. I Signed up for Power BI account using my university credentials [aakash.sukre@ucdconnect.ie](mailto:aakash.sukre@ucdconnect.ie)
2. After that in Power Bi, I select get data option and then give URL & the access key which credentials are come from the dataset file in the output folder of the blob storage.

**A screenshot of a computer

Description automatically generated**A screenshot of a computer

Description automatically generated**A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated**

1. After successfully connecting we can see all files here and we must choose the clean dataset file (hairloss\_cleandata12022023.csv) for the further data visualisation analysis dashboard.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

1. Now we must choose the extension and drilling option to get the proper dataset.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

Here we successfully configuered with the Power BI and The Dataset in Azure for Visualisation.

### A screenshot of a computer Description automatically generatedPower BI Visualisation:

Here, I have developed the dashboard using the power BI which relates to the Azure Blob Storage with my clean dataset file (hairloss\_cleandata12022023.csv).

Using the Power Bi Visualisation tool, I had shown the

1) Total Number of Patients 2) Level of Hair Fall using the

3) Iron Count Across Calcium Levels 4) Hair Health Analysis by Vitamin and Keratins Counts

5) Hair Fall Count Across Stress Levels 6) Protein and Keratins Sums by Hair Fall

Using this URL you can see my dashboard –

<https://app.powerbi.com/links/0vGNjuFUkO?ctid=420ec589-a866-4ad0-9a57-e6049e0d3bc0&pbi_source=linkShare&bookmarkGuid=679ce133-483e-4e02-a400-14182f2c3123>

### Deployment:

Here we can see that names which has been successfully deployed.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| 4. Suitability of Tools for Problem Solving |
|  |

Azure Cloud Services:

Strengths: Azure ensures cost effectiveness by offering a secure and scalable cloud environment. Sufficient training is necessary for the best possible use.

Azure Data Factory (ADF):

Strengths: Complex data workflows and smooth integration with other Azure services are made possible by ADF's pipeline architecture.

Azure Storage:

Strengths: Scalable and affordable storage with options for redundancy to ensure data integrity is provided by Azure Blob Storage.

Power BI:

Strengths: Power BI offers dynamic dashboard creation, real-time analysis capabilities, and user-friendly data visualisation tools.

When combined, the chosen tools create a complete ecosystem that can effectively handle the challenges associated with healthcare data management. Their security features, scalability, and integration abilities match the complexity of the project. Obstacles do arise, but they can be addressed with appropriate training and oversight procedures, guaranteeing maximum efficiency.

|  |
| --- |
| 5. Software features |
|  |

The software provides a broad feature set for efficient healthcare data management and analysis, and it consists of Azure Data Factory (ADF), Azure Storage, and Power BI. Structured data processing is made easier with ADF pipelines, data organisation and security are guaranteed by Azure Storage, and dynamic and informative data visualisations are made possible with Power BI. The project's successful execution is facilitated by the tools' seamless integration, which offers a reliable method for organising and evaluating medical information pertaining to hair health.

A yellow bar chart with a white background

Description automatically generated

**A green square with a white x

Description automatically generated**A green square with white text

Description automatically generatedA blue and white logo

Description automatically generated**A blue factory building with a few windows

Description automatically generated with medium confidenceA blue triangle with white background

Description automatically generated**

**A blue and white logo

Description automatically generated**

**Azure ADF AZURE STORAGE POWER BI**

|  |
| --- |
| 6. Worked Example: Managing Hair Loss Dataset with Azure Tools |
|  |

1. Setting Up the Azure Cloud:

Steps Done:

* Utilising the university email, I made an Azure account.
* 94.53 Euro in Azure credits were obtained.
* organised a special resource group called "HairLossApp-RG."
* Azure Data Factory (ADF) and Azure Storage were among the relevant
* resources that were provisioned within the resource group.

2. Actions for Azure Data Factory (ADF):

Steps Done:

* implemented ADF pipelines, such as "HairLossP1" and "Set Variable."
* Excel and cleaned CSV input and output datasets configured.
* Data flows like IngestData, DataPreprocessing, and CleanDataset have been developed.

3. Steps for Using Azure Storage:

Steps Done:

* Create a blob container and storage account.
* The unprocessed hair loss dataset (hair\_loss\_dataset.xlsx) was uploaded to Azure Storage.
* implemented access controls, redundancy, monitoring, and data backup in Azure Storage.

4. Integrating Power BI:

Steps Done:

* used the university login to create a Power BI account.
* To access the cleaned hair loss dataset, connect Power BI to Azure Blob Storage.
* Created visualisations, such as graphs and charts, using the imported dataset to highlight important hair health metrics.

6. Data Visualisations:

Example: Power BI

|  |
| --- |
| 7. Conclusion |
|  |

In conclusion, the Azure tool-implemented cloud-based healthcare data warehouse management solution for the hair loss dataset has shown to be a reliable and effective strategy. The project provided insightful information about factors influencing hair health while skilfully navigating the complexities of big data in healthcare. Azure Data Factory's well-organised pipelines made data processing easy and guaranteed the integrity of the dataset during ingestion, pre-processing, and cleaning. With vital features like redundancy and access controls, Azure Storage offered a safe platform for storing both raw and processed data. With sensible resource group optimisation and data storage configuration strategies in place, the pay-as-you-go model of Azure Cloud Services guaranteed cost-effectiveness. The combination of Power BI and Azure Blob Storage made it possible to create dynamic and perceptive data visualisations, which improved the interpretability of important hair health metrics. Power BI's collaboration features encouraged teamwork even more by making published reports available to stakeholders. All things considered, this project serves as an excellent example of the scalability, integration, and usefulness of Azure tools in the management of healthcare data, providing a solid basis for wise decision-making and creative breakthroughs in the field.

|  |
| --- |
| 8. References |
|  |

[1] Hair Loss Dataset on Kaggle: Brijlal Dhankour. (n.d.). Hair Loss Dataset. Retrieved from

<https://www.kaggle.com/datasets/brijlaldhankour/hair-loss-dataset/code>

[2] Microsoft Azure Data Factory Tutorials: Microsoft. (n.d.). Azure Data Factory Tutorials. Retrieved from <https://learn.microsoft.com/en-us/azure/data-factory/data-factory-tutorials>

[3] Azure Data Factory Tutorial Google Search: Google. (n.d.). Azure Data Factory Tutorial. Retrieved from <https://www.google.com/search?q=azure+data+factory+tutorial&oq=azure+data+factory+&gs_lcrp=EgZjaHJvbWUqBwgBEAAYgAQyBggAEEUYOTIHCAEQABiABDIHCAIQABiABDIHCAMQABiABDIHCAQQABiABDIGCAUQRRg8MgYIBhBFGDwyBggHEEUYPdIBCDYyNjNqMGo3qAIAsAIA&sourceid=chrome&ie=UTF-8>

[4] Azure Storage QuickStart Guide: Microsoft. (n.d.). QuickStart: Azure Blob storage using the Azure portal. Retrieved from

<https://learn.microsoft.com/en-us/azure/storage/blobs/storage-quickstart-blobs-portal>

[5] Azure Storage Tutorial: Edureka. (n.d.). Azure Storage Tutorial. Retrieved from <https://www.edureka.co/blog/azure-storage-tutorial/>

[6] Power BI Tutorial for Beginners: Data Camp. (n.d.). Power BI Tutorial for Beginners. Retrieved from

<https://www.datacamp.com/tutorial/tutorial-power-bi-for-beginners>

[7] Building an ETL Data Pipeline using Azure Data Factory: Analytics Vidhya. (2022, June). Building an ETL Data Pipeline using Azure Data Factory. Retrieved from

<https://www.analyticsvidhya.com/blog/2022/06/building-an-etl-data-pipeline-using-azure-data-factory/>