Configuration Manual
Achieving Green Data Centres for Sustainable Cloud Computing

MSc Research Project
MSc in Cloud Computing

Mariam Ishaq
Student ID: X23259817

School of Computing
National College of Ireland

Supervisor: Sean Heeney

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Mariam Ishaq
Student ID:	X23259817
Programme:	Programme Name
Year:	2024
Module:	MSc Research Project
Supervisor:	Sean Heeney
Submission Due Date:	12/12/2024
Project Title:	Configuration Manual
Word Count:	—
Page Count:	8

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	12th December 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Mariam Ishaq
X23259817

1 Introduction

This configuration manual presents a clear and detailed approach outlining how the cloud-based energy-efficient task scheduling system is supposed to be configured and utilised. It including informs about AWS settings, script of simulation, used machine learning algorithms, and measures of performance evaluation.

2 Configuration Manual

Step 1: Log into AWS Management Console

- Open the AWS Management Console in your browser.
- Navigate to the **EC2** dashboard.

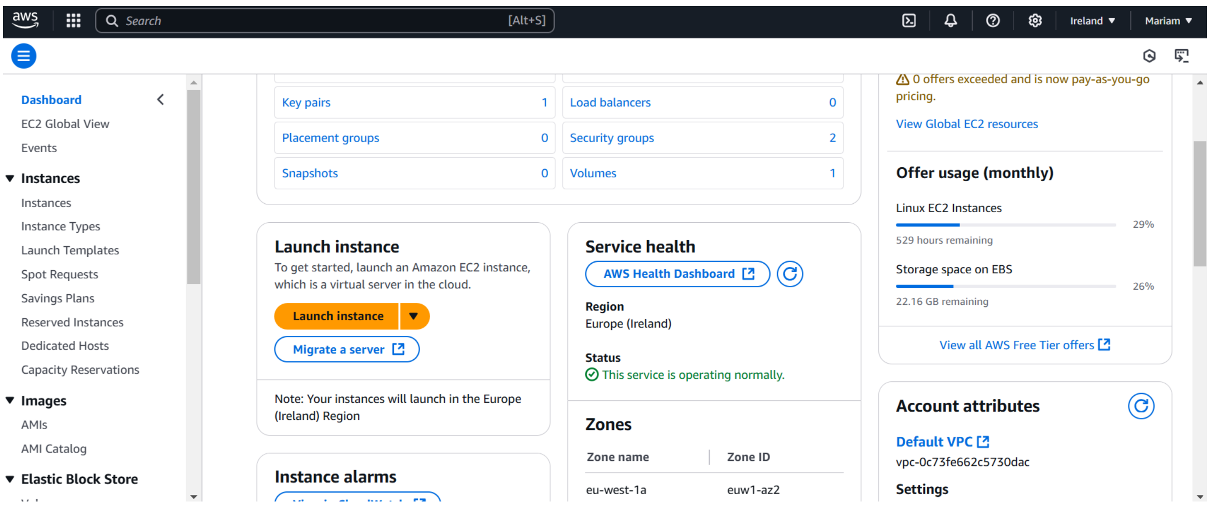


Figure 1: EC2 Dashboard - Launch Instance button

Step 2: Launch a New EC2 Instance

- On the EC2 dashboard, click **Launch Instance**.

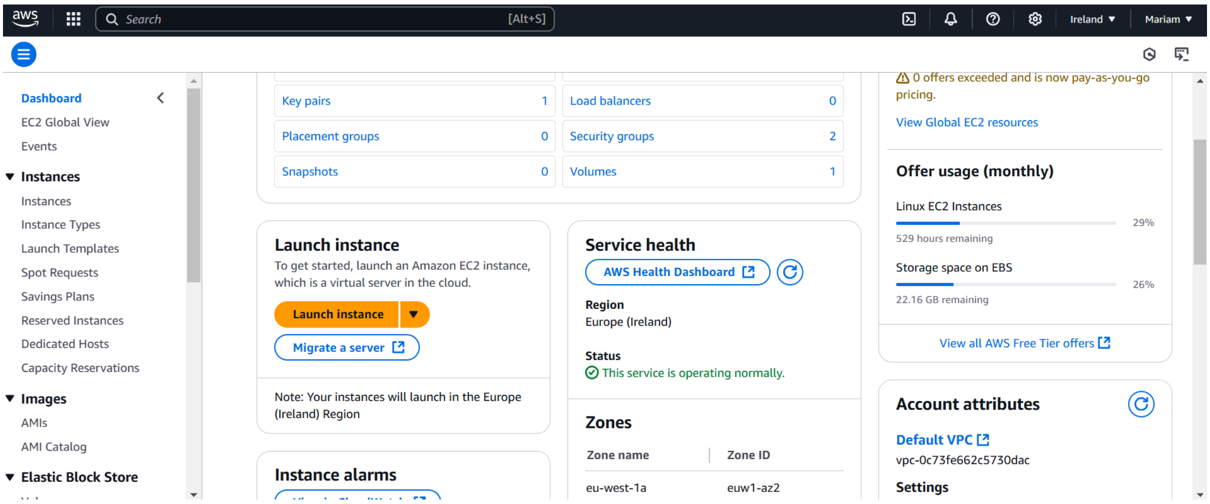
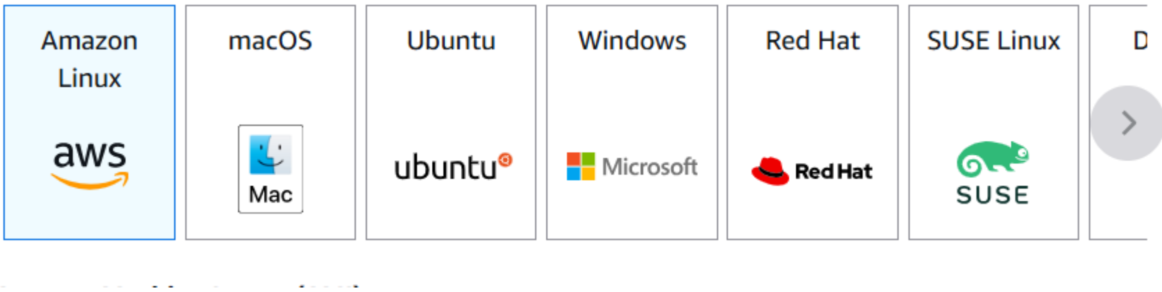
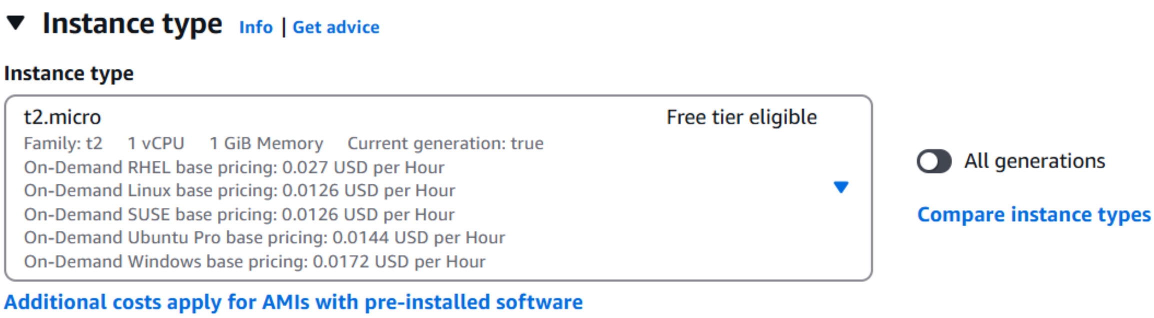


Figure 2: EC2 Dashboard - Launch Instance button

- **Choose an AMI**:
  - Select **Amazon Linux 2 AMI (Amazon Machine Image)**.



- Select **t2.micro** for basic setup



- **Note**: The **t2.micro** is eligible for the **AWS Free Tier**.

Click **Next: Configure Instance Details**.

Step 3: Configure Instance Details

Network:

- Select the default **VPC**
- **Subnet**: Choose a subnet **eu-west-1a**
- **Auto-assign Public IP** is set to **Enable**

IAM Role:

- Leave **IAM Role** as None

Advanced Details:

- Leave default

Click **Next: Add Storage**.

Step 4: Add Storage

- **Configure Storage**:
  The default 8 GB storage is **General Purpose SSD**) is sufficient for the research. Show the default storage configuration screen.
- Click **Next: Add Tags**.

Step 5: Add Tags

- **Add Name Tag**: Under **Key**, input `Name` and set the **Value** to `Mariam_123`.

Click **Next: Configure Security Group**.

Step 6: Configure Security Group

- **Create a New Security Group**:
  - Choose **Create a new security group** and configure it.
  - Add an **SSH rule**:
    * **Type**: SSH
    * **Port Range**: 22
    * **Source**: Custom (your IP address or `0.0.0.0/0` for global access)
- Add additional rules for HTTP/HTTPS if you plan to run a web server.
- Click **Review and Launch**.

Step 7: Review and Launch

- **Review** all configuration details:
  - Instance Type: `t2.micro`
  - AMI: Amazon Linux 2
  - VPC/Subnet: Default or custom selection
  - Security Group: As configured above
- **Select Key Pair**:
  Click **Create a new key pair**, download the `.pem` file, and save it securely.
- Click **Launch** to start the instance.

Step 8: Access Your EC2 Instance

- **Obtain the Public IP**:
  - Once the instance is running, go to the **Instances** page in EC2 and find your instance.
  - Note the **Public IPv4 DNS**

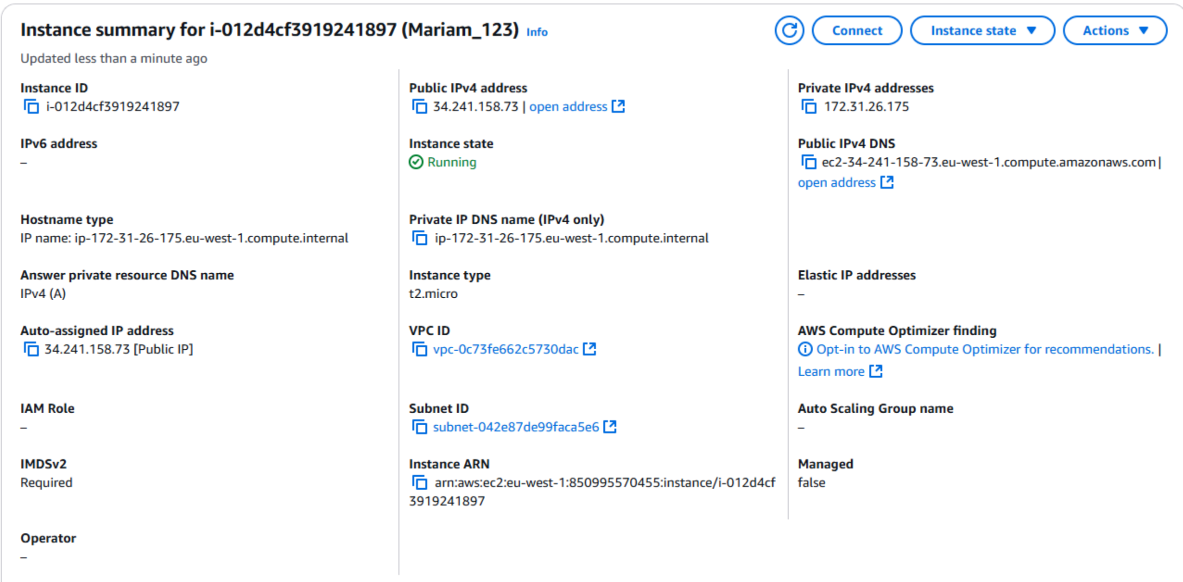


Figure 3: EC2 Instance Configuration Details

- **SSH Access**: Open a terminal and use SSH to access the instance:

  ssh -i /path/to/key.pem ec2-user@34.241.158.73

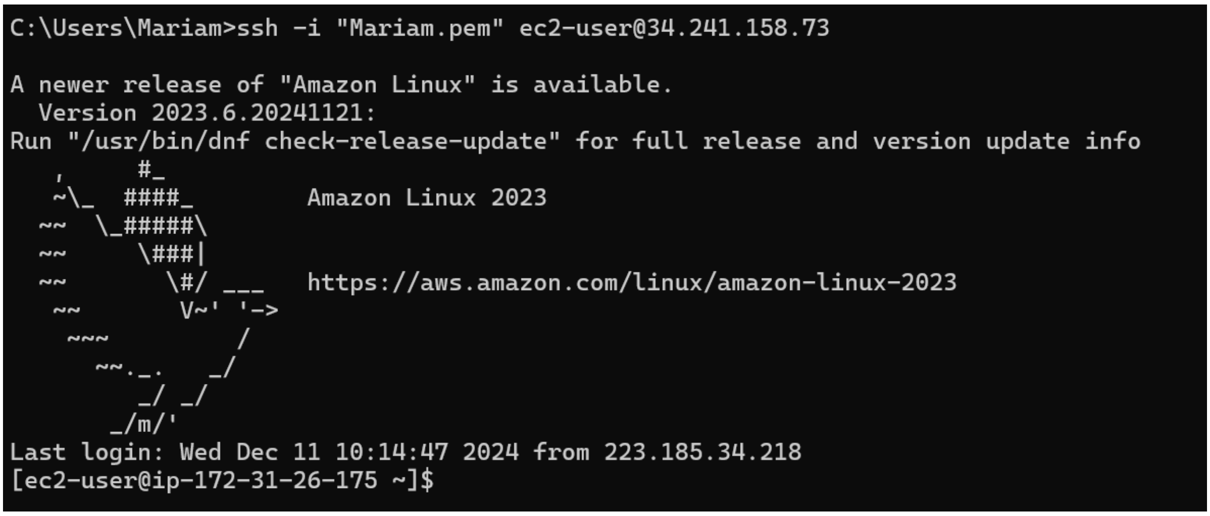


Figure 4: ssh command to access EC2

Step 9: Access the Scripts

This document contains configuration information for the four core scripts employed in the project. Thus, one script is the devoted to a particular function of miming energy efficiency while the other script is devoted to the miming of resources allocation. on management in cloud computing platforms. The scripts include data preparation before building the model, and the model building processinclude: comparison, cloud simulation by SimPy and energy consumption forecast.

Script 1: Data Preprocessing and Feature Engineering
The first script deals with the data preprocessing step and covers cleaning, feature extraction/selection steps. tion, and transformation. It is used to preprocess the data to feed them into the machine learning models due to the practice of the missing values and feature scaling and selection.
Configuration and Setup

- Install necessary libraries: [language=bash] pip install pandas numpy scikit-learn
- Input Data:
  - The script expects a CSV file containing raw data with columns like CPU usage, memory usage, energy consumption, and workload.
  - Modify the input path in the script to point to your local or cloud-based dataset.
- Feature Engineering:
  - The script includes functions for:
    * Handling missing values (using mean imputation or removal).
    * Normalizing continuous features (using Min-Max scaling).

      * Selecting relevant features (based on correlation matrix).

Execution Steps

1. Load and clean the dataset.
2. Normalize their format and feature engineer estimated power consumtion
3. Separate the data into two parts for training and one for the tests.
4. This needs to be stored for utilizing in model training.

Script 2: Model Comparison
This script features the results of three models: XGBoost, MLP Regressor, and Genetic Algorithm using Mean Squared Error (MSE) as the HP mesure) metric. They compare their forecasting prowess for energy usage and resources requirements in cloud computing systems.
Configuration and Setup

- Install required libraries: [language=bash] pip install xgboost scikit-learn deap matplotlib
- Input Data:
  - Use the preprocessed dataset generated by Script 1. The input data should include normalized features and target labels (e.g., energy consumption).
- Model Configuration:
  - XGBoost:
    * The script allows customization of hyperparameters such as learning rate, depth of trees, and regularization.
  - MLP Regressor:
    * Modify the number of layers, neurons per layer, and activation function as needed.
  - Genetic Algorithm:
    * Customize task size, mutation rate, and selection method.

Execution Steps

1. Train each model using the processed dataset.
2. Evaluate the models using Mean Squared Error (MSE).
3. Plot the results to compare model performance.
4. Adjust model parameters and rerun if needed to optimize performance.

Script 3: Cloud Simulation with SimPy
This script mimics operating resource in a cloud computing data center. It models the Processing capacity loading graph, memory loading graph, scheduling graph and energy consumption SimPy. The aim is to achieve the goals of power consumption rationalization while providing the services with proper quality in the clouds.
Configuration and Setup

- Install necessary libraries: pip install simpy numpy
- Cloud Resource Setup:
  - Describe the task types, the resources needed for completion and time needed to complete the tasks.
  - Introducing dynamic script so that the moving schedule can operate effectively in relation to energy preservation. Execution Steps
- Task Scheduling:
  - Define the task types, their resource requirements, and duration.
  - The script supports dynamic scheduling to optimize energy efficiency.

Execution Steps

1. Start the SimPy environment and start all cloud resources.
2. Simulate resource allocation and task scheduling.
3. Calculate energy consumption based on CPU and memory usage.
4. Implement failure handling and rerun simulations if necessary.

Script 4: Energy Consumption Prediction with XGBoost
This script aims to the extend of using the XGBoost algorithm to predict the energy consumption in the cloud data centre provided by the workload and resource utilization as inputs to the quantitative formula. The script trains the model based on the historical information and then makes prospect for usage of energy in the future.
Configuration and Setup

- Install required libraries: [language=bash] pip install xgboost scikit-learn pandas
- Input Data:
  - The dataset should include features like CPU usage, memory usage, and workload, along with energy consumption as the target variable.
  - Ensure the data is preprocessed and scaled (using the output from Script 1).
- Model Configuration:

- Modify XGBoost hyperparameters such as learning rate, tree depth, and regularization to fine-tune the model.
- Optionally, include cross-validation to optimize model performance.

Execution Steps

1. Load the preprocessed dataset.
2. Train the XGBoost model on the data.
3. Evaluate model performance using MSE.
4. Make predictions for future energy consumption and assess model accuracy.

3 Conclusion

This system enables users to balance the cloud resources to achieve power saving. In this concept, users are able to predict and minimize the energy utilization in cloud data centers through machine learning models as well as cloud simulation.

Caution Make sure that the instance is correct with a sufficient level of security, I once limited the SSH access to specific IP. Continually assess system performance and fine-tune parameters of the models to enhance the overall performance.