SIT314/SIT729 – Week 9 Group Activity  
Using Cloud Computing for IoT Application Scalability

short line

# Overview

# This activity is to think about the use of Cloud Computing load balancing and auto-scaling.

# Tasks

For an IoT health application that has the following parts:

* Client that collects health data
* API that allows data to be uploaded from the client and store in a MongoDB Database.
* A MongoDB Database.
* A microservice which calculates the average heartrate of a user.
* A microservice which calculates the yearly fitness of a user.
* A microservice that can provide the fitness level for all the runners in a marathon.
* A microservice which will detect when individuals are having health issues.

Think about the use of Cloud Computing and answer the following questions.

1. What parts of this application could be replicated on Cloud?

All components except for the client-side device (which is likely a physical wearable or mobile device) can be replicated on the cloud. This includes:

* - The API server
* -The MongoDB database
* - All four microservices (average heartrate, yearly fitness, marathon fitness level, health issue detection)

These services benefit from elasticity, reliability, and availability offered by cloud platforms like AWS, Azure, or GCP.

1. What Load-Balancing policy would you use?

* A **Round Robin** policy is suitable for stateless microservices (like the ones performing calculations) to evenly distribute requests.
* For stateful services like the MongoDB database, **Least Connections** or session-based sticky load balancing could be more appropriate to maintain continuity and performance.

1. What Auto-Scaling policy would you use?

**CPU-based auto-scaling** for microservices: Increase instances when CPU usage exceeds 70%.

**Request-rate-based auto-scaling** for the API server: Scale when request rate crosses a threshold (e.g., 1000 requests/minute).

**Memory-based auto-scaling** for the MongoDB database to handle spikes in storage or query demands.