CSCI 5902 Advanced Cloud Architecting (Fall 2023)
Assignment 2 – EC2

Scenario: You're the Chief Cloud Architect at Spacetech Galactic, a pioneer in the era of space tourism. Their magnum opus, the Spacetech Holodeck, lets millions of customers around the world experience the thrill of virtual space travel from their homes. From piloting state-of-the-art starships to venturing into the Martian wilderness and even simulating the surreal tranquility of zero gravity, all these experiences are now accessible at the click of a button. The Holodeck application hosts an intricate web of high-definition, interactive 3D models of planets, spaceships, and extraterrestrial lifeforms. Add to it the real-time physics engine, that simulates zero gravity, spaceship thrusters, and even warp speeds, the server demand becomes substantial. With extreme variability in user traffic due to weekends, holidays, or when a new starship model is unveiled, the requirement for computational resources can surge at a moment's notice.

As the application's popularity extends to the farthest corners of the globe, the demand for consistently low latency becomes non-negotiable to ensure smooth and immersive user experiences. Further complexity is added by the multiplayer mode, allowing users from diverse geographical locations to interact and embark on interstellar adventures together. This necessitates maintaining high-throughput, low-latency connections between instances across the globe.

The vast cosmos of data collected from the users and the multitude of static content required for the VR experiences necessitates a thoughtful strategy for data storage and processing. Not to mention, the non-critical workloads, like generating analytical reports and predictive models of customer behavior, could be processed during off-peak hours to optimize resources and costs, but would need an intelligent system to manage these operations.

The mission for Spacetech Galactic has always been to explore new frontiers while keeping an eye on the operational costs. The architecture would have to balance the high computational demands of the Holodeck application and the company's economic sustainability. Hence, an intricate understanding of instance pricing models and their appropriate utilization would be crucial.

The physics simulation engine, the heart of the Holodeck experience, comes with its own set of challenges. It has strict licensing requirements that necessitate the use of dedicated hardware for its operation. This adds another layer of complexity to the architectural design.
Spacetech Galactic always stands at the bleeding edge of technology. Regular system updates, patches, and maintenance activities are part and parcel of the operation. However, these shouldn't cause significant downtimes, disrupting the out-of-the-world experiences of the users.

Now, you have the challenging yet exciting task of designing a comprehensive, high-performing, secure, and cost-effective cloud architecture using Amazon EC2 and associated AWS services.

Can you ensure that Spacetech Galactic continues its successful journey through the virtual cosmos? Remember, in space, everyone can hear you code. Good luck!

Tasks (Total marks: 15):

1. **Instance Selection and Auto Scaling:** Evaluate the requirements of the application and determine the type and size of EC2 instances that would be most appropriate for your needs. Consider factors like the application's CPU, memory, storage, and network performance requirements. **(2 marks)**

2. **Multi-AZ and Multi-Region Deployment:** Implement a Multi-AZ and Multi-Region deployment to ensure high availability and fault tolerance. **(1 mark)**

3. **EBS & EFS:** Plan a strategy for data storage using Amazon EBS and/or EFS. **(1.5 marks)**

4. **Spot and Reserved Instances:** Incorporate spot instances and reserved instances in your architecture. **(2 marks)**

5. **AMI:** Explain how you would use Amazon Machine Images (AMIs) to quickly deploy and replicate your application. **(1.5 mark)**

6. **Cost Management:** Provide a rough estimate of the costs of running this infrastructure and discuss the strategies you would use to manage these costs. **(1.5 marks)**

7. **EC2 Placement Groups:** Design an architecture where the application has a need for low-latency, high throughput communication between instances. **(1 mark)**

8. **Instance Store:** Design a scenario where temporary, high-IOPS storage is required and an instance store would be used. **(1 mark)**

9. **Dedicated Hosts/Instances:** Plan for scenarios where the application has to comply with strict licensing terms (BYOL) or meet dedicated hardware requirements. **(1.5 mark)**

10. **EC2 Metadata and User Data:** Describe how you would use EC2 metadata and user data to handle configuration tasks and pass information to instances at launch time. **(1 mark)**

11. **Optimization and Performance:** Describe how to use tools like AWS Compute Optimizer and Trusted Advisor for identifying optimal EC2 instance types and for maintaining cost efficiency. **(1 mark)**

How to do this assignment?
Write a detailed response for each task, explaining how your proposed design caters to the task's needs and the rationale behind your architectural decisions.

If you want to design and draw your architecture diagram, you can use the free on-line tool DrawIO: https://app.diagrams.net. Drawing the diagram is NOT mandotary for this assignment.

Marking rubric:
• Full marks: A well-written, clear, detailed, and reasonable answer that meets all the requirements of the question and demonstrates a deep understanding of the concepts and compute services used.
• Half marks: A partially complete or somewhat unclear answer that addresses some of the requirements of the question and demonstrates a basic understanding of the concepts and AWS compute services used.
• Zero point: An incomplete or inaccurate answer that does not meet the requirements of the question and demonstrates a lack of understanding of the concepts and AWS compute services used.

Academic Standards:
The assignment must adhere to academic standards, including proper citation and referencing of sources, adherence to academic writing standards, and originality of work. Plagiarism or any form of academic misconduct is not tolerated and may result in an academic integrity offense. The report must demonstrate a thorough understanding of the concepts and storage services used and must be original work. It should be well-organized, clear, concise, and free of errors in grammar and spelling.

Note:
Remember, there's no single "correct" answer for this assignment. We're looking for creative, well-reasoned cloud architectures that meet the needs of the scenario. This assignment is not only about demonstrating your understanding of Amazon EC2 and related AWS services but also about showcasing your problem-solving, decision-making, and architectural design skills.

FAQ:
Q: How many pages should the assignment be?
A: There is no specific page requirement, but the answers should be long and clear enough to address all the questions thoroughly and clearly.

Q: Can I use external sources to support my report?
A: Yes, you may use external sources such as research papers, articles, or official documentation from cloud providers to support your answers. However, you should ensure that all sources are properly cited and referenced in the report to avoid plagiarism.

Q: What citation and reference format should I use for the assignment?
A: You should use either the APA or IEEE citation and reference format.

Q: Is there any specific format for the assignment?

A: No, there is no specific format, but make sure that your assignment is well-organized, properly formatted, and easy to read and understand, as this will help the marker to evaluate your work more effectively.

For any questions or concerns, reach out to the course TA.

Good luck :)