

PLEDGE TO PROGRESS

Sustainability Hackathon

Sponsored By



Your Team Name :

GirlHackToFuture

Your team bio :

Cloud developer in security .

Creating a API to reduce the energy used in data based
compression and duplication as per user needs and prompting the
developer about the compatibility issues to assigned

Date :24/04/2023

Problem Statement?

Why did you decide to solve this Problem statement?

As data is the new revolution and everyday millions of data is stored by every organization it is essential to store the data compressed and optimise it whenever required.

The data stored in the data center can be optimised to reduce energy consumption and reduce the carbon-footprint created by the software development sector.

User Segment & Pain Points

Which user /advertiser segment would be early adopter of your product & why?

Datacentre users who require database servers to store data in the data centre.

If the data is stored in compressed form and that to optimised accordingly it can reduce the energy consumption.

Pre-Requisite

What are the alternatives/competitive products for the problem you are solving?

Limited functionality: Some software solutions may offer limited functionality or may not address all of the areas that need optimization. It's important to evaluate the specific needs of the cloud server and choose a solution that offers comprehensive optimization capabilities.

Sharing improvement plans for compatibility before the software is created and then create a software to tackle the compatibility issue .

Security concerns: Optimizing cloud server data can involve sensitive information, and it's important to choose a solution that offers robust security features to protect against data breaches or unauthorized access.

Cloud already provides various security solutions

Lack of customization: Some software solutions may offer limited customization options, which can limit their effectiveness in optimizing cloud server data for specific use cases or configurations. It's important to choose a solution that offers customization options to meet the specific needs of the cloud server.

We can customize the API using which type of data is to be compressed and then delete the redundant data as per the administrator commands ,which is not available in the software's available in the market

Tools or resources

Azure tools or resources which are likely to be used by you for the prototype, if your idea gets selected

Azure AI to incorporate and scale the tool used .

Azure AI – to check various algorithms testing and check for the status .

Developer optimizing tool to prompt the developer with the compatability issues .

Azure Cosmos DB

Any Supporting Functional Documents

Present your solution, talk about methodology, architecture & scalability

This API will check for the compatability of the software and data set with the cloud enviornment and then compressing the data accordingly .

<https://learn.microsoft.com/en-us/azure/cdn/cdn-improve-performance>

CDN (Content Delivery Network): A CDN is a network of servers that are distributed across different geographic locations. By using a CDN, you can store copies of frequently accessed data in multiple locations, which can improve performance and reduce the amount of time it takes to transfer data. This can also help reduce energy consumption by reducing the amount of data that needs to be transferred across the internet.

1.Checking which duplication software to use so that the redundant data can be deleted and Hashing: One of the most common methods of deduplication is to use hashing. In this approach, a hash function is applied to each piece of data in the dataset, and the resulting hash value is stored. When new data is added to the dataset, it is compared against the existing hash values. If a match is found, the data is considered redundant and is not stored.

2.Content-Defined Chunking: This method breaks the data into chunks of fixed size, called blocks, based on the content rather than a set block size. Each block is assigned a hash value, and duplicates are identified by matching the hash values.

3.Delta Encoding: In delta encoding, each piece of data is compared against the previous piece of data in the dataset. If the new data is similar to the previous data, only the differences between the two are stored. This can be useful for datasets that are updated frequently, as only the changes need to be stored.

4.Compression: Compression algorithms can be used to remove redundant data by compressing the dataset and identifying patterns that can be stored more efficiently.

Any Supporting Functional Documents

Also checking the compression algorithms

- 1.Zstandard API: Zstandard is a real-time compression algorithm that provides high compression ratios while still maintaining fast compression and decompression speeds. The Zstandard API can be used to integrate Zstandard compression into a database or other storage system, allowing data to be stored more efficiently and with less energy consumption.
- 2.Snappy API: Snappy is a compression algorithm that is optimized for speed and is commonly used for compressing data that is transmitted over a network. The Snappy API can be used to integrate Snappy compression into a database or other storage system, allowing data to be stored more efficiently and with less energy consumption.
- 3.LZ4 API: LZ4 is a compression algorithm that provides fast compression and decompression speeds while still achieving good compression ratios. The LZ4 API can be used to integrate LZ4 compression into a database or other storage system, allowing data to be stored more efficiently and with less energy consumption.
- 4.Brotli API: Brotli is a newer compression algorithm that was developed by Google and is designed to provide high compression ratios while still maintaining fast compression and decompression speeds. The Brotli API can be used to integrate Brotli compression into a database or other storage system, allowing data to be stored more efficiently and with less energy consumption.

Key Differentiators & Adoption Plan

How is your solution better than alternatives and how do you plan to build adoption?

Selection of the best algorithm for data compression and data duplication using Azure AI

We can customize the API using which type of data is to be compressed and then delete the redundant data as per the administrator commands ,which is not available in the software's available in the market

Create an interface for developers that prompts the compatibility software issue and provide solution on it which reduces the end of life o software so that it can be used for a longer duration of time reducing the e-waste issue as well .

Sharing improvement plans for compatibility before the software is created and then create a software to tackle the compatibility issue .

GitHub Repository Link & supporting diagrams, screenshots, if any

How far it can go?

Selection of the best algorithm for data compression and data duplicationa using Azure AI

Create an interface for developers that prompts the compatibility software issue and provide solution on it which reduces the end of life o software so that it can be used for a longer duration of time reducing the e-waste issue as well .

TECHGIG

Thank You

Team member names