Curtin Summer Project Plan

Week 1:

Literature review on large-scale data management techniques and practices.

Data Compression and Storage Optimization

And so on.

Understand its specific requirements and challenges related to data management, including p

Week 2:

Getting hands on experience on common compression algorithms like gzip and zlib (Algorithm

DEFLATE algorithm, Huffman coding, and so on

Work on similar algorithms at code level

Deliverable: Research report summarizing findings and potential areas of innovation.

Week 3:

Develop a conceptual framework for data management.

Research machine learning algorithms for data optimization.

Study well-known machine learning algorithms such as Autoencoders and GANs to analyse d

Design preliminary models using machine learning algorithms (autoencoders, GANs) for data

Implement and debug sample codes in Python.

Week 4:

Design data partitioning strategies and investigate intelligent caching mechanisms.

Enhancing parallel processing and minimizing data transfer overhead.

Learn about caching technique, e.g., LRU (Least Recently Used) Caching

Deliverable: Prototype of data compression and partitioning strategies.: Initial design documen

Week 5:

Evaluate the performance of different compression algorithms on the selected use case's data

Task 5.1: Select a subset of the use case data for testing.

Task 5.2: Implement a benchmarking framework to evaluate different compression algorithms

Task 5.3: Run the compression algorithms on the test data, measuring key performance metri

Task 5.4: Analyze the results to identify the most effective compression algorithm for the spec

Deliverable: A detailed report on the performance of each compression algorithm, with recom

Develop intelligent data partitioning strategies based on clustering or decision trees to distribu

Task 5.5: Research and select suitable machine learning models for intelligent data partitionin

Task 5.6: Design a prototype model to partition the data based on factors like data type, acces

Task 5.7: Develop a simulation environment to test the partitioning strategy on a sample of the

Task 5.8: Analyse the effectiveness of the partitioning model in terms of data retrieval efficien

Deliverable: A prototype of the data partitioning model and a simulation report evaluating its e

Implement and refine the algorithms and models, considering efficiency and scalability for larg

Task 5.9: Integrate the selected compression algorithm with the data partitioning model.

Task 5.10: Run a series of tests to ensure that the integrated system works efficiently on a lar

Task 5.11: Identify any bottlenecks or inefficiencies and refine the algorithms and models acc

Task 5.12: Document the refinements and their impact on the overall system performance.

Deliverable: An integrated system of compression and partitioning models, along with a docur

Expected Outcomes:

By the end of Week 5, you should have:

A clear understanding of which compression algorithms are most suitable for your use case.

A prototype model for intelligent data partitioning.

An integrated system that combines efficient data compression with smart partitioning strategi

Week 6:

Start implementing data management systems.

Develop experimental setups for performance evaluations.

Initial testing and refinement of models.

Deliverable: Prototype of data compression and partitioning strategies.

Detailed Tasks may consider as:

Start implementing data management systems.

Task 6.1: Finalize the design of the data management system based on the insights and resul

Task 6.2: Begin the implementation process by setting up the necessary infrastructure, includi

Task 6.3: Install and configure the software components of the data management system, ens

Task 6.4: Conduct initial tests to ensure that the system components are functioning correctly

Deliverable: An operational draft of the data management system ready for further testing and

Develop experimental setups for performance evaluations.

Task 6.5: Design experimental setups to evaluate the performance of the data management s

Task 6.6: Prepare datasets that will be used for testing. Ensure these datasets are representa

Task 6.7: Develop scripts or programs to automate the testing process. This automation will e

Task 6.8: Validate the experimental setup by running preliminary tests to check if it is capable

Deliverable: A comprehensive experimental setup ready to be used for in-depth performance

By the end of Week 6, the project should have:

A functioning prototype of the data management system, incorporating the previously develop

A detailed and validated experimental framework capable of rigorously assessing the perform

Week 7:

Conduct initial experiments and evaluations.

Gather preliminary results and identify areas for optimization.

Week 8:

Document the findings, including performance measurements, insights gained, and recommen

Summarize the techniques, algorithms, and models used in the project and