Unit 4 – Cloud Programming Models

Introduction

- Cloud programming models are a set of programming abstractions that allow developers to build and deploy applications on cloud computing platforms
- These models provide a high-level view of the cloud infrastructure, and hide the details of how the applications are actually implemented
- This makes it easier for developers to focus on the business logic of their applications, and less concerned with the underlying cloud infrastructure.

Thread Programming

- Thread programming is a programming model that allows multiple threads of execution to run concurrently within a single process
- Thread programming can be used to improve the performance of applications that are deployed on cloud servers
- By running multiple threads of execution on different cloud servers, applications can take advantage of the parallel processing capabilities of the cloud infrastructure

Thread Programming (contd..)

Benefits

- Improved performance
- Scalability
- Reliability

Challenges

- Complexity
- Concurrency issues

Thread Pool

- A thread pool is a collection of threads that are managed by the application
- When an application needs to execute a task, it submits the task to the thread pool.
- The thread pool then chooses a thread from its collection to execute the task.

Task Programming

- A task-based programming model allows applications to break down their work into a series of tasks
- These tasks can then be executed in parallel on different cloud servers
- This can improve the performance of applications by allowing them to take advantage of the parallel processing capabilities of the cloud infrastructure

Task Based Programming

- It Provides a set of APIs that can be used to create and manage tasks
- These APIs can be used to submit tasks to the cloud, track the progress of tasks, and collect the results of tasks
- If an application needs to be able to execute a large number of tasks in parallel, task based programming is used

Workflow Management System

- Provides a graphical user interface that can be used to create and manage workflows
- A workflow is a series of tasks that are executed in a specific order
- Workflow management systems can be used to automate the execution of tasks, which can improve the efficiency of applications
- If an application needs to be able to automate the execution of tasks, then a workflow management system is used

Benefits

- Improved performance
- Scalability
- Reliability

Challenges

- Complexity
- Concurrency issues

Map Reduced Programming

- MapReduce is a programming model and an associated implementation for processing and generating large data sets
- It is designed to scale up from single machines to thousands of machines, each of which can have multiple processing cores
- Supported by a number of cloud computing platforms, such as Amazon Web Services, Microsoft Azure, and Google Cloud Platform

MRP – Working Principle

- The data is split into smaller chunks that can be processed by individual machines.
- Each machine runs the map function on its chunk of data. The map function takes a piece of data and produces a set of key-value pairs.
- The key-value pairs from all the machines are then shuffled and redistributed.
- Each machine runs the reduce function on its set of key-value pairs. The reduce function takes a set of key-value pairs and produces a smaller set of key-value pairs.
- The results from all the machines are then combined to produce the final output.

MRP - Usages

- Log Analysis
- Data Mining
- Machine Learning

MRP - Efficiency

- Depends on a number of factors, including the size of the dataset, the number of machines used, and the type of computation being performed.
- Most efficient for large datasets that can be divided into smaller chunks that can be processed in parallel.

- Factors affecting MRP:
  - Data Size
  - Number of Machines
  - Types of Computation

MRP – Parallel Efficiency

- The ratio of the actual execution time of a job to the theoretical execution time of the same job on a single machine
- It is a measure of how well MapReduce can take advantage of multiple machines to improve performance

- Improving Parallel Efficiency:
  - Choosing right number of machines
  - Optimizing the map and reduce functions
  - Using a distributed file system

Enterprise Batch Processing using MRP

- The process of running large-scale data processing jobs on a regular schedule

- Benefits:
  - Scalability
  - Fault tolerance
  - Ease of use
  - Cost Effective