Java Concepts in MapReduce Programming

Key Java concepts are essential in MapReduce programming using Hadoop.

by A.Pardha Preetham

Object-Oriented Programming (OOP)

Core Principle

Java is the foundation for MapReduce programs in Hadoop.

Role in MapReduce

Classes and objects play a central role in the MapReduce framework.

Java I/O Streams

Data Reading

Input data is read from a text file using Hadoop's InputFormat.

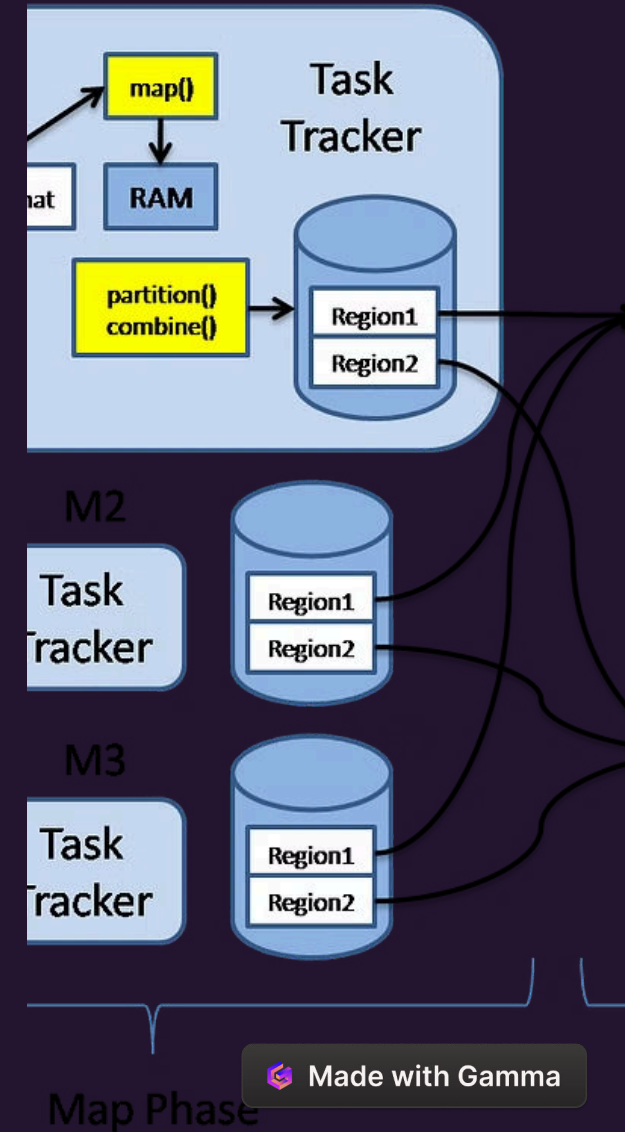### Data Writing

Results are written using Hadoop's OutputFormat.

MapReduce API

1 ──── **Core Programming**

Hadoop MapReduce API includes classes like Mapper and Reducer.

2 ──── **Usage in Programs**

These classes are extended and used in MapReduce programs.

Java Collections Framework

Usage

java.util package and its collections framework are often used.

Example

HashMap or other collections might be used in programs.

Concurrency (Multithreading)

1 **Parallel Processing**

Hadoop manages parallel processing of data using MapReduce.

2 **Developer's Role**

Understanding concurrency concepts is beneficial.

SQL-Exception

ClassNotFound

Memory

Framework

Exception Handling

Crucial Mechanism

Java's exception handling is crucial in MapReduce programming.

Error Management

Dealing with potential errors during job execution.

Serialization and Deserialization

Data Transmission

1 Data is serialized before being sent between nodes.

Crucial Process

2 Converting data into a format for easy transmission.

File

Data Base

Memory

Configuration Management

1

Key-Value Pairs

2

Customizing MapReduce job behavior.