CS565: Cloud Computing

Reading Assignment 2

Solutions   **Roll No.**: 1801CS31   **Date**: 22-Apr-2022

Ans 1:

भारतीय प्रौद्योगिकी संस्थान पटना
INDIAN INSTITUTE OF TECHNOLOGY PATNA
Supplementary Sheet   Sheet No. [E]

Invigilator's Signature

Name : Mahesh Babu
Roll Number : 1801CS36
Course Number and Name : Cloud Computing (CS565)

1. Limitations of server-client architecture:
    a) If all the clients simultaneously request data from the server, it may get overloaded
    b) Cost of setting and maintaining is high.
    c) If server fails, none of the requests can be fulfilled.
    d) The high latency may cause the results to obsolete.

• How edge stream processing addresses these issues.

Edge stream processing applies the stream processing paradigm to the edge computing architecture. Instead of relying on the cloud to process sensor data, the edge stream processing system relies on distributed edge compute nodes which are near the data sources to process data and trigger actuators.

The server-client architecture infrastructure cannot afford massive data streams whereas in edge stream processing, many IOT stream applications will run concurrently, consuming these live data streams to quickly derive insights and make decision.

Ans 2:
<Answer goes here>

Ans 3:

③ Failure recovery mechanism on top of dynamic dataflow Abstraction

As the overlay is self-organizing and self repairing, the dataflow graph for each IoT application can be automatically recovered by restarting failed operators on another node. DART System has 2 Components: a set of distributed schedulers and a set of workers. A general approach is checkpointing periodically for all operator States to a persistent storage Systems and failover node retrieves Checkpointed state upon failure. Since, this is Slow approach we design a parllel recovery approach where periodically, larger than state memory is divided replicated and checkpointed to each node's leaf set nodes by using erasurecode. Once any failure happens, backup nodes take over and retrives state fragments from a Subset of leaf. Set nodes to recompute state and resume processing. Recovery Process is fast because many nodes can leverage the dataflow graph to recompute the lost state in parllel upon failures.

Ans 4:

4) What are the key phases in the execution pipeline for processing an IoT stream application?

Ans) Key phases of IoT stream application include

(i) Query Parsing and Optimization

The user code is parsed into logical execution plan represented by DAG → Vertices are stream operators
Edges are data flow

(ii) Operator replacement

DAG is converted to physical execution plan with multiple tasks in parallel. System places all operators on distributed edge nodes to minimize query latency and maximize throughput

(iii) Compute & shuffle
operators compute local data and shuffle results from one stage to next stage.

PHASE-2

User Code | Input Info & Config

Operator

Logical execution plan

Phase 1: Query Parsing and Optimization

Task/Instance

Physical execution plan

Phase 2: Operator Placement

Job scheduler

Central Data Warehouse

WAN/LAN

Edge Node

Shuffle

Phase 3: Compute and Shuffle