Course: Cloud Computing and Distributed Systems

Course Code: noc21-cs15

Session: 2020-21

Duration: 8 Weeks

Assessment procedures: Weekly Assignment (25%) + proctored certification Exam (75%)

Curriculum of the Course:

1. Introduction to Cloud Computing: Why Clouds, What is a Cloud, What's new in today's Clouds, Cloud computing vs. Distributed computing, Utility computing, Features of today's Clouds: Massive scale, AAS Classification: HaaS, IaaS, PaaS, SaaS, Data-intensive Computing, New Cloud Paradigms, Categories of Clouds: Private clouds, Public clouds
2. Virtualization: What's virtualization, Benefits of Virtualization, Virtualization Models: Bare metal, Hosted hypervisor
3. Types of Virtualizations: Processor virtualization, Memory virtualization, Full virtualization, Para virtualization, Device virtualization
4. Hotspot Mitigation for Virtual Machine Migration: Enterprise Data Centers, Data Center Workloads, Provisioning methods, Sandipiper Architecture, Resource provisioning, Black-box approach, Gray-box approach, Live VM Migration Stages, Hotspot Mitigation
5. Server Virtualization: Methods of virtualization: Using Docker, Using Linux containers, Approaches for Networking of VMs: Hardware approach: Single-root I/O virtualization (SR-IOV), Software approach: Open vSwitch, Mininet and its applications
6. Software Defined Network: Key ideas of SDN, Evolution of SDN, SDN challenges, Multi-tenant Data Centers: The challenges, Network virtualization, Case Study: VL2, NVP
7. Geo-distributed Cloud Data Centers: Inter-Data Center Networking, Data center interconnection techniques: MPLS, Google's B4 and Microsoft's Swan
8. Leader Election in Rings (Classical Distributed Algorithms): LeLann-Chang-Roberts (LCR) algorithm, The Hirschberg and Sinclair (HS) algorithm
9. Leader Election (Ring LE & Bully LE Algorithm): Leader Election Problem, Ring based leader election, Bully based leader election, Leader Election in Industry Systems: Google's Chubby and Apache Zookeeper
10. Design of Zookeeper: Race condition, Deadlock, Coordination, Zookeeper design goals, Data model, Zookeeper architecture, Sessions, States, Usecases, Operations, Access Control List (ACL), Zookeeper applications: Katta, Yahoo! Message Broker
11. Time and Clock Synchronization in Cloud Data Centers: Synchronization in the cloud, Key challenges, Clock Skew, Clock Drift, External and Internal clock synchronization, Christians algorithm, Error bounds, Network time protocol (NTP), Berkley's algorithm, Data center time protocol (DTP), Logical (or Lamport) ordering, Lamport timestamps, Vector timestamps
12. Global State and Snapshot Recording Algorithms: Global state, Issues in Recording a Global State, Model of Communication, Snapshot algorithm: Chandy-Lamport Algorithm

13. Distributed Mutual Exclusion: Mutual Exclusion in Cloud, Central algorithm, Ring-based Mutual Exclusion, Lamport's algorithm, Ricart-Agrawala's algorithm, Quorum-based Mutual Exclusion, Maekawa's algorithm, Problem of Deadlocks, Handling Deadlocks, Industry Mutual Exclusion: Chubby
14. Consensus in Cloud Computing and Paxos: Issues in consensus, Consensus in synchronous and asynchronous system, Paxos Algorithm
15. Byzantine Agreement: Agreement, Faults, Tolerance, Measuring Reliability and Performance, SLIs, SLOs, SLAs, TLAs, Byzantine failure, Byzantine Generals Problem, Lamport-Shostak-Pease Algorithm, Fischer-Lynch-Paterson (FLP) Impossibility
16. Failures & Recovery Approaches in Distributed Systems: Local checkpoint, Consistent states, Interaction with outside world, Messages, Domino effect, Problem of Livelock, Rollback recovery schemes, Checkpointing and Recovery Algorithms: Koo-Toueg Coordinated Checkpointing Algorithm
17. Design of Key-Value Stores: Key-value Abstraction, Key-value/NoSQL Data Model, Design of Apache Cassandra, Data Placement Strategies, Snitches, Writes, Bloom Filter, Compaction, Deletes, Read, Membership, CAP Theorem, Eventual Consistency, Consistency levels in Cassandra, Consistency Solutions
18. Design of HBase: What is HBase, HBase Architecture, Components, Data model, Storage Hierarchy, Cross-Data center Replication, Auto Sharding and Distribution, Bloom Filter, Fold, Store, and Shift
19. Peer to Peer Systems in Cloud Computing: Napster, Gnutella, FastTrack, BitTorrent, DHT, Chord, Pastry and Kelips.
20. MapReduce: Paradigm, Programming Model, Applications, Scheduling, Fault-Tolerance, Implementation Overview, Examples
21. Introduction to Spark: Resilient Distributed Datasets (RDDs), RDD Operations, Spark applications: Page Rank Algorithm, GraphX, GraphX API, GraphX working
22. 3.Introduction to Kafka: What is Kafka, Use cases for Kafka, Data model, Architecture, Types of messaging systems, Importance of brokers

List of students enrolled

S. No	Name of Student
1	Chhavi Jangid
2	Lalit Kumawat
3	Naman Joshi
4	Prashu Jain
5	Sarthak Sharma
6	Shruti Dubey