☁ Cloud Computing Questions

1. **Why SaaS is ideal for non-technical users?**
    - a) It requires Linux knowledge
    - b) It's free software
    - **c) It offers ready-to-use apps over the internet (ans)**
    - d) It requires cloud programming

---

2. **Storage virtualization is primarily used to:**
    - a) Increase CPU performance
    - **b) Combine multiple storage devices into a single logical view (ans)**
    - c) Reduce network latency
    - d) Enhance data encryption speed

---

3. **Which of the following best defines storage virtualization?**
    - a) Creating multiple physical storage devices from one logical device
    - **b) Masking the physical storage resources to appear as a single storage unit (ans)**
    - c) Compressing data for efficient transmission
    - d) Storing data only in the cloud

---

4. **What PaaS allows developers to do?**
    - a) Write kernel modules

- **b) Build and deploy apps without managing infrastructure (ans)**
- c) Manually manage cloud servers
- d) Patch firewalls

---

5. **Which of the following is different between IaaS and SaaS?**
   - a) Both offer only data storage
   - **b) IaaS gives infrastructure; SaaS gives end-user software (ans)**
   - c) SaaS is offline; IaaS is online only
   - d) IaaS offers databases directly

---

6. **Outline which steps is involved in monitoring cloud resources.**
   - a) Logging out users
   - **b) Checking dashboards, alerts, usage metrics (ans)**
   - c) Turning off VMs randomly
   - d) Installing on-premise servers

---

7. **Choose how utility computing is linked to cloud pricing.**
   - a) Monthly fixed payment
   - **b) Pay-per-use based on resource consumption (ans)**
   - c) Free-tier use
   - d) Unlimited subscriptions

---

8. **[What is] the goal of elastic computing in cloud environments.**
    - a) Make systems rigid
    - b) Reduce bandwidth
    - **c) Adjust computing power dynamically (ans)**
    - d) Boost energy usage

---

9. **Which is the purpose of a resource management system?**
    - a) Stores passwords
    - **b) Allocates and optimizes use of cloud resources (ans)**
    - c) Encrypts emails
    - d) Deletes files

---

10. **Choose the correct classification of cloud computing platforms based on their function.**
    - a) SaaS for infrastructure, PaaS for storage, IaaS for games
    - **b) IaaS for infrastructure, PaaS for application development, SaaS for end-user software (ans)**
    - c) PaaS for end-user apps, SaaS for virtualization, IaaS for databases
    - d) IaaS for coding tools, SaaS for networking, PaaS for emails
11. **What is the SLA in a cloud context?**
    - a) Server Log App
    - b) Security Log Audit

- **c) Agreement on service quality and availability (ans)**
- d) System Language Adapter

---

12. **SLA management is vital for enterprises because its .**
    - a) Avoids vendor lock-in
    - **b) Ensures performance guarantees, uptime, and penalties (ans)**
    - c) Allows free internet
    - d) Tracks customer IDs

---

13. **Tell the function of remote administration in the cloud.**
    - **a) Control systems from any location via network tools (ans)**
    - b) Manually open data centers
    - c) Run machines offline
    - d) Turn off firewalls

---

14. **Tell the role of a data management system in cloud computing.**
    - a) Manages fonts
    - b) Tracks disk errors
    - **c) Handles storage, backup, replication, and access (ans)**
    - d) Deletes old apps

---

15. **What is the different between monitoring and administration?**
    - a) Both are unrelated
    - **b) Monitoring = observe; Admin = configure & control (ans)**
    - c) Admin only backs up data
    - d) Monitoring is always offline

---

16. **Outline key limitations of cloud computing.**
    - a) Unlimited bandwidth
    - b) Manual scaling only
    - **c) Data security, internet dependency, vendor lock-in (ans)**
    - d) Free lifetime use

📊 Big Data, Hadoop & NoSQL Questions

17. **Which of the following is NOT a common "V" of Big Data?**
    - a) Volume
    - b) Velocity
    - c) Veracity
    - **d) Validity (ans)**

---

18. **What is the primary purpose of Hadoop Distributed File System (HDFS)?**

- a) To manage relational databases
- **b) To provide a distributed, fault-tolerant storage system for large files (ans)**
- c) To process real-time streaming data
- d) To run SQL queries on structured data

---

19. **Which of the following is a key characteristic of NoSQL databases compared to traditional relational databases?**
    - a) Strict schema enforcement
    - **b) Horizontal scalability and flexible schemas (ans)**
    - c) ACID compliance for all operations
    - d) Primarily used for small, structured datasets

---

20. **Google File System (GFS) is a precursor to which open-source distributed file system?**
    - a) Apache Cassandra
    - b) Apache Spark
    - **c) Hadoop Distributed File System (HDFS) (ans)**
    - d) MongoDB

---

21. **Which of the following is a common issue encountered when handling large datasets (Big Data)?**
    - a) Limited storage capacity on a single machine
    - b) Difficulty in processing and analysing data in a timely manner

- c) Challenges in data governance and security
- **d) All of the above (ans)**

---

22. **Which type of NoSQL database stores data in key-value pairs?**
    - a) Document-oriented
    - b) Graph
    - **c) Key-value (ans)**
    - d) Column-family
23. **What is a "big data spreadsheet" often used for?**
    - a) Storing and analyzing small, structured datasets on a single machine
    - **b) Providing a visual interface for exploring and manipulating large datasets, often backed by distributed systems (ans)**
    - c) Performing complex statistical modeling on real-time data streams
    - d) Managing traditional relational database tables

---

24. **Which of the following scenario may not be a good fit for HDFS in the Big Data paradigm?**
    - a) HDFS is not suitable for scenarios requiring multiple/simultaneous writes to the same file
    - **b) HDFS is suitable for storing data related to applications requiring low latency data access (ans)**

- c) HDFS is suitable for storing data related to applications requiring high latency data access
- d) None of the above

---

25. **Hadoop (a big data tool) works with a number of related tools. Choose from the following the common tools included in Hadoop:**
    - a) MySQL, Google API and MapReduce
    - b) MapReduce, Scala and Hummer
    - **c) MapReduce, HBase and Hive (ans)**
    - d) MapReduce, Hummer and Heron

---

26. **The data node and name node in HADOOP are:**
    - **a) Worker Node and Master Node respectively (ans)**
    - b) Both Worker Nodes
    - c) other Master Nodes
    - d) None of the above

---

27. **Which of the following is a component of Hadoop?**
    - a) YARN
    - b) HDFS
    - c) MapReduce
    - **d) All of the options (ans)**

---

28. **Which of the following type of virtualization is found in hypervisors such as Microsoft's Hyper-V?**

- a) Paravirtualization
- **b) Full virtualization (ans)**
- c) Emulation
- d) None of the mentioned

---

29. **What is HDFS Block in Hadoop?**

- **a) It is the logical representation of data (ans)**
- b) It is the physical representation of data
- c) Both of the above
- d) None of the above

---

30. **What is HDFS?**

- **a) Storage layer (ans)**
- b) Batch processing engine
- c) Resource management layer
- d) None of the above

---

31. **Which of the following statement/s is/are true? (i) Facebook has the world's largest Hadoop cluster. (ii) Hadoop 2.0 allows live stream processing of real-time data**

- **a) Neither (i) nor (ii) (ans)**
- b) Both (i) and (ii)
- c) (i) only

- d) (ii) only

---

32. **What is the term used for a collection of large, complex data sets that cannot be processed using traditional data processing tools?**
    - **a) Big Data (ans)**
    - b) Small Data
    - c) Mini Data
    - d) Medium Data

---

33. **What is the process of transforming structured and unstructured data into a format that can be easily analyzed?**
    - a) Data Mining
    - **b) Data Processing (ans)**
    - c) Data Integration
    - d) Data Warehousing

---

34. **Which of the following is a tool used for processing and analyzing Big Data?**
    - a) PostgreSQL
    - **b) Hadoop (ans)**
    - c) MySQL
    - d) Oracle

---

35. **Which of the following is not a common challenge associated with Big Data?**

- a) Data Quality
- b) Data Integration
- c) Data Privacy
- **d) Data Duplication (ans)**

---

36. **Apply the map-reduce model to count words in a document. What phase comes first?**

- a) Reduce
- b) Filter
- **c) Map (ans)**
- d) Shuffle

---

37. **Calculate the total HDFS storage needed for a 1GB file with 3 replicas.**

- a) 1GB
- b) 2GB
- **c) 3GB (ans)**
- d) 4GB

---

38. **Infer how parallelism helps speed up matrix multiplication.**

- a) Decreases speed
- **b) Processes parts simultaneously (ans)**

- c) Needs more memory
- d) Adds complexity only

---

39. **Choose the most fault-tolerant model to solve a big data storage problem.**
    - a) NoSQL
    - b) RDBMS
    - **c) HDFS (ans)**
    - d) CSV

---

40. **Use HDFS when .**
    - a) Files are small
    - b) Real-time response is critical
    - **c) Batch processing is needed (ans)**
    - d) RAM is high

---

41. **Choose the best approach to implement a distributed solution for sorting 1TB of data.**
    - a) Single-threaded sorting
    - **b) HDFS + MapReduce (ans)**
    - c) Python loop
    - d) Manual indexing

---

42. **Choose the principle used to operate a distributed system.**

- a) Centralization
- **b) Redundancy (ans)**
- c) Single thread
- d) Static IP

---

43. **Choose the two key components of the HDFS (Hadoop Distributed File System) architecture.**
    - **a) NameNode and DataNode (ans)**
    - b) Server and Router
    - c) Block and Bit
    - d) Heap and Stack

---

44. **Interpret the NoSQL data model from a JSON structure. What does it represent?**
    - a) Array
    - **b) Key-value pair (ans)** (More accurately, a document, which is a collection of key-value pairs)
    - c) Columnar
    - d) Tuple

---

45. **Illustrate the difference between GFS and HDFS.**
    - a) HDFS is for Google
    - b) GFS is Hadoop's
    - **c) GFS is proprietary, HDFS is open-source (ans)**
    - d) Both are same

---

46. **Choose what is improved by applying the concept of sharding in NoSQL databases.**

- a) Improves data encryption
- b) Improves user interface design
- **c) Improves query performance and scalability (ans)**
- d) Schema rigidity

---

47. **Calculate replication factor storage: File = 500MB, RF = 2. Total?**

- a) 250MB
- b) 500MB
- **c) 1GB (ans)**
- d) 2GB

---

48. **Which of the following is a big data spreadsheet that handles 10 million rows?**

- a) Crashes
- **b) Loads slowly (ans)** (Or crashes; traditional spreadsheets are not built for this scale.)
- c) Uses chunked memory model
- d) Requires reboot

---

49. **Which is sollution of the bottleneck problem in distributed programming?**

- a) Increase single-thread load
- **b) Use parallel tasks (ans)**
- c) Use one loop
- d) Disable threads

---

50. **Use Hadoop for processing which type of data?**
    - a) Small binary logs
    - **b) Massive structured data (ans)** (and unstructured/semi-structured)
    - c) HTML content only
    - d) Temporary RAM cache

---

51. **Implement parallelism in data filtering using...**
    - a) Manual typing
    - b) SQL only
    - **c) MapReduce (ans)**
    - d) Email alerts

---

52. **Choose the feature used to operate a fault-tolerant distributed storage system.**
    - a) Encryption only
    - **b) Data replication (ans)**
    - c) Manual backup
    - d) Data compression

---

53. **Choose the last phase in a high-level workflow of the MapReduce programming model.**

- a) Map
- b) Shuffle
- **c) Reduce (ans)**
- d) Cleanup

---

54. **Interpret the error message in HDFS when DataNode fails.**

- a) Backup enabled
- b) Data loss
- **c) Data replicated (ans)** (This isn't an error message, but the system's *response* is to use a replica.)
- d) File deleted

---

55. **Illustrate how big data affects traditional storage methods.**

- a) Makes it easier
- b) Slows down access
- **c) Requires distributed models (ans)**
- d) Reduces RAM

---

56. **Apply a hash function in a NoSQL DB for .**

- **a) Indexing (ans)** (Especially for distributing data in sharding)
- b) Validation
- c) Replication
- d) Sorting

---

57. **Calculate map task distribution for 1GB file across 4 nodes.**
    - **a) 4 tasks (ans)** (This is a simplified answer; it depends on the block size, but multiple tasks are definite.)
    - b) 2 tasks
    - c) 1 task
    - d) 8 tasks

---

58. **Scalability using distributed file systems as .**
    - a) Less data allowed
    - **b) Adds more nodes easily (ans)**
    - c) Fewer users supported
    - d) Manual sync

---

59. **Tell slow write problems in big data NoSQL.**
    - a) Use SSD
    - **b) Use indexes (ans)** (This is a *cause* of slow writes, as indexes must be updated.)
    - c) Use write buffers

- d) Disable write-ahead log

---

60. **Use BigQuery or similar tools to process .**
    - a) 10 rows
    - b) 100 rows
    - **c) Petabyte-scale data (ans)**
    - d) KB logs only

---

61. **Implement a reducer in Hadoop that counts frequencies. What does it emit?**
    - a) Input key
    - **b) Sum of values (ans)**
    - c) File name
    - d) Memory size

---

62. **Choose the correct use case to operate columnar storage for .**
    - a) Fast random read
    - b) Reduced redundancy
    - **c) Fast analytics (ans)**
    - d) Faster email

---

63. **Tell which option represents parallel write architecture in distributed file systems.**
    - a) Sequential blocks

- **b) Parallel disk writes (ans)**
- c) Only NameNode writes
- d) FIFO writes

---

64. **Interpret Hadoop job failure. What might be the cause?**
    - a) Browser crash
    - **b) Mapper issue (ans)**
    - c) RAM upgrade
    - d) HDD replaced

---

65. **Choose the way for NoSQL handles schema changes.**
    - a) Fixed schema
    - b) No schema at all
    - **c) Flexible schema (ans)**
    - d) Uses CSV only

---

66. **Apply replication in GFS for .**
    - a) Performance
    - b) Security
    - **c) Fault tolerance (ans)**
    - d) Compression

---

67. **Calculate which value is storage of overhead in HDFS with 3x replication for 2GB file.**

- **a) 4GB (ans)** (Total = 2GB * 3 = 6GB. Overhead = Total - Original = 6GB - 2GB = 4GB)
- b) 6GB
- c) 2GB
- d) 3GB

---

68. **Choose the role of NameNode in HDFS.**

- **a) Stores metadata (ans)**
- b) Stores actual data
- c) Compresses file
- d) Schedules jobs

---

69. **Use Spark instead of MapReduce when .**

- a) You want slower performance
- **b) You want real-time processing (ans)** (or faster in-memory batch processing)
- c) Batch-only is enough
- d) You use spreadsheets

---

70. **Implement big data analysis on...**

- a) CPU-only system
- b) Local desktop
- **c) Clustered architecture (ans)**

- d) XML parser

---

71. **Choose the purpose to operate distributed cache for purpose.**
    - a) Executing batch jobs
    - **b) Reducing database load and improving performance (ans)**
    - c) File sorting
    - d) Compression

---

72. **Multiple mappers work in Hadoop as .**
    - a) Single-thread
    - **b) Parallel threads (ans)** (or processes, depending on the node)
    - c) Serial blocks
    - d) RAM-only

---

73. **Choose the correct interpretation: log reports from a Hadoop job where high Map time means .**
    - a) Efficiency
    - **b) Inefficient mapping or data skew (ans)**
    - c) Success
    - d) Disk error

---

74. **In the context of big data stores, which of the following best demonstrates how partitioning improves performance and scalability in distributed systems?**

- a) By compressing large datasets to reduce storage requirements
- b) By duplicating data across multiple servers to ensure redundancy
- **c) By dividing a large dataset into smaller, manageable segments distributed across different nodes (ans)**
- d) By encrypting data to enhance security during transmission