

CS219 Midterm #2, Winter 2021

Name: _____

Student ID: _____

I. Multiple Choices: Select *all* the answers that you think are correct (3 points each; 51 points)

1. Which statement is correct with Chubby? Your answer: _____
(A) Chubby is intended to provide fine-grained locking; (B) Chubby clients can subscribe to a range of events when they create a handle; (C) While file data or meta-data is to be changed, master sends invalidations for the data to each client that may have cached it while blocking the modification; (D) A Chubby session is maintained by both periodic handshakes called KeepAlives and a range of events.
2. Which statement is correct with Google Spanner? Your answer: _____
(A) Fine-grained replication configuration for data is automatically controlled by the Spanner database system rather than by applications; (B) TrueTime API uses both GPS and atomic clock as the time reference but does not expose clock uncertainty directly; (C) Spanner's data model is not purely relational since each row may not have names; (D) Its implementation supports both read-write transactions and read-only transactions through the globally meaningful commit timestamp.
3. Which statement is correct with Facebook Haystack? Your answer: _____
(A) Post-CDN caching is still effective since a request that misses in the CDN will still hit in the internal Facebook cache. (B) The Haystack Directory load balances reads across logical volumes and writes across physical volumes; (C) The mapping of photo IDs to the file-system metadata (i.e., file, offset and size in bytes) critical for retrieving the photo is kept in the hard disk by the Haystack store; (D) Each Store machines maintain an index file, which serves as a checkpoint of data structures used to locate needles efficiently on disk.
4. Which technique is used by Dynamo to achieve high availability for writes? Your answer: _____
(A) Merkle trees; (B) consistent hashing; (C) slopped quorum and hinted handoff; (D) vector clock with reconciliation during reads.
5. What techniques are NOT used by Azure storage? Your answer: _____
(A) There are two replication engines at two different layers of Stream and Partition; (B) A new stream can be constructed by concatenating extents from existing streams; (C) metadata and commit log are the streams that the partition layer can read randomly from any starting point to the last record of a stream; (D) For an extent, every append is replicated three times across the extent's replicas.

6. What statements are true with Facebook's f4 storage system? Your answer: _____
 (A) The request rate for the content of f4 is higher than that for content in Haystack; (B) Each volume has a data file, an index file and a journal file; (C) Individual f4 cells are prone to disk, host and rack failures, although geo-replicating XOR volumes brings tolerance to data center failures; (D) f4 encodes data to lower replication to 2.1x while Haystack may use 3.6x.

7. What is true with Megastore? Your answer: _____
 (A) It supports wide-area replication and scalability, but not consistent view of data; (B) It uses Paxos to replicate primary user's data across datacenters on every read and write; (C) It uses two-phase commit for atomic commitment; (D) ACID semantics are ensured across entity groups.

8. What technique helps to better scale the parameter server? Your answer: _____
 (A) It supports independent parameter namespace, thus allowing a worker group to isolate its shared parameters from others; (B) identical consistency model is provided to hide latency and synchronization cost; (C) efficient asynchronous communication is used; (D) Any form of vector clock implementation can already scale to track status of a large number of nodes and parameters after machine failures and partitions.

9. What techniques are used by Petuum to support both data-parallel and model-parallel machine learning algorithms? Your answer: _____
 (A) Bounded asynchronous consistency model helps to ensure data-parallel convergence guarantees without explicit network synchronization; (B) Global model state can be accessed via distributed shared-memory interface; (C) Workers have to access the data in the batch order; (D) scheduler enables user-fined application consistency rules, thus permitting fine-grained control over parallel ordering of model-parallel updates.

10. What techniques are used to build the scalable and fault-tolerant platform of Pregel? Your answer: _____
 (A) It leverages the bulk synchronous parallel model with a sequence of super-steps; (B) It uses an edge-centric approach to reduce synchronization and communication cost; (C) The vertices can compute in parallel, with each running the same user-defined function; (D) It expresses the graph algorithm as a chained MapReduce innocations to reduce the communication and associated serialization cost.

11. What techniques are used by DistBelief system? Your answer: _____
 (A) It uses the parameter server architecture to enable model parallelism via multithreading within a machine and through message passing across machines, while hiding the detailed synchronization and communication from the programmer; (B) Both online Downpour SGD and batch Sandblaster L-BFGS have similar speed compared with conventional SGD and L-BFGS. (C) MapReduce is also effective for deep network training, similar to DistBelief; (D) Both optimization algorithms of Downpour SGD and Sanblaster L-BFGS use data parallelism to concurrently process different training samples in each model replica, and then combine the results to optimize objective function.

12. What are the solutions used by TensorFlow? Your answer: _____
 (A) It offers flexible and different dataflow-based programming abstractions to adapt applications on heterogeneous hardware platforms; (B) It does not support multiple concurrent executions on overlapping subgraphs; (C) Mutable states are owned by individual vertices, and cannot be shared between different executions of the graph; (D) It uses user-defined checkpointing to ensure fault tolerance.
13. What are used by Adam system? Your answer: _____
 (A) It optimizes and balances both computation and communication for deep learning training; (B) It decouples durability from the update processing path to boost throughput serving to training nodes; (C) Only two copies of each parameter shard are stored on different parameter servers; (D) It minimizes memory copies of data by using pointers to pass data to neurons.
14. What are addressed by ByteScheduler? Your answer: _____
 (A) Communication-computation dependency; (B) Tensor partitioning; (C) heterogenous runtime environment; (D) slow machines.
15. What ideas are exploited in BytePS? Your answer: _____
 (A) It uses summation service to remove the bottlenecks of existing parameter server design; (B) It uses priority scheduling together with pipelining; (C) It addresses RDMA related issues; (D) All reduce, which aggregates every GPU's gradients before GPUs update their own local parameters, is never optimal in any setup, compared with parameter server.
16. What are adopted by Gavel? Your answer: _____
 (A) Many scheduling policies can be formulated as optimization problems whose objective is a job's achieved throughput; (B) round-based scheduling mechanism to ensure allocations returned by polies; (C) It arbitrates cluster accelerators among jobs to optimize the cluster objective using a two-step process; (D) It supports hierarchical scheduling policies and solves hierarchical objectives via water filling.
17. What are used by HiveD? Your answer: _____
 (A) Buddy cell allocation handles dynamic binding and ensures safety in response to dynamic workloads; (B) It defines a hierarchy of multi-level cell structures; (C) High-priority jobs cannot preempt low-priority jobs to ensure sharing safety; (D) It never proactively binds to a faulty physical cell. Therefore, third-party scheduler in the VC cannot see the faulty hardware and avoid using GPUs in the cell.

II. (21 points) You are asked to compare four cloud storage systems, Google BigTable, Amazon Dynamo, Facebook Haystack, and Microsoft Azure.

(A) (4 points) Are these storage systems supporting the same data models? Describe the specific data models these systems support.

(B) (6 points) Which consistency model does each of the above system support? Justify your answer.

(C) (6 points) What are the differences in their design (Google, Amazon, Haystack, and Azure) to improve system availability?

(D) (5 points) Can you ever achieve both availability and consistency at the same time? If your answer is Yes, Under what conditions? If your answer is No, justify your answer.

1. Describe at least two ideas that achieve transaction properties in cloud storage systems.
2. Describe two differences when using Chubby and the Internet DNS as a name service.
3. Which approach works better to scale the design in the cloud storage system, scale up (increasing the depth of the hierarchy) or scale out (adding more elements in the existing hierarchy)? Briefly justify your answer.
4. Can you describe the main ideas in concurrency control when providing relational database like transaction properties on top of the cloud key-value storage system?

- Can you identify **two** scenarios where Paxos is used to ensure consistent data replications in cloud systems? Can you also identify a scenario where two-phase commit protocol can be used to ensure atomic commitment in cloud systems?
- Does Facebook need both Haystack and F4 as the storage systems, or one will replace the other eventually? Identify the conditions when the given system (i.e., Haystack or F4) is used.
- What is the main limitation of Megastore compared with Spanner? How does Spanner's design address the limitations of Megastore?