Big Data L4: Distributed File Systems True/False Questions

#	Question	Answer
151	A hash function must always be pre-image resistant and collision resistant.	True
152	A small change in the input of a good hash function should cause a large, unpredictable change in output.	True
153	RAID distributes storage across multiple computers in a network.	False
154	A RAID 1 setup increases reliability by mirroring data across multiple disks.	True
155	RAID 5 provides full redundancy at the cost of halving the storage capacity.	False
156	RAID uses parity bits to recover lost data in case of single disk failure.	True
157	In RAID, striping without redundancy scales capacity but sacrifices fault tolerance.	True
158	Distributed file systems were invented mainly to improve disk capacity on a single machine.	False
159	Extreme replication (copying all data to all nodes) is an efficient storage solution.	False
160	In HDFS, data is divided into blocks, each typically up to 128MB in size.	True
161	In HDFS, the Name Node stores both data and file contents.	False
162	The Name Node in HDFS stores only metadata about file locations and block mappings.	True
163	Data Nodes in HDFS store metadata like filenames and directory structures.	False
164	A Data Node in HDFS manages physical storage and maintains checksums.	True
165	Failure of the HDFS Name Node is recoverable with simple replication.	False
166	Data Node failure in HDFS is tolerated as long as enough replicas of blocks exist.	True
167	In HDFS, data blocks can be appended to existing files.	True
168	HDFS strictly follows POSIX compliance for filesystem operations.	False
169	The HDFS architecture favors a write-once, read-many access pattern.	True
170	In HDFS, the replication factor directly impacts fault tolerance.	True
171	HDFS minimizes network communication by trying to schedule tasks close to where data resides.	True
172	Each file in HDFS is stored as a single large block without partitioning.	False
173	Clients in HDFS directly send write requests to the Data Nodes without consulting the Name Node.	False

#	Question	Answer
174	The heartbeat mechanism in HDFS is used by Data Nodes to signal their availability to the Name Node.	True
175	In HDFS, the Name Node uses checkpoints to periodically back up its metadata state.	True
176	Append-only operations in HDFS help simplify consistency and replication.	True
177	Files written to HDFS can be modified in-place after the initial write.	False
178	HDFS is primarily optimized for storing many small files.	False
179	In HDFS, storing fewer large files is preferred over storing many small files.	True
180	MapReduce and HDFS are completely independent and cannot be used together.	False
181	In the CAP theorem, HDFS sacrifices availability for strong consistency and partition tolerance.	False
182	The HDFS Name Node ensures consistent namespace management in the distributed file system.	True
183	If the HDFS Name Node fails, the system can continue serving client requests as long as the Data Nodes are alive.	False
184	HDFS is resilient to network partitioning up to the replication factor.	True
185	HDFS guarantees that once data is written, it cannot be changed later.	True
186	Communication cost is a major consideration in the design of distributed file systems.	True
187	Data locality in HDFS means moving computation closer to where data is stored rather than moving data.	True
188	The CAP theorem states that consistency, availability, and partition tolerance can all be fully achieved simultaneously.	False
189	When using HDFS, the Name Node always knows which Data Nodes are still alive via periodic heartbeats.	True
190	HDFS replicates both the metadata and the actual data blocks across multiple nodes.	False
191	If a block is lost in HDFS but two replicas remain, the system automatically recovers the missing block.	True
192	The typical maximum block size in HDFS is 8 KB to handle small file updates.	False
193	RAID 5 can recover from the loss of any one disk through parity computations.	True
194	In HDFS, the "checksum" attached to a data block ensures correctness of the block's contents.	True
195	HDFS is designed to support highly concurrent small write operations from many users.	False
196	"Striping" in RAID refers to distributing parts of a file across multiple disks for speed.	True
197	In a distributed file system, minimizing communication overhead is often more important than minimizing storage usage.	True
198	HDFS was designed primarily for interactive computing with frequent file edits.	False

#	Question	Answer
199	RAID 1 can improve read performance because multiple disks can serve read requests in parallel.	True
200	HDFS embraces a "write-once, read-many" principle to simplify data management and improve scalability.	True