Big Data L1: True/False Questions

#	Question	Answer
1	Big Data is primarily about using the trendiest new tools.	False
2	The main focus of Big Data is solving scaling issues in storage, computation, and communication.	True
3	If you can't fit your dataset on a single laptop, it can operationally be considered "Big Data."	True
4	The CS view of data treats data as immutable, unchangeable "givens."	False
5	The DS view of data considers bits and bytes as something to be efficiently moved and transformed.	False
6	More data improves parameter estimation and recall.	True
7	The five "V"s of Big Data are Volume, Velocity, Variety, Veracity, and Value.	True
8	A huge amount of data automatically guarantees accurate insights without additional effort.	False
9	High dimensionality (more features/columns) can actually bless data science models by enabling richer analysis.	True
10	In Big Data, "Value" refers to how easy it is to store data.	False
11	A single ant can easily carry a Colocasia gigantea leaf without help.	False
12	Facebook generates petabytes of new data every month.	False
13	In scaling, more workers always linearly speed up completion for complex projects like building an operating system.	False
14	Big Data challenges often require coordination among multiple machines.	True
15	Increasing CPU clock speeds has remained the primary driver of computation improvements recently.	False
16	Communication overhead can significantly hurt the scaling of complex tasks.	True
17	Main memory access (RAM) is often the first performance bottleneck in data processing.	True
18	Storing 540TB of data costs less today than it did to store 1MB of data in the 1980s.	True
19	In Big Data, the cost of storage is usually the main bottleneck.	False
20	Big Data strategies suggest distributing processing and minimizing communication.	True
21	Pre-fetching and pipelining techniques aim to accelerate sequential memory reads.	True
22	Adding more developers to a late software project will always speed it up.	False

#	Question	Answer
23	A professional software engineer writes about 100,000 production-quality lines of code per year.	False
24	A modern operating system can contain up to 100 million lines of code.	True
25	Tasks like collecting seashells are easy to parallelize.	True
26	Building a cathedral is a task that scales linearly with more workers.	False
27	Communication demands in team projects scale linearly with the number of people.	False
28	"More oxen, not bigger oxen" reflects the philosophy behind scaling in Big Data.	True
29	CPUs are getting infinitely faster every year because of Moore's Law.	False
30	The latency to access SSD storage is significantly higher than RAM.	True
31	Veracity in Big Data refers to the speed at which new data arrives.	False
32	A single computer can easily process a YouTube-scale data stream without distributed systems.	False
33	Storage prices per GB have gone down significantly over the past decades.	True
34	Minimizing RAM-to-CPU communication improves overall system performance.	True
35	YouTube uploads 500 minutes of video every hour.	False
36	Communication latency has improved faster than CPU speeds.	False
37	Minimizing unnecessary data movement is a critical principle for scaling Big Data systems.	True
38	Increasing the number of features (columns) in a dataset often worsens statistical model performance.	False
39	A GPU can help in parallel computation tasks better than a traditional CPU in some cases.	True
40	Hadoop and Spark are both tools used for distributed computation.	True
41	In Big Data systems, local network bandwidth is usually faster than RAM access.	False
42	Dremel and Parquet are associated with column-oriented storage formats.	True
43	Large teams can speed up software development by simply dividing the work into independent blocks.	False
44	Collecting large-scale noisy data guarantees better machine learning model performance.	False
45	In the context of Big Data, storage capacity is no longer the primary scaling challenge.	True
46	Increases in data storage speed have kept pace with increases in CPU performance.	False
47	A well-structured distributed system seeks to minimize global coordination where possible.	True
48	The Pyramid of Giza was built using perfect parallelism without communication overhead.	False

#	Question	Answer
49	Distributed computing often requires a shared coordination infrastructure (like a "yoke" in oxen).	True
50	In Big Data, bigger computers are always a better solution than more computers.	False