Latency for Common Operations

OPERATION	NOTE	LATENCY	SCALED LATENCY
L1 cache reference	Level-1 cache, usually built onto the microprocessor chip itself.	0.5 ns	Consider L1 cache reference duration is 1 sec
Branch mispredict	During the execution of a program, CPU predicts the next set of instructions. Branch misprediction is when it makes the wrong prediction. Hence, the previous prediction has to be erased and new one calculated and placed on the execution stack.	5 ns	10 s
L2 cache reference	Level-2 cache is memory built on a separate chip.	7 ns	14 s
Mutex lock/unlock	Simple synchronization method used to ensure exclusive access to resources shared between many threads.	25 ns	50 s
Main memory reference	Time to reference main memory i.e. RAM.	100 ns	3m 20s
Compress 1K bytes with Snappy	Snappy is a fast data compression and decompression library written in C++ by Google and used in many Google projects like BigTable, MapReduce and other open source projects.	3,000 ns	1h 40 m
Send 1K bytes over 1 Gbps network		10,000 ns	5h 33m 20s
Read 1 MB sequentially from memory	Read from RAM.	250,000 ns	5d 18h 53m 20s

Round trip within same datacenter	We can assume that the DNS lookup will be much faster within a data center than it is to go over an external router.	500,000 ns	11d 13h 46m 40s
Read 1 MB sequentially from SSD disk	Assumes SSD disk. SSD boasts random data access times of 100000 ns or less.	100,000 ns	23d 3h 33m 20s
Disk seek	Disk seek is the method to get to the sector and head in the disk where the required data exists.	10,000,000 ns	231d 11h 33m 20s
Read 1 MB sequentially from disk	Assumes regular disk, not SSD. Check the difference in comparison to SSD!	20,000,000 ns	462d 23h 6m 40s
Send packet CA->Netherlands->CA	Round trip for packet data from U.S.A to Europe and back.	150,000,000 ns	3472d 5h 20m