Table 1: Description of the 34 Spark configuration parameters.

Configuration Parameters—Description	Range	Default
spark.reducer.maxSizeInFlight—Maximum size of map outputs to fetch simultaneously from each reduce task, in MB.	2–128	48
spark.shuffle.file.buffer—Size of the in-memory buffer for each shuffle file output stream, in KB.	2–128	32
spark.shuffle.sort.bypassMergeThreshold—Avoid merge-sorting data if there is no map-side aggregation.	100–1000	200
spark.speculation.interval—How often Spark will check for tasks to speculate, in millisecond.	10–100	100
spark.speculation.multiplier—How many times slower a task is than the median to be considered for speculation.	1–5	1.5
spark.speculation.quantile—Percentage of tasks which must be complete before speculation is enabled.	0–1	0.75
spark.broadcast.blockSize—Size of each piece of a block for TorrentBroadcastFactory, in MB.	2–128	4
spark.io.compression.codec—The codec used to compress internal data such as RDD partitions, and so on.	snappy, lzf,lz4	snappy
spark.io.compression.lz4.blockSize—Block size used in LZ4 compression, in KB.	2–128	32
spark.io.compression.snappy.blockSize—Block size used in snappy, in KB.	2–128	32
spark.kryo.referenceTracking—Whether to track references to the same object when serializing data with Kryo	true,false	true
spark.kryoserializer.buffer.max—Maximum allowable size of Kryo serialization buffer, in MB.	8–128	64
spark.kryoserializer.buffer—Initial size of Kryo's serialization buffer, in KB.	2–128	64
spark.driver.cores—Number of cores to use for the driver process.	1–30	1
spark.executor.cores—The number of cores to use on each executor.	4–30	core #
spark.executor.instances—The number of executors for static allocation.	6–10	2
spark.driver.memory—Amount of memory to use for the driver process, in MB.	1024–36864	1024
spark.executor.memory—Amount of memory to use per executor process, in MB.	7168–36864	1024
spark.storage.memoryMapThreshold—Size of a block above which Spark maps when reading a block from disk, in MB.	50–500	2
spark.network.timeout—Default timeout for all network interactions, in second.	20–500	120
spark.locality.wait—How long to launch a data-local task before giving up, in second.	1–10	3
spark.scheduler.revive.interval—The interval length for the scheduler to revive the worker resource, in second.	2–50	1
spark.task.maxFailures—Number of task failures before giving up on the job.	1–8	4
spark.shuffle.compress—Whether to compress map output files.	true,false	true
spark.memory.fraction—Fraction of (heap space - 300 MB) used for execution and storage.	0.5–1	0.75
spark.shuffle.spill.compress—Whether to compress data spilled during shuffles.	true,false	true
spark.speculation—If set to "true", performs speculative execution of tasks.	true,false	false
spark.broadcast.compress—Whether to compress broadcast variables before sending them. Generally a good idea.	true,false	true
spark.rdd.compress—Whether to compress serialized RDD partitions.	true,false	false
spark.serializer—Class to use for serializing objects that are sent over the network or need to be cached in serialized form.	java,kryo	java
spark.memory.storageFraction—Amount of storage memory immune to eviction, expressed as a fraction of the size of the region set aside by *spark.memory.fraction*.	0.5–1	0.5
spark.default.parallelism—The largest number of partitions in a parent RDD for distributed shuffle operations.	8–50	#
spark.memory.offHeap.enabled—If true, Spark will attempt to use off-heap memory for certain operations.	true,false	false
spark.memory.offHeap.size—The absolute amount of memory which can be used for off-heap allocation, in MB.	10–1000	0