

60 Days of Apache Spark Series

A Comprehensive Guide to Apache Spark

Charchit Patidar

Serial No.	Topic	Link
1	Create a Spark RDD using Parallelize	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7010137498794307584-bnIP?utm_source=share&utm_medium=member_desktop
2	Read Text files into single RDD	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7010144664242585600-e063?utm_source=share&utm_medium=member_desktop
3	Read/Load CSV File into RDD and Data frame.	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-learning-like-activity-7010149084934152192-RMx8?utm_source=share&utm_medium=member_desktop
4	Repartition() vs Coalesce() in Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-learning-like-activity-7010235933048410112-r84I?utm_source=share&utm_medium=member_desktop
5	Dynamic Shuffle Partitions in Spark SQL	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7010278382391214080-GvL2?utm_source=share&utm_medium=member_desktop
6	Broadcast Variables in Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7010605955298050048-eHmW?utm_source=share&utm_medium=member_desktop
7	Caching Spark DataFrame — How & When	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7011015073418657792-ZweW?utm_source=share&utm_medium=member_desktop
8	Spark speculative execution	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7011330464040857600-j8qq?utm_source=share&utm_medium=member_desktop
9	Persistence in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7011729682115039232-kqoC?utm_source=share&utm_medium=member_desktop
10	Apache Spark Web UI - Understanding Spark Execution	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-activity-7012099556536442880-Eyj9?utm_source=share&utm_medium=member_desktop
11	Apache Spark Window Functions	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-activity-7013169169911209984-SRRy?utm_source=share&utm_medium=member_desktop
12	Apache Spark - How to See Record Count Per Partition in a DataFrame	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-

		7014098797723086848-MxTa?utm_source=share&utm_medium=member_desktop
13	Apache Spark - Support Recursive Folder Lookup (Spark 3.0 new feature)	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7014289643470827520-5BA?utm_source=share&utm_medium=member_desktop
14	Broadcast Join in Apache Spark : an Optimization Technique	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7015911270658830336-iQBq?utm_source=share&utm_medium=member_desktop
15	Pivot and Unpivot a DataFrame in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7016051375037980672-N1A7?utm_source=share&utm_medium=member_desktop
16	RDD V/s Dataframe V/s Dataset	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7016439690669023233-78?utm_source=share&utm_medium=member_desktop
17	Sort() and OrderBy() in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7016825725831843840-sBGh?utm_source=share&utm_medium=member_desktop
18	DAG in Apache Spark - Achieve Fault Tolerance through DAG	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7017138394254241792-92Wy?utm_source=share&utm_medium=member_desktop
19	Stages and their Tasks in Apache Spark Job	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7017471404010864640-YbT?utm_source=share&utm_medium=member_desktop
20	Accumulators in Apache Spark : Shared Variables	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7017807116199493632-afva?utm_source=share&utm_medium=member_desktop
21	Handling of Bad Records in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7018161954254970880-hXZH?utm_source=share&utm_medium=member_desktop
22	Read CSV into Dataframe & Various parameters while reading	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7018524342258978816-LhLQ?utm_source=share&utm_medium=member_desktop
23	Read Json into Dataframe & Various parameters while reading	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7018818782483668992-uwBK?utm_source=share&utm_medium=member_desktop
24	groupByKey() v/s reduceByKey() in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7019566208068833280-QO3?utm_source=share&utm_medium=member_desktop

25	Read/Write from Database Using Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7019913495827664896-D17T?utm_source=share&utm_medium=member_desktop
26	Spark Deploy Modes in YARN - Client vs Cluster	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7020268334323560450-QLZs?utm_source=share&utm_medium=member_desktop
27	Understanding Stack Function in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7020630722042351616-N3sf?utm_source=share&utm_medium=member_desktop
28	Transpose Dataframe without Aggregation in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7021000659227332608-ftSt?utm_source=share&utm_medium=member_desktop
29	Exploring Apache Spark's Column Methods	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7021347947866968065-HE7d?utm_source=share&utm_medium=member_desktop
30	Joins in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7021680136919912448-3ztI?utm_source=share&utm_medium=member_desktop
31	Overview of Spark Streaming	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7022027425291128832-zqbi?utm_source=share&utm_medium=member_desktop
32	DStream in Apache Spark (Spark Streaming)	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7022389812850556928-KewW?utm_source=share&utm_medium=member_desktop
33	Spark Structured Streaming	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7022782399616880640-EMPj?utm_source=share&utm_medium=member_desktop
34	Checkpointing in Spark Streaming - PART 1	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7023152337242255361-xyN0?utm_source=share&utm_medium=member_desktop
35	Checkpointing in Spark Streaming - PART 2	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7023492076201791488-e9Em?utm_source=share&utm_medium=member_desktop
36	Difference between Spark Checkpointing and Persist	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7023960160310108160-5Eax?utm_source=share&utm_medium=member_desktop
37	Resource Allocation in Apache Spark - PART 1	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-

		7024337647699443712-cZDX?utm_source=share&utm_medium=member_desktop
38	Resource Allocation in Apache Spark - PART 2	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7024684936112590849-Ceot?utm_source=share&utm_medium=member_desktop
39	Narrow and Wide Transformation in Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7025047323697172480-5ARy?utm_source=share&utm_medium=member_desktop
40	Drop function in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7025409711596339200-k4FO?utm_source=share&utm_medium=member_desktop
41	Describe and Summary functions in Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7025779649607589888-TdPn?utm_source=share&utm_medium=member_desktop
42	Data Serialization : An Optimization Technique	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7026104288779137025-Qwcs?utm_source=share&utm_medium=member_desktop
43	Apache Spark SQL Sampling	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7027653199595544576-kSrz?utm_source=share&utm_medium=member_desktop
44	Read XML file in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7028044240513355776-vtxC?utm_source=share&utm_medium=member_desktop
45	Read/Write parquet file format in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7028757894380625920-syAE?utm_source=share&utm_medium=member_desktop
46	Output Modes in Spark Streaming	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7030874502678941696-Papa?utm_source=share&utm_medium=member_desktop
47	Output Sinks in Spark Structured Streaming	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-activity-7031961476730945536-WGGK?utm_source=share&utm_medium=member_desktop
48	Data Locality in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7032755041811779584-19FW?utm_source=share&utm_medium=member_desktop
49	8 Performance Optimization Techniques Using Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-activity-7034035761813348352-gzkP?utm_source=share&utm_medium=member_desktop
50	Apache Spark column functions in Dataframe	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-

		7034508650685636608-CSSn?utm_source=share&utm_medium=member_desktop
51	Adaptive Query Execution -Coalescing Post Shuffle Partitions	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-scala-spark-activity-7035523253775319040-0XKv?utm_source=share&utm_medium=member_desktop
52	Adaptive Query Execution - Converting join strategy	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7035985703763156992-dbNK?utm_source=share&utm_medium=member_desktop
53	Adaptive Query Execution - Optimizing Skew Join	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7036373836753227777-xAqx?utm_source=share&utm_medium=member_desktop
54	Transform Function in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7036605226103054336-fQ4G?utm_source=share&utm_medium=member_desktop
55	Zip function in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7038482021236543488-75bg?utm_source=share&utm_medium=member_desktop
56	Predicate Push-Down and Partition Pruning Optimization Technique	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7039586673885646848-WPtK?utm_source=share&utm_medium=member_desktop
57	Performance Tuning Config Options	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-activity-7042068886116073472-z14c?utm_source=share&utm_medium=member_desktop
58	Troubleshooting common errors in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7044686342127067136-xBaS?utm_source=share&utm_medium=member_desktop
59	Convert File Formats in Apache Spark (JSON/CSV/AVRO/PARQUET)	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-spark-learningeveryday-activity-7045092991425794050-rLQd?utm_source=share&utm_medium=member_desktop
60	Garbage Collection Tuning in Apache Spark	https://www.linkedin.com/posts/charchit-patidar_60daysofspark-lastpostofsparkseries-spark-activity-7046487352705220608-J93S?utm_source=share&utm_medium=member_desktop