

Topic	Type	links
Spark course	Udemy (Mandatory)	https://deloittedevelopment.udemy.com/course/apache-spark-programming-in-python-for-beginners/learn/lecture/35799518#overview
Data Modelling	Case- Study (Mandatory)	https://www.techshashank.com/data-warehousing/shipping-dimensional-modeling
Spark Optimizations	Customized Doc	Master Doc with all the details --> https://medium.com/codex/apache-spark-optimization-techniques-and-tuning-52cc446c07f5 - -> https://medium.com/inspiredbrilliance/spark-optimization-techniques-a192e8f7d1e4
Storage Formats	Knowledge Articles	https://www.upsolver.com/blog/the-file-format-fundamentals-of-big-data https://www.adaltas.com/en/2020/07/23/benchmark-study-of-different-file-format/ https://techmagie.wordpress.com/category/big-data/data-formats/
Databricks	Git Hub Repo	https://github.com/ryanchynoweth44/DatabricksContent
Replication Strategies	Knowledge Articles	https://www.striim.com/blog/the-7-data-replication-strategies-you-need-to-know/
Data Consistency	Knowledge Articles	https://www.scaler.com/topics/data-consistency-in-dbms/
Data partitioning	Knowledge Articles	https://medium.com/data-engineer-things/data-partitioning-and-bucketing-examples-and-best-practices-15bcadd35479
Database Types	Knowledge Articles	https://medium.com/@muddaprince456/types-of-databases-bfdab96cfe3
Data pipeline Architecture	Knowledge Articles	https://streamsets.com/blog/data-pipeline-architecture-deep-dive/ https://www.upsolver.com/blog/data-pipeline-architecture-building-blocks-diagrams-and-patterns
HDFS	Knowledge Articles	https://www.databricks.com/glossary/hadoop-distributed-file-system-hdfs#:~:text=HDFS%20(Hadoop%20Distributed%20File%20System,handle%20and%20store%20big%20data.
Batch and Streaming Architecture	Lucidchart	https://lucid.app/lucidchart/2d443cd4-7874-4dab-a299-54b0c10fb3ca/view?page=0_0#
Day 1 ppt	One drive	DE Deck.pptx