Explicación del metadatata en Hive:

<https://www.quora.com/What-is-Hive-Metastore>

Creación de tablas en Hive con diferentes formatos de archivos de entrada:

<https://community.cloudera.com/t5/Support-Questions/how-to-load-double-quotes-data-of-fields-in-hive-table/td-p/189295>

<https://stackoverflow.com/questions/30637536/how-to-load-csv-data-with-enclosed-by-double-quotes-and-separated-by-tab-into-hi>

Explicación de DAG en Apache Spark:

<https://data-flair.training/blogs/dag-in-apache-spark/>

Explicación DAG:

<https://es.wikipedia.org/wiki/Grafo_ac%C3%ADclico_dirigido>

Explicación Impala, buenas prácticas y comparación de rendimiento con herramientas similares (hive, sparksql, presto, Greenplum):

<https://medium.com/@adirmashiach/apache-impala-my-insights-and-best-practices-8e0507089935>

<https://blog.cloudera.com/apache-impala-leads-traditional-analytic-database/>

Explicación de Hive e Impala y principales diferencias entre ellos:

<https://data-flair.training/blogs/impala-vs-hive/>

Comparación de estadísticas al ejecutar mismas consultas en diferentes herramientas: Impala, Presto y Hive en MR3:

<https://mr3.postech.ac.kr/blog/2019/03/22/performance-evaluation-0.6/>

Explicación de Hive, Impala, Presto y Spark, Pros y contras de cada herramienta, recomendaciones de para que usar cada herramienta:

<https://www.janbasktraining.com/blog/how-to-compare-hive-spark-impala-and-presto/>

Explicación de apache tez:

<https://es.cloudera.com/products/open-source/apache-hadoop/apache-tez.html>

Bloques de datos en HDFS <https://data-flair.training/blogs/data-block/>