Classic Data Engineering/Big Data Papers

This assignment lists some of the classic big data papers that have had the most impact and influence in the data engineering field. Their order is chronological:

[The Google File System](#)

[MapReduce: Simplified Data Processing on Large Clusters](#)

[Bigtable: A Distributed Storage System for Structured Data](#)

[Spark: Cluster Computing with Working Sets](#)

These will give you a great fundamental understanding of the underpinning technologies that support data engineering and big data.