# Key Points To Remember

The main ideas we discussed in this chapter are summarized below:

- The 3Vs of Big Data (Volume, Velocity and Variety) are driving the adoption of Apache Hadoop:
  - Industry expects 44ZB of data by 2020
  - Existing data architectures make data inaccessible, incomplete, irrelevant, and expensive

- Top use case families that have emerged around Hadoop are: Data Discovery, Single View, Predictive Analytics, Active Archive, ETL Onboard, Data Enrichment, etc.

- Hadoop is an OLAP data platform and it evolved as part of the database industry. Hadoop is:
  - a scalable, fault tolerant, open source framework for the distributed storing (HDFS) and processing (YARN) of large sets of data on commodity hardware
  - a clean-room implementation of two papers from Google
  - optimized for schema-on-read use cases
  - the basis of the enterprise "data lake" architecture

- Hadoop is an open source project governed by the Apache Software Foundation (ASF)

- Apache Bigtop is to Hadoop what Debian is to Linux

- ODPi helps ASF with harmonizing industry expectations around the Hadoop ecosystem.

Learn About Verified Certificates