

The overall layout of this survey article is as follows:

- 1. Introduction**
- 2. Scaling**
  - (i) Horizontal Scaling**
  - (ii) Vertical Scaling**
- 3. Horizontal Scaling Platforms**
  - 3.1 Peer-to-Peer Networks**
  - 3.2 Apache Hadoop**
    - 3.2.1. MapReduce*
    - 3.2.2 MapReduce Wrapper*
    - 3.2.3 Limitations of MapReduce*
  - 3.3 Spark: Next Generation Data Analysis Paradigm**
- 4. Vertical Scaling Platforms**
  - 4.1 High Performance Computing (HPC) Clusters**
  - 4.2 Multicore CPU**
  - 4.3 Graphics Processing Unit (GPU)**
  - 4.4 Field Programmable Gate Arrays (FPGA)**
- 5. Comparison of Different Platforms**
  - (i) System/Platform dependent Characteristics**
    - i. Scalability**
    - ii. Data I/O Performance**
    - iii. Fault Tolerance**
  - (ii) Application/Algorithm dependent Characteristics**
    - i. Real-Time Processing**
    - ii. Data Size Supported**
    - iii. Iterative Tasks Support**
- 6. Choice of a Platform for Big Data Analytics?**
  - (i) Data Size**
  - (ii) Speed or Throughput Optimization**
  - (iii) Training/Applying a Model**
  - (iv) Practical Implications**
- 7. K-Means Clustering on Different Platforms**
  - (i) K-means on MapReduce**
  - (ii) K-means on MPI**
  - (iii) K-means on GPU**
  - (iv) K-means on other platforms**
- 8. Conclusion and Future Directions**