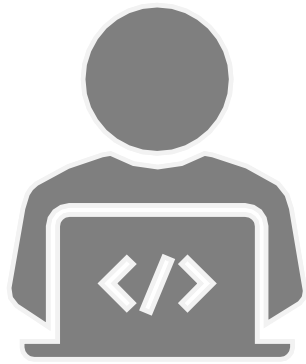




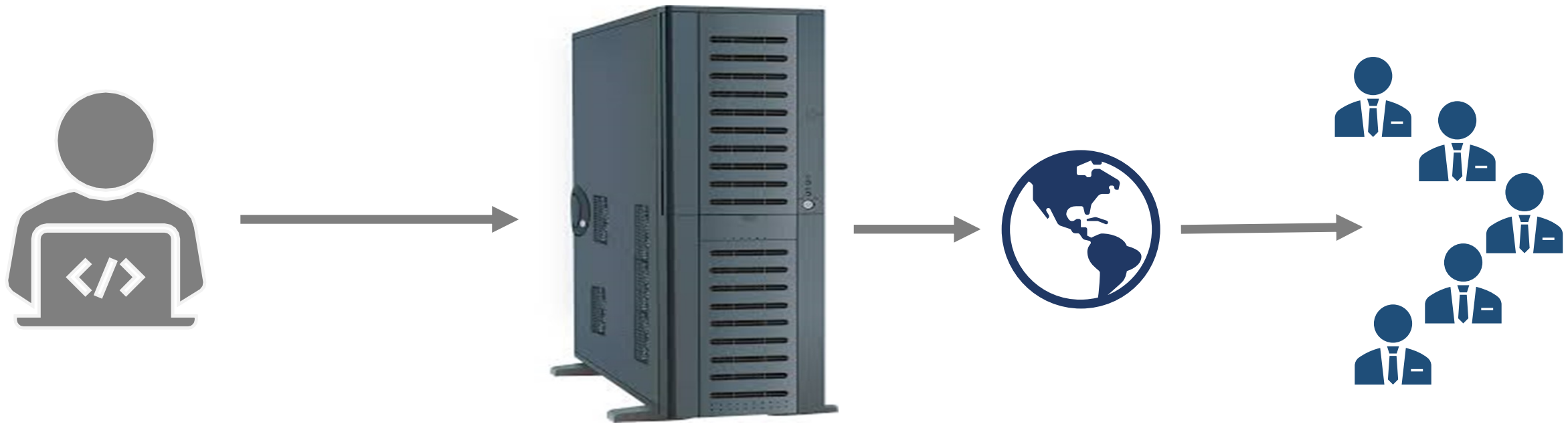
Apache Spark

History of Distributed Computing

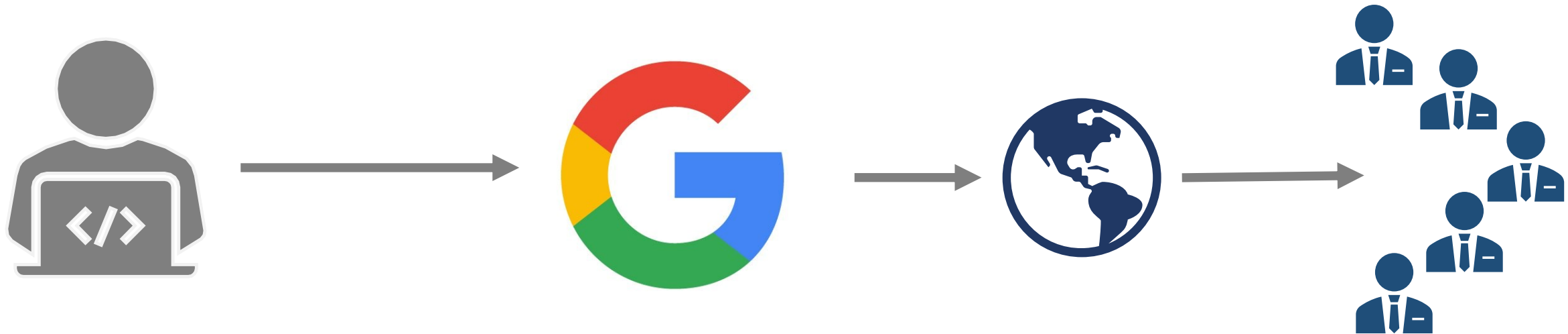
History and Evolution



History and Evolution



History and Evolution



History and Evolution



1. Data Collection and Ingestion
2. Data Storage and Management
3. Data Processing and Transformation
4. Data Access and Retrieval

History and Evolution



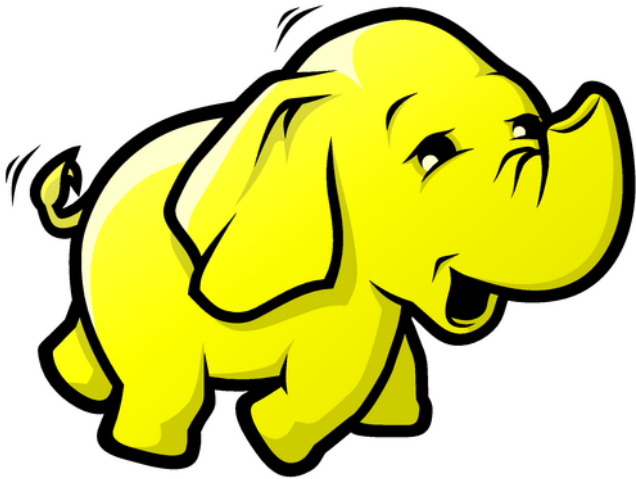
1. Google File System – 2003

<https://ai.google/research/pubs/pub51>

2. MapReduce – 2004

<https://ai.google/research/pubs/pub62>

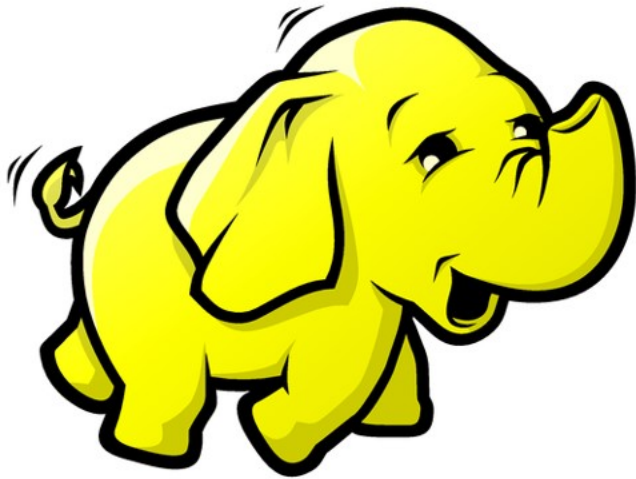
History and Evolution



HADOOP

1. HDFS - Hadoop Distributed File System
2. Hadoop MapReduce

History and Evolution



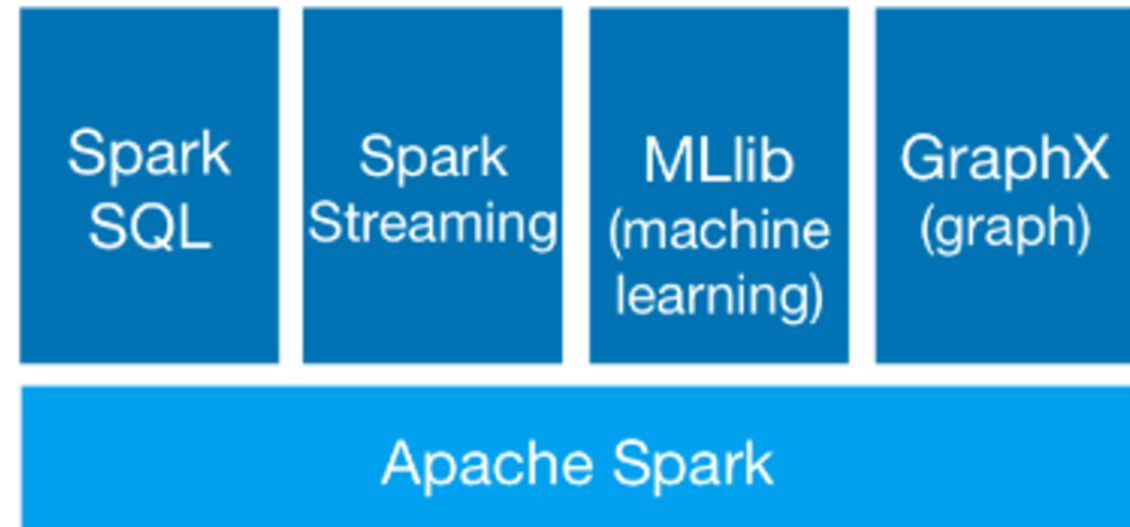
HADOOP



Map Reduce
HDFS



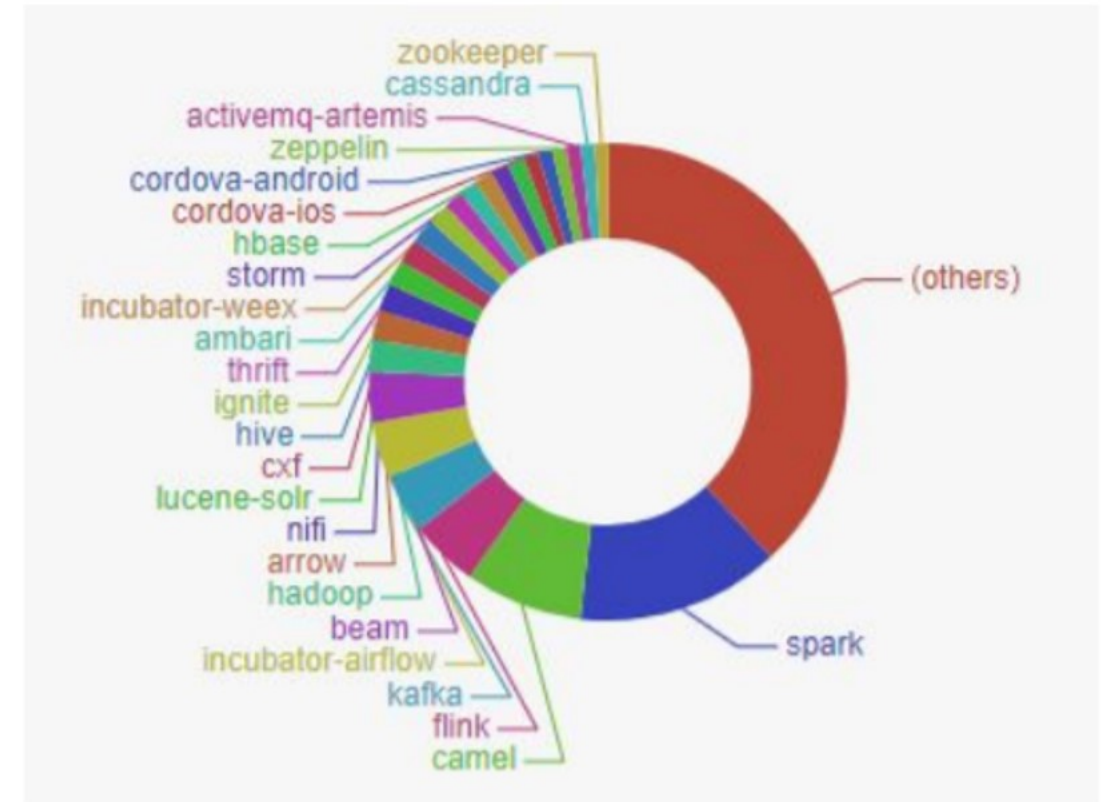
History and Evolution



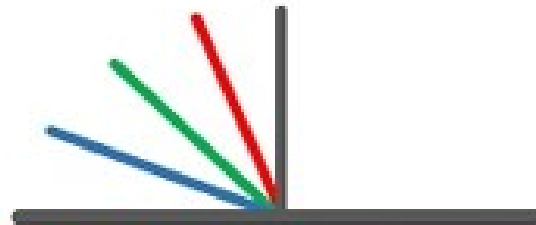
History and Evolution



Most Active Sources: Visits



<https://files-dist.s3.amazonaws.com/AnnualReports/FY2019+Annual+Report.pdf>



LEARNING JOURNAL

www.learningjournal.guru