

Universidad de las Américas Puebla



Difference between Spark and MPI

Parallel Computing

Octavio Palomeque Gasperin

151884

February 2019

Difference between Spark and MPI

	MPI	Spark
What is it?	<p>Message Passing Interface is a distributed memory parallel model implementation.</p> <p>Anyway, performance aside, MPI can be used to develop pretty much any parallel code that runs over multiple machines. Note, if you want, you can mix MPI with OpenMP to use threads within machines. This is called hybrid programming.</p>	<p>Spark is a general-purpose distributed data processing engine that is suitable for use in a wide range of circumstances. On top of the Spark core data processing engine, there are libraries for SQL, machine learning, graph computation, and stream processing, which can be used together in an application.</p>
When to use it?	<p>Used to develop parallel scientific applications because they are usually tightly synchronous code and well load balanced. Why this is important is that having stragglers makes MPI program inefficient.</p>	<p>Scenarios where you'd like to use Hadoop but either it's not efficient or harder to implement what you want.</p>
What platform supports?	<p>It works at operating system level, and must be implemented using C in the programming because is at a very low level of the machine.</p>	<p>Programming languages supported by Spark include: Java, Python, Scala, and R. Application developers and data scientists incorporate Spark into their applications to rapidly query, analyze, and transform data at scale. Tasks most frequently associated with Spark include ETL and SQL batch jobs across large data sets, processing of streaming data from sensors, IoT, or financial systems, and machine learning tasks.</p>

References

Carol McDonald. Spark 101: What Is It, What It Does, and Why It Matters. <https://mapr.com/blog/spark-101-what-it-what-it-does-and-why-it-matters/>

Saliya Ekanayake. When does we use Hadoop, Spark, Flink, MPI, PVM, and OpenMP, and why? <https://www.quora.com/When-do-we-use-Hadoop-Spark-Flink-MPI-PVM-and-OpenMP-and-why>

Wes Kendall. MPI Hello World. <http://mpitutorial.com/tutorials/mpi-hello-world/>

<https://spark.apache.org/>