Discussion 2

What are the other frameworks like Spark for real time analytics? Briefly, compare 1 other framework with Spark.

* Apache Flink

Apache Flink was designed to handle large volumes of data and provide insights into the data as soon as possible (real-time analytics). It is important to state both Spark and Flink are open sources for the batch processing as well as the stream processing at the massive scale (this provides fault tolerance). Both tools are solutions to several Big Data problems.

Comparison:

* Computation model: Flink is based on the operator based computational model. Spark is based on the micro batch model.
* Streaming: Flink uses streams for all workloads. Spark uses micro-batches for all workloads.
* Speed: Flink is faster than Spark.
* Memory: Flink provides automatic memory management while Spark provides configurable memory management.
* Duplication: Both tools process every record exactly one time hence eliminates duplication.
* Latency: Flink streaming run-times achieves low latency and high throughput. Spark has high latency as compared to Flink.
* Performance: Flink uses native closed loop iterations operators which makes ML and graph processing faster. The stream processing is not that efficient than Flink as it uses micro-batch processing.

Overall, it looks like Flink might be a better framework than Spark following the outlined comparison. It all depends on the specific scenarios and when to use which tool. However, more research and personal experience might be needed in order to determine this.

Sources:

<https://www.velotio.com/engineering-blog/apache-flink-a-solution-for-real-time-analytics>

<https://data-flair.training/blogs/comparison-apache-flink-vs-apache-spark/>