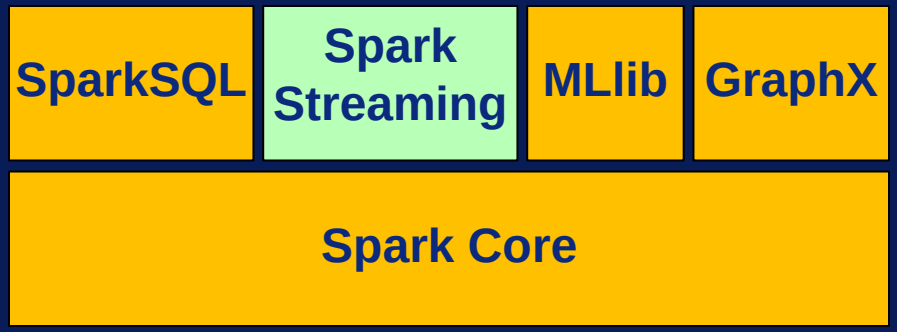
Spark Streaming

After this video you will be able to..

- Summarize how Spark reads streaming data
- List several sources of streaming data supported by Spark
- Describe Spark’s sliding windows



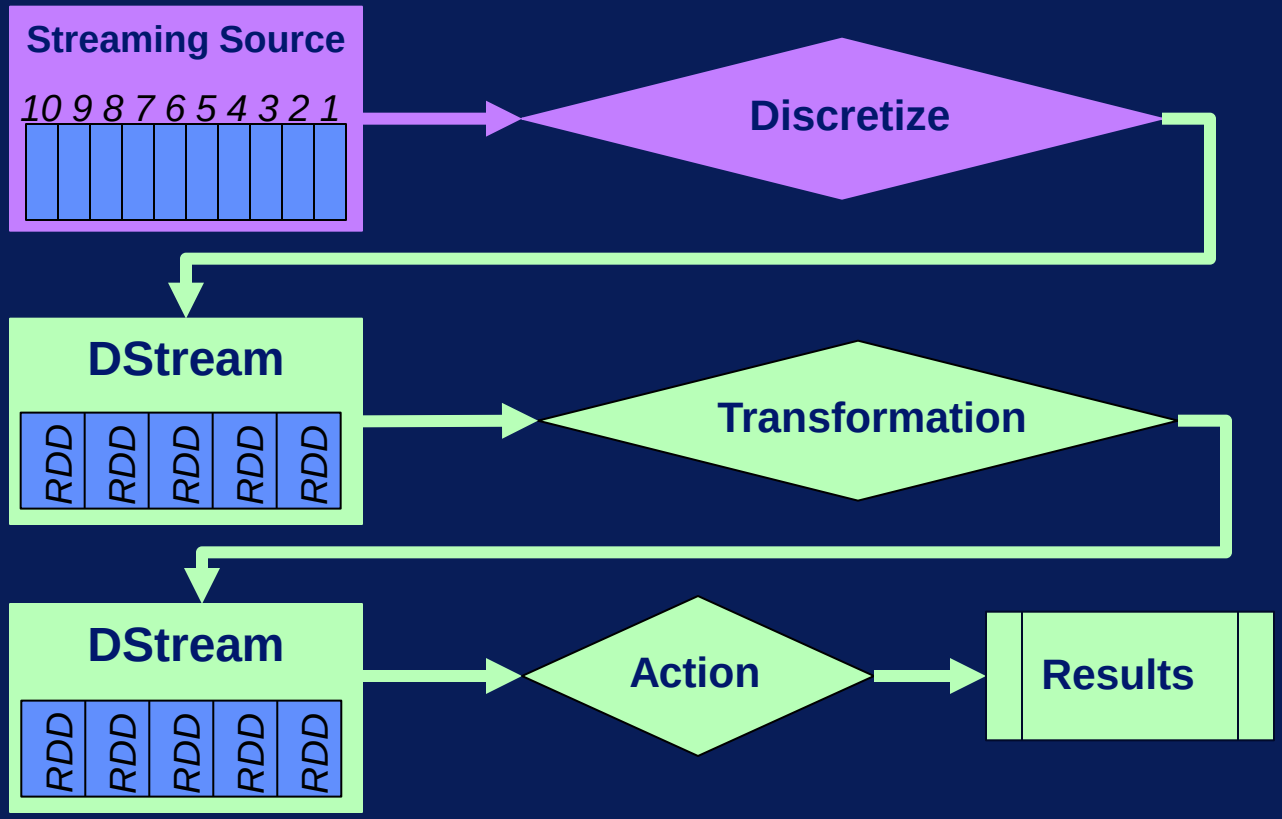
Spark Streaming

- Scalable processing for real-time analytics
- Data streams converted to discrete RDDs
- Has APIs for Scala, Java, and Python

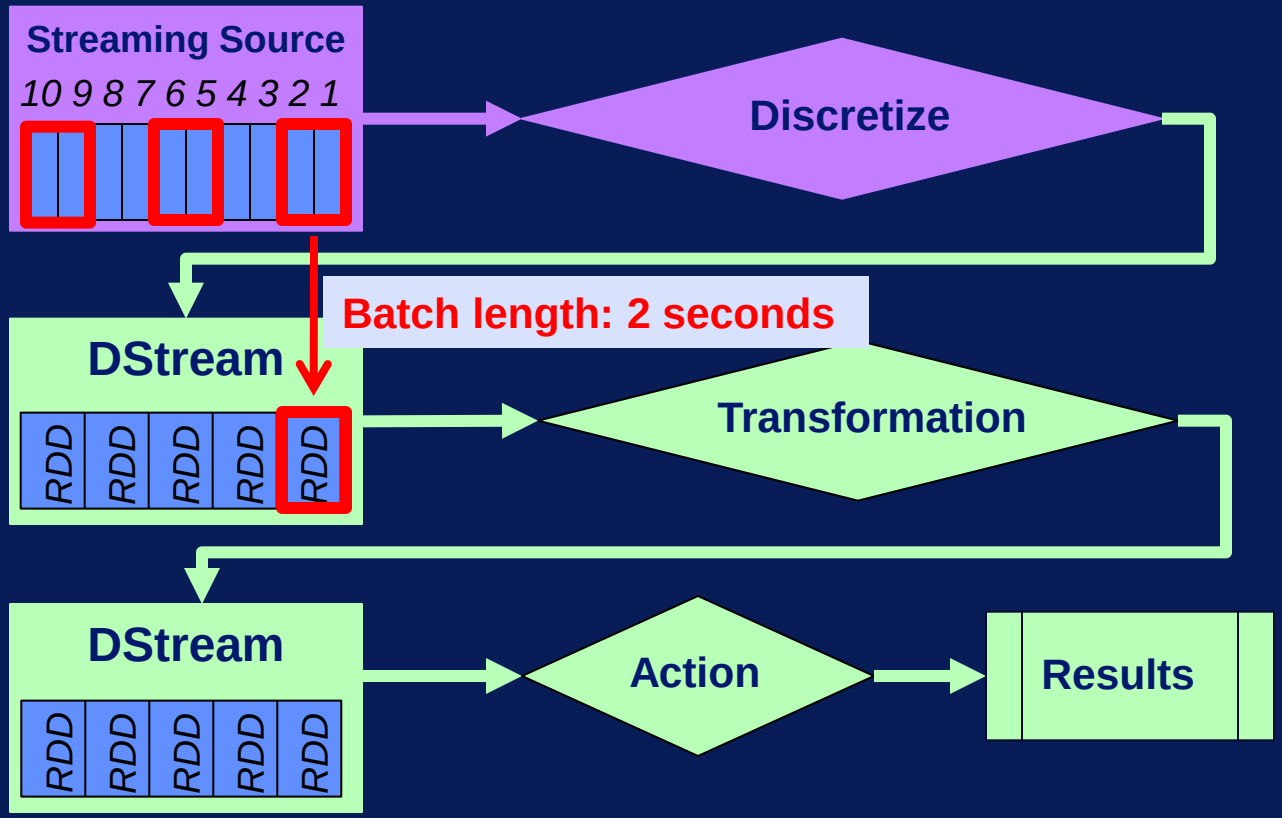
Spark Streaming Sources

- Kafka
- Flume
- HDFS
- S3
- Twitter
- Socket
- …etc.

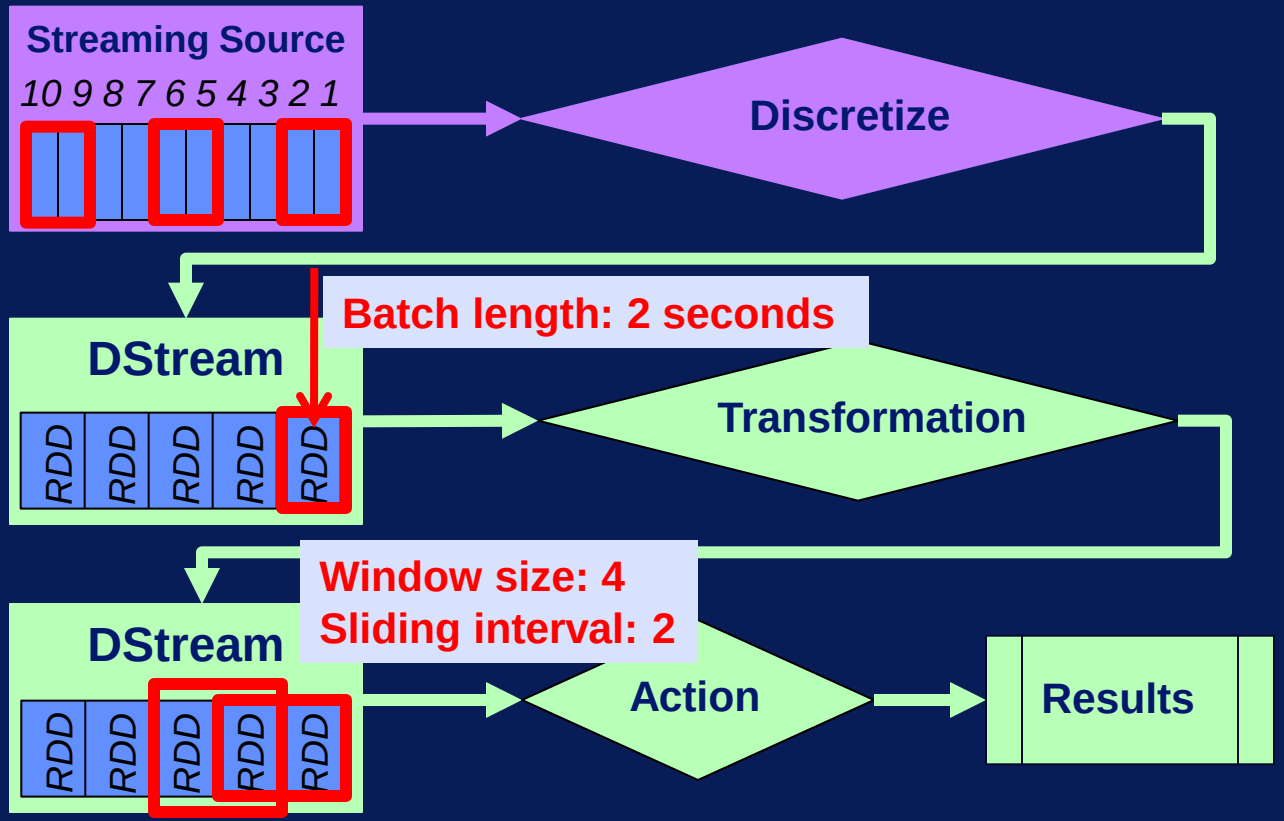
Creating and Processing DStreams



Creating and Processing DStreams



Creating and Processing DStreams



Main Take-Aways

- Spark uses DStreams to make discrete RDDs from streaming data.
  - Same transformations and calculations applied to batch RDDs can be applied
- DStreams can create a sliding window to perform calculations on a window of time.