

# Apache Spark Overview

**Vadim Bichutskiy**

George Mason University  
[vadim.bichutskiy@gmail.com](mailto:vadim.bichutskiy@gmail.com)

**Keywords:** Apache Spark, Cluster Computing, Big Data, Analytics, Data Science

Apache Spark is a fast and general engine for large-scale data processing that is one of the hottest technologies in Big Data. Spark improves upon MapReduce by allowing for in-memory computation, unifying batch, streaming and interactive workloads, and providing high-level APIs in Java, Scala, Python, and R. Spark includes components for SQL, streaming, machine learning, and graph processing. This talk will give an overview of Spark including its improvements over MapReduce and its fundamental abstractions.