



## Managed Solutions Apache Spark®

Apache Spark™ is a high performing engine for large-scale analytics and data processing. While Apache Spark™ provides advanced analytics capabilities, it requires a fast, distributed back-end data store. Apache Cassandra® is the most modern, reliable and scalable choice for that data store. Spark when fully integrated with the key components of Cassandra, provides the resilience and scale required for big data analytics.

## Managed Spark Offering



### Collocated Data Engine

Spark engine resides in the operational database, so there is no need for extracting, transforming, and loading into a new environment.



### Monitoring

We provide integrated Spark management and monitoring through the Instaclustr Management Console as well as provide 24x7 monitoring and support to our Spark customers.



### Performance

Spark has an advanced DAG execution engine that supports cyclic data flow and in-memory computing. It runs programs up to 100x faster than Hadoop MapReduce in memory, or 10x faster on disk.



### Unified Memory Management

Spark comes with the shared memory pool for execution and caching.



## Automated Provisioning of Spark Jobserver and Apache Zeppelin

To provide easy access to your Spark processing engine, Instaclustr's Spark cluster can include Spark Jobserver (REST API) and Apache Zeppelin (analyst notebook UI).



## Reliability

Instaclustr Managed Spark architecture and support offering enables you to use the power of Spark from your application, with the confidence to meet your availability and processing requirements.

Instaclustr Managed Apache Spark provides a reliable and managed platform, collocated with your Apache Cassandra® data store, to leverage the power of Apache Spark for stream or batch analytics.

# About Instaclustr

Instaclustr is the open source-as-a-service company delivering reliability at scale through our integrated data platform for technologies such as [Apache Cassandra®](#), [Apache Kafka®](#), [Apache Spark™](#), [Elasticsearch](#), and [Redis](#).

Our expertise stems from delivering more than 70 million node hours under management. We provide a range of consulting, enablement, and integration, and support services relating to open source technologies.

Our integrated data platform, built on open source technologies, powers mission-critical, highly available applications for our customers and help them achieve scalability, reliability, and performance for their applications.

