**Intro to Apache Spark Workshop**

Robert Sanders, Clairvoyant LLC

This workshop introduces participants to Apache Spark, an open source distributed computing framework used to process large volumes of data. We will primarily focus on Core Apache Spark and do a deep dive into the implementation of Resilient Distributed Datasets (RDDs) in the framework. Details will be provided on how to work with the API’s as well as how to develop and execute Spark batch jobs. The session will include examples and exercises to help participants gain hands on knowledge of Spark so that they walk away with a better understanding of how Apache Spark can benefit their organization.

**Laptop Required**