## Apache Spark Nedir?

[Apache Spark](http://spark.apache.org/),  büyük veri kümeleri üzerinde paralel olarak işlem yapmamızı sağlayan Scala ile geliştirilmiş açık kaynak kodlu kütüphanedir

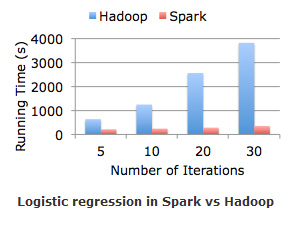


* HDFS ile verileri birden fazla makinada saklayabilir ve yönetebiliriz.
* MapReduce  ile büyük verileri paralel olarak işleyebiliriz

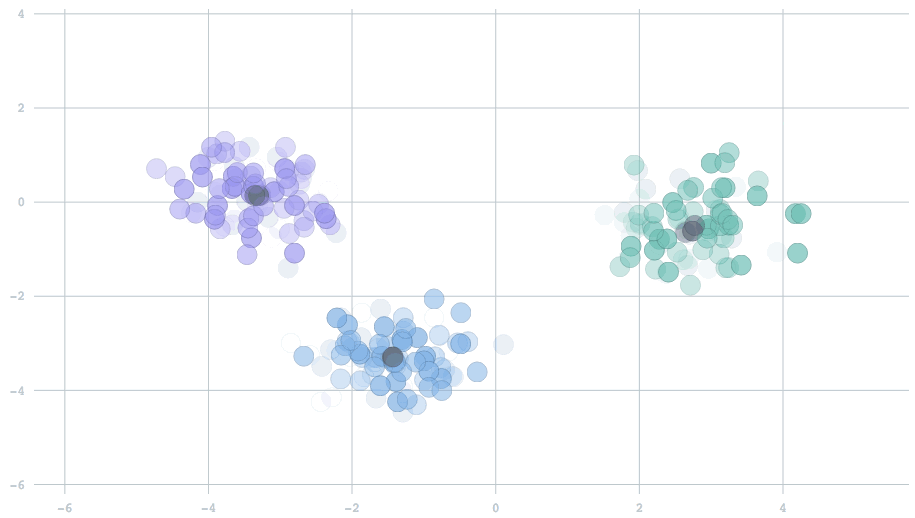
Genel olarak Spark ı , MapReduce alternatifi olarak kullanabiliriz . Biz Hadoop kullanarak verileri yine HDFS de saklayabiliriz fakat Apache Spark ile bu verileri daha **kolay** ve daha **hızlı** bir biçimde işleyebiliriz

### Spark ın özellikleri nelerdir?

* Kullanımı kolaydır. Daha önce MapReduce ile geliştirdiğimiz projeleri Apache Spark ile daha az eforla geliştirebiliyoruz
* MapReduce a göre daha hızlıdır . Spark , kendi sitesinde MapReduce a göre memory işlemlerini 100 kat daha hızlı yaptığını iddaa etmektedir . Geliştirdiğim projelerde Spark ın daha hızlı oldugunu gördüm fakat bu kadar farkı yakalayamadım . Ama genel olarak daha hızlı diyebiliriz



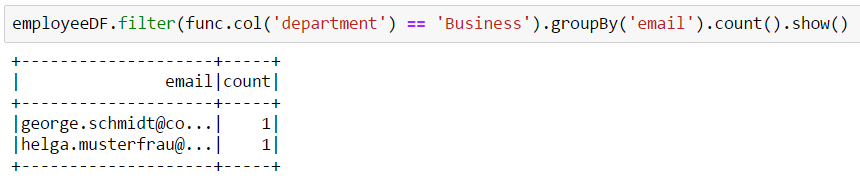
* Java  , Scala , Python ile geliştirilebilir . Ben şu an Java ile geliştiriyorum fakat aradığım bazı örneklerde en çok Scala ile geliştirilmiş kodlarla karşılaşıyorum
* MLib kütüphanesi sayesinde machine learning (makine öğrenmesi) uygulamaları yazabilirsiniz.



* Spark Streaming ile verileri anlık olarak işleyebilirsiniz. Yapmış oldugumuz Spark Streaming projesinde anlık olarak gelen verileri 5 dakikalık bloklara ayırarak, verileri üzerinde çeşitli analizler yaptık. Bu sayede akan veri üzerinde analiz yapmış olduk



* Spark SQL, DataFrame gibi yapılarla büyük veriler üzerinde SQL tabanlı analizler yapabilirsiniz. Geliştirdiğimiz projelerde bu yapılar işlerimizi oldukça kolaylaştırıyor



* Farklı kaynaklarda tutulan büyük verileri analiz edebilirsiniz. HDFS, Kafka,Cassandra, HBase, S3 …  Geliştirdiğimiz projelerde HDFS ve Kafka verilerilerini rahatlıkla analiz edebildik
* Kurulumu basittir. Local bilgisayarlarınızda eğer benim gibi java kullanıyorsanız basit bir kütüphane ekleme işlemi ile geliştirmeye başlayabilirsiniz. Eğer birden fazla makinaya spark kurmak istiyorsanız yine kurulum adımları basitleştirilmiştir.