# Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing

## 中文翻译

<http://shiyanjun.cn/archives/744.html>

<http://spark.apachecn.org/paper/zh/spark-rdd.html#21-rdd-%E6%8A%BD%E8%B1%A1>

<http://talentluke.iteye.com/blog/2236879>

<http://blog.sciencenet.cn/blog-425672-520947.html>

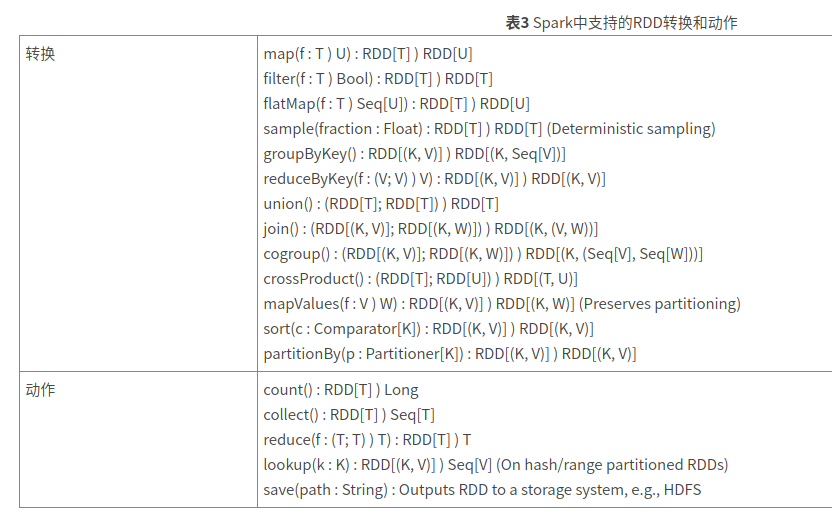
## 英文原文

<https://pdos.csail.mit.edu/6.824/papers/zaharia-spark.pdf>

<https://www2.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-82.pdf>

# RDD

## RDD操作



## 每个RDD包括的内容

### 一组RDD分区（partition，即数据集的原子组成部分）

### 对父RDD的一组依赖，这些依赖描述了RDD的Lineage

### 一个函数，即在父RDD上执行何种计算

### 元数据，描述分区模式和数据存放的位置。