Spark

RDDs never change; thus the adjective *immutable*. RDDs are *resilient* because of Spark’s built-in fault recovery mechanics. RDDs are different: they prov

ide fault tolerance by logging the transformations used to build a dataset (how it came to be) rather than the dataset itself . For example, in the previous section, the process of loading the text file yielded the licLines RDD. Then you applied the filter function to licLines, which produced

the new bsdLines RDD. Those transformations and their ordering are referred

to as *RDD lineage*.

There are two types of RDD operations: transformations and actions. *Transformations*

(for example, filter or map) are operations that produce a new RDD by performing

some useful data manipulation on another RDD. *Actions* (for example, count or

foreach) trigger a computation in order to return the result to the calling program or

to perform some actions on an RDD’s elements.