CLOUD COMPUTING APPLICATIONS

RDD Fault Tolerance

Roy Campbell & Reza Farivar

RDD Fault Tolerance

RDDs maintain *lineage* information that can be used to reconstruct lost partitions

Ex:

```
messages = textFile(...).filter(_.startsWith("ERROR"))
                        .map(_.split('\t')(2))
```

HDFS File → Filtered RDD → Mapped RDD

filter
(func = _.contains(...))

map
(func = _.split(...))