# foreachRDD

# The `foreachRDD(`*`func`*`)` Operation

- The most generic output operator that applies a function, *func*, to each RDD generated from the stream.
- This function should push the data in each RDD to an external system, such as saving the RDD to files, or writing it over the network to a database.
- Note that the function *func* is executed in the driver process running the streaming application, and will usually have RDD actions in it that will force the computation of the streaming RDDs.

# Mistakes

```
# Mistake Number 1
def sendRecord(rdd):
    connection = createNewConnection()  # executed at the driver
    rdd.foreach(lambda record: connection.send(record))
    connection.close()
dstream.foreachRDD(sendRecord)
```

```
# Mistake Number 2
def sendRecord(record):
    connection = createNewConnection()
    connection.send(record)
    connection.close()
dstream.foreachRDD(lambda rdd: rdd.foreach(sendRecord))
```

# Optimization Strategies

```
# Optimization Strategy 1
def sendPartition(iter):
    connection = createNewConnection()
    for record in iter:
        connection.send(record)
    connection.close()
dstream.foreachRDD(lambda rdd: rdd.foreachPartition(sendPartition))
```

```
# Optimization Strategy 2
def sendPartition(iter):
    connection = ConnectionPool.getConnection()
    for record in iter:
        connection.send(record)
    ConnectionPool.returnConnection(connection)
dstream.foreachRDD(lambda rdd: rdd.foreachPartition(sendPartition))
```

# To the Code!