```
In [1]:   import findspark
          from pyspark.streaming import StreamingContext
          from pyspark.sql import SparkSession
          import math

          findspark.init()
```

```
In [85]:  try: ssc.stop(True, True)
          except: pass
          try: spark.stop()
          except: pass
```

```
printjoin []
printjoin []
printjoin []
printjoin []
printjoin []
printjoin []
printjoin []
printjoin []
printjoin []
```

```
In [77]:  spark=SparkSession.builder.appName("SparkStreaming-06").master('local[1]').getOrCreate()
          sc=spark.sparkContext
          # .config("spark.driver.allowMultipleContexts","true")
          spark
```

Out[77]: **SparkSession - in-memory**

SparkContext

Spark UI

Version	v2.4.8
Master	local[1]
AppName	SparkStreaming-06

```
In [78]:  ssc=StreamingContext(sc, 1)
          # ssc.checkpoint('checkpoint-06')
          ssc # 1=1 second
```

Out[78]: `<pyspark.streaming.context.StreamingContext at 0x1f4eb8fb308>`

```
In [79]:  stream1=ssc.queueStream([
              sc.parallelize([('a',1), ('b',2), ('c',3)]),
              sc.parallelize([(2,'h'),(2,'d'),(1,'j'),(2,'e'),(1,'l'),(3,'m')])
          ],oneAtATime=True)
```

```
In [80]:  stream2=ssc.queueStream([
              sc.parallelize([('a',True), ('b',False), ('d',True)]),
              sc.parallelize([(2,'e'),(2,'d'),(1,'j'),(2,'k'),(1,'l'),(3,'m')])
          ],oneAtATime=True)
```

```
In [81]:  stream3=ssc.queueStream([
```

```
        sc.parallelize([('a',1), ('b',2), ('c',3)]),
        sc.parallelize([('a',True), ('b',False), ('d',True)]),
    ],oneAtATime=True)
```

In [82]:
```
dataset = sc.parallelize([('e',True), ('b',False), ('d',True)])
```

In [83]:
```python
def printjoin(rdd):
    print("printjoin", rdd.collect())

def printtransformjoin(rdd):
    print("printtransformjoin", rdd.collect())

stream1.join(stream2).foreachRDD(printjoin)
# dataset.transform(lambda rdd: rdd.join(stream2)).foreachRDD(printtransformjoin)
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
C:\Users\DMS240~1\AppData\Local\Temp/ipykernel_20792/2188843946.py in <module>
      6
      7 stream1.join(stream2).foreachRDD(printjoin)
----> 8 dataset.transform(lambda rdd: rdd.join(stream2)).foreachRDD(printtransformjoin)

AttributeError: 'RDD' object has no attribute 'transform'
```

In [84]:
```
ssc.start()
```

```
printjoin [('a', (1, True)), ('b', (2, False))]
printjoin [(2, ('h', 'e')), (2, ('h', 'd')), (2, ('h', 'k')), (2, ('d', 'e')), (2, ('d',
'd')), (2, ('d', 'k')), (2, ('e', 'e')), (2, ('e', 'd')), (2, ('e', 'k')), (1, ('j',
'j')), (1, ('j', 'l')), (1, ('l', 'j')), (1, ('l', 'l')), (3, ('m', 'm'))]
printjoin []
printjoin []
printjoin []
```

In [ ]: