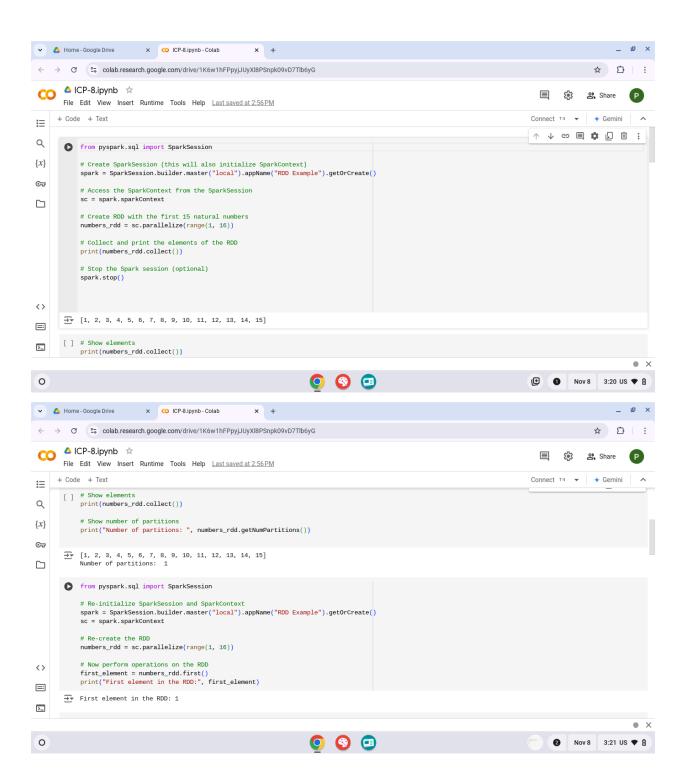
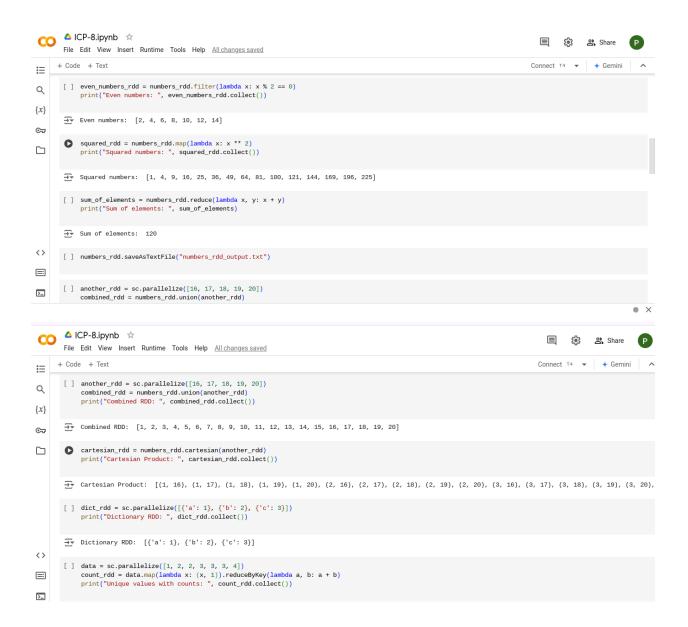
ICP-8 REPORT





```
↑ ↓ ⊖ 🗏 ‡ 🖟 🖽 :
Q
       [] data = sc.parallelize([1, 2, 2, 3, 3, 3, 4])
count_rdd = data.map(lambda x: (x, 1)).reduceByKey(lambda a, b: a + b)
print("Unique values with counts: ", count_rdd.collect())
{x}
        \rightarrow Unique values with counts: [(1, 1), (2, 2), (3, 3), (4, 1)]
# Create RDD by combining multiple text files (file1.txt and file2.txt)
              rdd_from_files = sc.textFile("file1.txt,file2.txt")
              # Show the content from the text files
             print("RDD from text files:", rdd_from_files.collect())
        🚌 RDD from text files: ['This is the first file.', 'It contains some text data.', 'Spark is awesome.', 'This is the second file.', 'It also contains text data
        [ ] # Use the `take` action to retrieve the first 5 lines
first_five_lines = rdd_from_files.take(5)
print("First 5 lines of the RDD:", first_five_lines)
\equiv
        🛨 First 5 lines of the RDD: ['This is the first file.', 'It contains some text data.', 'Spark is awesome.', 'This is the second file.', 'It also contains tex
>_
CO A ICP-8.ipynb 🌣
                                                                                                                                                                 🗏 🌼 😩 Share
        File Edit View Insert Runtime Tools Help All changes saved
                                                                                                                                                              + Code + Text
∷
Q
         # Create a DataFrame
             data = [("sachin", 10), ("rohit", 45), ("virat", 18)]
df = spark.createDataFrame(data, ["Name", "Jersey"])
{x}
              # Show the DataFrame
೦ಘ
             df.show()
| Name|Jersey|
              |sachin|
| rohit|
| virat|
                           18
CO LCP-8.ipynb 🌣
                                                                                                                                                                   ■ Share
         File Edit View Insert Runtime Tools Help All changes saved
      + Code + Text
                                                                                                                                                                ∷
        <del>→</del>*
Q
         # RDD Example (low-level operations)
              # NUD Example = (low-level operations) rdd_example = sc.parallelize([("sachin", 10), ("rohit", 45), ("virat", 18)]) df_example = spark.createDataFrame([("sachin", 10), ("rohit", 45), ("virat", 18)], ["Name", "jersey"])
\{x\}
              df_example.show()
◎
        +----+
| Name|jersey|
|sachin| 10|
              | rohit| 45|
| virat| 18|
```

My github link: https://github.com/PraveenDondapati/bda.git