Unit -4 Spark (Work Report)

Submitted By:

2021UCA1536 - Aparna Srivastava

2021UCA1804 - Pulkit Batra

2021UCA1814 - Muskan Sinha

In Unit 4, we delve into Spark, a versatile distributed computing framework designed for speed and ease of use. Spark enables processing of large-scale data sets across clusters of computers, providing capabilities for data manipulation, machine learning, and real-time analytics. Throughout this unit, we will learn how to harness the power of Spark to handle big data efficiently, perform complex data transformations, and implement advanced analytics algorithms to derive valuable insights from massive datasets.

1. RDD Exploration

Assignment 1 focuses on exploring Resilient Distributed Datasets (RDDs) within Spark, which serve as the foundational data structure in Spark's programming model. The goal of this assignment is to gain a deeper understanding of RDDs and their operations through practical exploration.

The assignment tasks may include:

1. **Creating RDDs**: Students may be tasked with creating RDDs from various data sources such as text files, CSV files, or through parallelizing collections.
2. **Transformations**: Exploring RDD transformations involves applying operations like `map`, `filter`, `flatMap`, `reduceByKey`, etc., to manipulate the data within RDDs. Students might be asked to perform transformations to preprocess data, filter out irrelevant information, or aggregate data based on certain keys.
3. **Actions**: Students will also explore RDD actions, which trigger the execution of transformations and return results to the driver program. Examples of actions include `collect`, `count`, `take`, `reduce`, etc. They might be required to perform actions to retrieve results or save RDD data to external storage.
4. **Performance Optimization**: As part of the exploration, students may be encouraged to analyze the performance of RDD operations and identify opportunities for optimization. This could involve techniques such as partitioning, caching, and leveraging built-in optimizations provided by Spark.
5. **Error Handling and Debugging**: Dealing with errors and debugging Spark applications is an essential skill. Students may encounter scenarios where they need to troubleshoot errors, understand Spark's execution model, and optimize code for better performance.
6. **Documentation and Reporting**: Finally, students might be required to document their exploration process, including the RDD operations performed, insights gained, and any challenges faced during the assignment. Clear and concise reporting is vital for conveying the learnings and observations effectively.

Overall, Assignment 1 on RDD exploration provides students with hands-on experience in working with distributed datasets in Spark, honing their skills in data manipulation, performance optimization, and troubleshooting within the Spark framework.

Link to DataBricks Notebook

RDD Exploration

(https://databricks.com)

```
from pyspark import SparkContext


# Create a SparkContext
sc = SparkContext("local", "RDD Exploration")
```

```
ValueError: Cannot run multiple SparkContexts at once; existing SparkContext(app=Databricks Shell, master=local[8]) created by __init__ at /databricks/python_shell/dbruntime/spark_connection.py:127
```

```
# Create an RDD from a list
data = [1, 2, 3, 4, 5]
rdd = sc.parallelize(data)
```

```
# Perform some basic operations on the RDD
# 1. Count the number of elements
count = rdd.count()
print("Number of elements:", count)
```

```
Number of elements: 5
```

```
# 2. Sum all elements
total_sum = rdd.sum()
print("Sum of all elements:", total_sum)
```

```
Sum of all elements: 15
```

```
# 3. Calculate the mean
mean = total_sum / count
print("Mean of elements:", mean)
```

```
Mean of elements: 3.0
```

```
# 4. Find the maximum and minimum elements
max_element = rdd.max()
min_element = rdd.min()
print("Maximum element:", max_element)
print("Minimum element:", min_element)
```

```
Maximum element: 5
Minimum element: 1
```

```
# 5. Filter elements greater than 3
filtered_rdd = rdd.filter(lambda x: x > 3)
print("Elements greater than 3:", filtered_rdd.collect())
```

```
Elements greater than 3: [4, 5]
```

```
# 6. Map operation to square each element
squared_rdd = rdd.map(lambda x: x*x)
print("Squared elements:", squared_rdd.collect())
```

```
Squared elements: [1, 4, 9, 16, 25]

Sum of elements using reduce: 15
```

2.1 Spark with Sales Data

Link to Collab Notebook

```
pip install pyspark
```

```
    Requirement already satisfied: pyspark in /usr/local/lib/python3.10/dist-packages (3.5.1)
    Requirement already satisfied: py4j==0.10.9.7 in /usr/local/lib/python3.10/dist-packages (from pyspark) (0.10.9.7)
```

```
import pyspark
```

```
import pandas as pd
pd.read_csv('Sales_data.csv')
```

	Sr.No.	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	Month	Sales	City	Hour
0	0	295665	Macbook Pro Laptop	1	1700.00	30-12-2019 00:01	136 Church St, New York City, NY 10001	12	1700.00	New York City	0
1	1	295666	LG Washing Machine	1	600.00	29-12-2019 07:03	562 2nd St, New York City, NY 10001	12	600.00	New York City	7
2	2	295667	USB-C Charging Cable	1	11.95	12-12-2019 18:21	277 Main St, New York City, NY 10001	12	11.95	New York City	18
3	3	295668	27in FHD Monitor	1	149.99	22-12-2019 15:13	410 6th St, San Francisco, CA 94016	12	149.99	San Francisco	15
4	4	295669	USB-C Charging Cable	1	11.95	18-12-2019 12:38	43 Hill St, Atlanta, GA 30301	12	11.95	Atlanta	12
...	...	...	...	...	...	...	...	...	...	...	...
185945	13617	222905	AAA Batteries (4-pack)	1	2.99	07-06-2019 19:02	795 Pine St, Boston, MA 02215	6	2.99	Boston	19
185946	13618	222906	27in FHD Monitor	1	149.99	01-06-2019 19:29	495 North St, New York City, NY 10001	6	149.99	New York City	19

```
from pyspark.sql import SparkSession
```

```
from pyspark.sql.functions import col, sum
from pyspark.sql.types import FloatType, IntegerType
```

```
spark=SparkSession.builder.appName('Practise').getOrCreate()
#Initialize spark session
```

```
spark
```

SparkSession - in-memory

SparkContext

Spark UI

Version
```
    v3.5.1
```
Master
```
    local[*]
```
AppName
```
    Practise
```

```
df_pyspark=spark.read.csv('Sales_data.csv')
```

```
df_pyspark=spark.read.option('header','true').csv('Sales_data.csv')
```

```
type(df_pyspark)
```

```
pyspark.sql.dataframe.DataFrame
def __init__(jdf: JavaObject, sql_ctx: Union['SQLContext', 'SparkSession'])

A distributed collection of data grouped into named columns.

.. versionadded:: 1.3.0

   versionchanged:: 3.4.0
```

```
df_pyspark.printSchema()
```

```
    root
     |-- Sr.No.: string (nullable = true)
     |-- Order ID: string (nullable = true)
     |-- Product: string (nullable = true)
     |-- Quantity Ordered: string (nullable = true)
     |-- Price Each: string (nullable = true)
     |-- Order Date: string (nullable = true)
     |-- Purchase Address: string (nullable = true)
```

```
 |-- Month: string (nullable = true)
 |-- Sales: string (nullable = true)
 |-- City: string (nullable = true)
 |-- Hour: string (nullable = true)
```

```
df_pyspark.select(['Product','Order ID'])
```

```
DataFrame[Product: string, Order ID: string]
```

```
df_pyspark.select(['Product','Order ID']).show()
```

```
+--------------------+--------+
|             Product|Order ID|
+--------------------+--------+
|  Macbook Pro Laptop|  295665|
|  LG Washing Machine|  295666|
|USB-C Charging Cable|  295667|
|    27in FHD Monitor|  295668|
|USB-C Charging Cable|  295669|
|AA Batteries (4-p...|  295670|
|USB-C Charging Cable|  295671|
|USB-C Charging Cable|  295672|
|Bose SoundSport H...|  295673|
|AAA Batteries (4-...|  295674|
|USB-C Charging Cable|  295675|
|     ThinkPad Laptop|  295676|
|AA Batteries (4-p...|  295677|
|AAA Batteries (4-...|  295678|
|USB-C Charging Cable|  295679|
|Lightning Chargin...|  295680|
|        Google Phone|  295681|
|USB-C Charging Cable|  295681|
|Bose SoundSport H...|  295681|
|    Wired Headphones|  295681|
+--------------------+--------+
only showing top 20 rows
```

```
df_pyspark['Product']
```

```
Column<'Product'>
```

```
df_pyspark.dtypes
```

```
[('Sr.No.', 'string'),
 ('Order ID', 'string'),
 ('Product', 'string'),
 ('Quantity Ordered', 'string'),
 ('Price Each', 'string'),
 ('Order Date', 'string'),
 ('Purchase Address', 'string'),
 ('Month', 'string'),
 ('Sales', 'string'),
 ('City', 'string'),
 ('Hour', 'string')]
```

```
### Adding columns in data frame
df_pyspark=df_pyspark.withColumn('Sales After 1 year',df_pyspark['Sales']*2)
```

```
df_pyspark.show()
```

```
+------+--------+--------------------+----------------+----------+----------------+--------------------+-----+------+-------------+-
|Sr.No.|Order ID|             Product|Quantity Ordered|Price Each|      Order Date|    Purchase Address|Month| Sales|         City|H
+------+--------+--------------------+----------------+----------+----------------+--------------------+-----+------+-------------+-
|     0|  295665|  Macbook Pro Laptop|               1|      1700|30-12-2019 00:01|136 Church St, Ne...|   12|  1700|New York City|
|     1|  295666|  LG Washing Machine|               1|       600|29-12-2019 07:03|562 2nd St, New Y...|   12|   600|New York City|
|     2|  295667|USB-C Charging Cable|               1|     11.95|12-12-2019 18:21|277 Main St, New ...|   12| 11.95|New York City|
|     3|  295668|    27in FHD Monitor|               1|    149.99|22-12-2019 15:13|410 6th St, San F...|   12|149.99|San Francisco|
|     4|  295669|USB-C Charging Cable|               1|     11.95|18-12-2019 12:38|43 Hill St, Atlan...|   12| 11.95|      Atlanta|
|     5|  295670|AA Batteries (4-p...|               1|      3.84|31-12-2019 22:58|200 Jefferson St,...|   12|  3.84|New York City|
|     6|  295671|USB-C Charging Cable|               1|     11.95|16-12-2019 15:10|928 12th St, Port...|   12| 11.95|     Portland|
|     7|  295672|USB-C Charging Cable|               2|     11.95|13-12-2019 09:29|813 Hickory St, D...|   12|  23.9|       Dallas|
|     8|  295673|Bose SoundSport H...|               1|     99.99|15-12-2019 23:26|718 Wilson St, Da...|   12| 99.99|       Dallas|
|     9|  295674|AAA Batteries (4-...|               4|      2.99|28-12-2019 11:51|77 7th St, Dallas...|   12| 11.96|       Dallas|
|    10|  295675|USB-C Charging Cable|               2|     11.95|13-12-2019 13:52|594 1st St, San F...|   12|  23.9|San Francisco|
|    11|  295676|     ThinkPad Laptop|               1|    999.99|28-12-2019 17:19|410 Lincoln St, L...|   12|999.99|  Los Angeles|
|    12|  295677|AA Batteries (4-p...|               2|      3.84|20-12-2019 19:19|866 Pine St, Bost...|   12|  7.68|       Boston|
|    13|  295678|AAA Batteries (4-...|               2|      2.99|06-12-2019 09:38|187 Lincoln St, D...|   12|  5.98|       Dallas|
|    14|  295679|USB-C Charging Cable|               1|     11.95|25-12-2019 09:39|902 2nd St, Dalla...|   12| 11.95|       Dallas|
|    15|  295680|Lightning Chargin...|               1|     14.95|01-12-2019 14:30|338 Main St, Aust...|   12| 14.95|       Austin|
|    16|  295681|        Google Phone|               1|       600|25-12-2019 12:37|79 Elm St, Boston...|   12|   600|       Boston|
|    17|  295681|USB-C Charging Cable|               1|     11.95|25-12-2019 12:37|79 Elm St, Boston...|   12| 11.95|       Boston|
|    18|  295681|Bose SoundSport H...|               1|     99.99|25-12-2019 12:37|79 Elm St, Boston...|   12| 99.99|       Boston|
|    19|  295681|    Wired Headphones|               1|     11.99|25-12-2019 12:37|79 Elm St, Boston...|   12| 11.99|       Boston|
+------+--------+--------------------+----------------+----------+----------------+--------------------+-----+------+-------------+-
only showing top 20 rows
```

```
### Drop the columns
df_pyspark=df_pyspark.drop('Sales After 1 year')



df_pyspark.show()
```

```
+------+--------+--------------------+----------------+----------+----------------+--------------------+-----+------+--------------+-
|Sr.No.|Order ID|             Product|Quantity Ordered|Price Each|      Order Date|    Purchase Address|Month| Sales|          City|H
+------+--------+--------------------+----------------+----------+----------------+--------------------+-----+------+--------------+-
|     0|  295665|  Macbook Pro Laptop|               1|      1700|30-12-2019 00:01|136 Church St, Ne...|   12|  1700| New York City|
|     1|  295666|  LG Washing Machine|               1|       600|29-12-2019 07:03|562 2nd St, New Y...|   12|   600| New York City|
|     2|  295667|USB-C Charging Cable|               1|     11.95|12-12-2019 18:21|277 Main St, New ...|   12| 11.95| New York City|
|     3|  295668|    27in FHD Monitor|               1|    149.99|22-12-2019 15:13|410 6th St, San F...|   12|149.99| San Francisco|
|     4|  295669|USB-C Charging Cable|               1|     11.95|18-12-2019 12:38|43 Hill St, Atlan...|   12| 11.95|       Atlanta|
|     5|  295670|AA Batteries (4-p...|               1|      3.84|31-12-2019 22:58|200 Jefferson St,...|   12|  3.84| New York City|
|     6|  295671|USB-C Charging Cable|               1|     11.95|16-12-2019 15:10|928 12th St, Port...|   12| 11.95|      Portland|
|     7|  295672|USB-C Charging Cable|               2|     11.95|13-12-2019 09:29|813 Hickory St, D...|   12|  23.9|        Dallas|
|     8|  295673|Bose SoundSport H...|               1|     99.99|15-12-2019 23:26|718 Wilson St, Da...|   12| 99.99|        Dallas|
|     9|  295674|AAA Batteries (4-...|               4|      2.99|28-12-2019 11:51|77 7th St, Dallas...|   12| 11.96|        Dallas|
|    10|  295675|USB-C Charging Cable|               2|     11.95|13-12-2019 13:52|594 1st St, San F...|   12|  23.9| San Francisco|
|    11|  295676|     ThinkPad Laptop|               1|    999.99|28-12-2019 17:19|410 Lincoln St, L...|   12|999.99|   Los Angeles|
|    12|  295677|AA Batteries (4-p...|               2|      3.84|20-12-2019 19:19|866 Pine St, Bost...|   12|  7.68|        Boston|
|    13|  295678|AAA Batteries (4-...|               2|      2.99|06-12-2019 09:38|187 Lincoln St, D...|   12|  5.98|        Dallas|
|    14|  295679|USB-C Charging Cable|               1|     11.95|25-12-2019 09:39|902 2nd St, Dalla...|   12| 11.95|        Dallas|
|    15|  295680|Lightning Chargin...|               1|     14.95|01-12-2019 14:30|338 Main St, Aust...|   12| 14.95|        Austin|
|    16|  295681|        Google Phone|               1|       600|25-12-2019 12:37|79 Elm St, Boston...|   12|   600|        Boston|
|    17|  295681|USB-C Charging Cable|               1|     11.95|25-12-2019 12:37|79 Elm St, Boston...|   12| 11.95|        Boston|
|    18|  295681|Bose SoundSport H...|               1|     99.99|25-12-2019 12:37|79 Elm St, Boston...|   12| 99.99|        Boston|
|    19|  295681|    Wired Headphones|               1|     11.99|25-12-2019 12:37|79 Elm St, Boston...|   12| 11.99|        Boston|
+------+--------+--------------------+----------------+----------+----------------+--------------------+-----+------+--------------+-
only showing top 20 rows
```

```
df_pyspark = df_pyspark.withColumnRenamed("Sr.No.", "SrNo")

# Data cleaning
# Convert columns to appropriate data types
df_pyspark= df_pyspark.withColumn("Quantity Ordered", df_pyspark["Quantity Ordered"].cast(IntegerType()))
df_pyspark = df_pyspark.withColumn("Price Each", df_pyspark["Price Each"].cast(FloatType()))
df_pyspark = df_pyspark.withColumn("Sales", df_pyspark["Sales"].cast(FloatType()))

# Handling missing values
df_pyspark = df_pyspark.dropna()

# Removing duplicates
df_pyspark = df_pyspark.dropDuplicates()

#  Calculate total sales amount for each product
product_sales = df_pyspark.groupBy("Product").agg(sum("Sales").alias("TotalSales"))

# Output the results to a new CSV file with overwrite mode
product_sales.coalesce(1).write.mode("overwrite").csv("output_path", header=True)


# Stop SparkSession
spark.stop()
```

2.2 Spark with BigData (Google Play Store Data)

Link to Databricks Notebook

Untitled Notebook 2024-03-17 00:20:50

(https://databricks.com)

Import Library

```
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql.types import StructType,StructField, StringType, IntegerType
from pyspark.sql.functions import *
```

Init Dataframe

```
df = spark.read.load('/FileStore/tables/googlestore.csv',format='csv',header='true',sep=',',escape="'" )
```

```
df.count()
```

```
Out[3]: 10841
```

```
df.show(10)
```

```
+--------------------+--------------+------+-------+----+-----------+----+-----+--------------+--------------------+------------------+------
------------+------------+
|                 App|      Category|Rating|Reviews|Size|   Installs|Type|Price|Content Rating|              Genres|      Last Updated|
Current Ver| Android Ver|
+--------------------+--------------+------+-------+----+-----------+----+-----+--------------+--------------------+------------------+------
------------+------------+
|Photo Editor & Ca...|ART_AND_DESIGN|   4.1|    159| 19M|    10,000+|Free|    0|      Everyone|        Art & Design|   January 7, 2018|
1.0.0|4.0.3 and up|
| Coloring book moana|ART_AND_DESIGN|   3.9|    967| 14M|   500,000+|Free|    0|      Everyone|Art & Design;Pret...|  January 15, 2018|
2.0.0|4.0.3 and up|
|U Launcher Lite â...|ART_AND_DESIGN|   4.7|  87510|8.7M| 5,000,000+|Free|    0|      Everyone|        Art & Design|    August 1, 2018|
1.2.4|4.0.3 and up|
|Sketch - Draw & P...|ART_AND_DESIGN|   4.5| 215644| 25M|50,000,000+|Free|    0|          Teen|        Art & Design|      June 8, 2018|Varies
with device|  4.2 and up|
|Pixel Draw - Numb...|ART_AND_DESIGN|   4.3|    967|2.8M|   100,000+|Free|    0|      Everyone|Art & Design;Crea...|     June 20, 2018|
1.1|  4.4 and up|
|Paper flowers ins...|ART_AND_DESIGN|   4.4|    167|5.6M|    50,000+|Free|    0|      Everyone|        Art & Design|    March 26, 2017|
1|  2.3 and up|
|Smoke Effect Phot...|ART_AND_DESIGN|   3.8|    178| 19M|    50,000+|Free|    0|      Everyone|        Art & Design|    April 26, 2018|
1.1|4.0.3 and up|
|    Infinite Painter|ART_AND_DESIGN|   4.1|  36815| 29M| 1,000,000+|Free|    0|      Everyone|        Art & Design|     June 14, 2018|
```

Schema

```
df.printSchema()
```

```
root
 |-- App: string (nullable = true)
 |-- Category: string (nullable = true)
 |-- Rating: string (nullable = true)
 |-- Reviews: string (nullable = true)
 |-- Size: string (nullable = true)
 |-- Installs: string (nullable = true)
 |-- Type: string (nullable = true)
 |-- Price: string (nullable = true)
 |-- Content Rating: string (nullable = true)
 |-- Genres: string (nullable = true)
 |-- Last Updated: string (nullable = true)
 |-- Current Ver: string (nullable = true)
 |-- Android Ver: string (nullable = true)
```

Data Cleaning

```
# Since Size of the app is a non significant attribute. we can drop that

df= df.drop("Size","Content Rating");
```

```
df.printSchema()
df.show(1)
```

```
root
 |-- App: string (nullable = true)
 |-- Category: string (nullable = true)
 |-- Rating: string (nullable = true)
 |-- Reviews: string (nullable = true)
 |-- Installs: string (nullable = true)
 |-- Type: string (nullable = true)
 |-- Price: string (nullable = true)
 |-- Genres: string (nullable = true)
 |-- Last Updated: string (nullable = true)
 |-- Current Ver: string (nullable = true)
 |-- Android Ver: string (nullable = true)

+--------------------+--------------+------+-------+--------+----+-----+------------+---------------+-----------+------------+
|                 App|      Category|Rating|Reviews|Installs|Type|Price|      Genres|   Last Updated|Current Ver| Android Ver|
+--------------------+--------------+------+-------+--------+----+-----+------------+---------------+-----------+------------+
|Photo Editor & Ca...|ART_AND_DESIGN|   4.1|    159| 10,000+|Free|    0|Art & Design|January 7, 2018|      1.0.0|4.0.3 and up|
+--------------------+--------------+------+-------+--------+----+-----+------------+---------------+-----------+------------+
only showing top 1 row
```

```
df=df.withColumn("Reviews",col("Reviews").cast(IntegerType()))
```

```
df.printSchema()
```

```
root
 |-- App: string (nullable = true)
 |-- Category: string (nullable = true)
 |-- Rating: string (nullable = true)
 |-- Reviews: integer (nullable = true)
 |-- Installs: string (nullable = true)
 |-- Type: string (nullable = true)
 |-- Price: string (nullable = true)
 |-- Genres: string (nullable = true)
 |-- Last Updated: string (nullable = true)
 |-- Current Ver: string (nullable = true)
 |-- Android Ver: string (nullable = true)
```

```
df.show(5)
```

```
+--------------------+--------------+------+-------+-----------+----+-----+--------------------+----------------+------------------+------------+
|                 App|      Category|Rating|Reviews|   Installs|Type|Price|              Genres|    Last Updated|       Current Ver| Android Ver|
+--------------------+--------------+------+-------+-----------+----+-----+--------------------+----------------+------------------+------------+
|Photo Editor & Ca...|ART_AND_DESIGN|   4.1|    159|    10,000+|Free|    0|        Art & Design| January 7, 2018|             1.0.0|4.0.3 and up|
| Coloring book moana|ART_AND_DESIGN|   3.9|    967|   500,000+|Free|    0|Art & Design;Pret...|January 15, 2018|             2.0.0|4.0.3 and up|
|U Launcher Lite â...|ART_AND_DESIGN|   4.7|  87510| 5,000,000+|Free|    0|        Art & Design|  August 1, 2018|             1.2.4|4.0.3 and up|
|Sketch - Draw & P...|ART_AND_DESIGN|   4.5| 215644|50,000,000+|Free|    0|        Art & Design|    June 8, 2018|Varies with device|  4.2 and up|
|Pixel Draw - Numb...|ART_AND_DESIGN|   4.3|    967|   100,000+|Free|    0|Art & Design;Crea...|   June 20, 2018|               1.1|  4.4 and up|
+--------------------+--------------+------+-------+-----------+----+-----+--------------------+----------------+------------------+------------+
only showing top 5 rows
```

```
df.createOrReplaceTempView("apps")
```

```
%sql select * from apps
```

Table

	App	Category	Rating	Review
1	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159
2	Coloring book moana	ART_AND_DESIGN	3.9	967
3	U Launcher Lite â€“ FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510
4	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	21564
5	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967
6	Paper flowers instructions	ART_AND_DESIGN	4.4	167
7	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178

10,000 rows | Truncated data

Top Reviews Given

```
%sql select App,sum(Reviews) from apps
group by App
order by sum(Reviews) desc
```

Table

	App	sum(Reviews)
1	Instagram	266241989
2	WhatsApp Messenger	207348304
3	Clash of Clans	179558781
4	Messenger â€“ Text and Video Chat for Free	169932272
5	Subway Surfers	166331958
6	Candy Crush Saga	156993136
7	Facebook	156286514

9,660 rows

```
df=df.withColumn("Installs",col("Installs").cast(IntegerType()))
```

Top 10 Installed Apps

```
%sql
select App,installs from apps
order by installs desc
limit 10
```

Table

	App	installs
1	Life Made WI-Fi Touchscreen Photo Frame	Free
2	Viber Messenger	500,000,000+
3	imo free video calls and chat	500,000,000+
4	imo free video calls and chat	500,000,000+
5	Google Duo - High Quality Video Calls	500,000,000+
6	UC Browser - Fast Download Private & Secure	500,000,000+
7	imo free video calls and chat	500,000,000+

10 rows

Top Paid Apps

Table

	App	Price
1	ðŸ’Ž I'm rich	$399.99
2	Ð‘Ð¸Ð»ÐµÑ‚Ñ‹ ÐŸÐ”Ð” CD 2019 PRO	$1.49
3	Ã‰galitÃ© et RÃ©conciliation	$2.99
4	Â¡Ay Caramba!	$1.99
5	weather HD	$1.99
6	tTorrent - ad free	$1.99
7	sugar, sugar	$1.20

800 rows