Final Project M3

CS 327E
November 13, 2017
1) What is the core abstraction in Spark SQL?

A) MapReduce
B) RDDs
C) DataFrames
D) Catalyst
2) Like RDDs, DataFrames are immutable.

A) True
B) False
3) Unlike RDDs, DataFrames require ___________

A) Schemas
B) Relational Databases
C) User Defined Functions
D) User Defined Types
4) How can you explore the data that is in a DataFrame?

A) With SQL expressions
B) Using the DataFrame API
C) By converting the DataFrame to an RDD and then running RDD operations over the RDD
D) All of the above
5) How many rows are returned by `distinct_table.show()`?

A) 2  
B) 3  
C) 4
Querying Postgres from Python

```python
conn = psycopg2.connect(database=rds_database, user=rds_user, password=rds_password, host=rds_host, port=rds_port)

select_stmt = "select count(*) from title_basic where start_year = 2017"

try:
    cur = conn.cursor()
    cur.execute(select_stmt)

    row = cur.fetchone()

    if row != None:
        record_count = row[0]
        print("record count is: " + str(record_count))

except Exception as e:
    print("Exception: " + e.message)

cur.close()
conn.close()
```
Querying Postgres from Python

```
conn = psycopg2.connect(database=rds_database, user=rds_user, password=rds_password, host=rds_host, port=rds_port)

select_stmt = "select title_id, primary_title from title basics where start_year = 2017"

try:
    cur = conn.cursor()
    cur.execute(select_stmt)

    rows = cur.fetchall()

    for row in rows:
        title_id = row[0]
        primary_title = row[1]
        print("title_id: " + title_id + ", primary_title: " + primary_title)

except Exception as e:
    print("Exception: " + e.message)

cur.close()
conn.close()
```
Milestone 3

http://www.cs.utexas.edu/~scohen/projects/m3-assignment.pdf