

## Quiz 8 - Spark SQL and Spark Streaming

1. What does the following filter line of code do?

```
df.filter(df["teamlevel"] > 1)
```

- **Filter each row to show only team levels larger than 1.**
- Filter each column to show only team levels larger than 1.
- Select the first two columns of the data and filter each column to show only team levels larger than 1.
- Select the first two columns of the data and displays only team levels greater than 1.

2. What does the following do?

```
df.select("userid", "teamlevel").show(5)
```

- Select the rows named "userid" and "teamlevel" and display first 5 rows.
- Display all rows except "userid" and "teamlevel".
- **Select the columns named "userid" and "teamlevel" and display first 5 rows.**
- Display all columns except "userid" and "teamlevel".

3. What does the 1 represent in the following line of code?

```
ssc = StreamingContext(sc,1)
```

- To create only one partition to manage the stream.
- To specific debug output.
- To create one single context.
- **A batch interval of 1 second.**

#### 4. What does the following code do?

```
window = vals.window(10, 5)
```

- **Creates a window that combines 10 seconds worth of data and moves by 5 seconds.**
- Creates 10 windows with 5 seconds worth of data in them.
- Creates 10 windows with 5 batch intervals inbetween.
- Creates a batch interval between 10 seconds and 5 seconds.