ELG 5166 – Cloud Analytics Exam

Personal Ethics & Academic Integrity Statement

Student name: **Hosam Mahmoud Ibrahim Mahmoud** Student ID: **300327269**

By typing in my name and student ID on this form and submitting it electronically, I am attesting to the fact that I have reviewed my work in its entirety.

I attest to the fact that this submission adheres to the fraud policies as outlined in the Academic Regulations in the University's Undergraduate Studies Calendar. I further attest that I have knowledge of and have respected the "Beware of Plagiarism" brochure found on the University of Ottawa website.

I, by typing in my name and student ID on this form and submitting it electronically,

- warrant that the work submitted herein is your work and not the work of others.
- acknowledge that I have read and understood the University Regulations on Academic Misconduct.
- acknowledge that it is a breach of University Regulations to give or receive unauthorized and/or unacknowledged assistance on a graded piece of work.

Q1)

- A **service level agreement** typically describes the parties involved, the nature of the service rendered, quality or performance metrics for the services, how and when the contract may be terminated, and how to resolve service quality issues.
- In their SLAs, most cloud service providers, for example, frequently guarantee 99.99% (often referred to as four nines) or even 99.999% uptime for their services. And if they don't deliver, they'll give service credits (points that can be redeemed for cloud services).
  Reference: [What is a Service Level Agreement (SLA) in IT Service Delivery (invgate.com)]

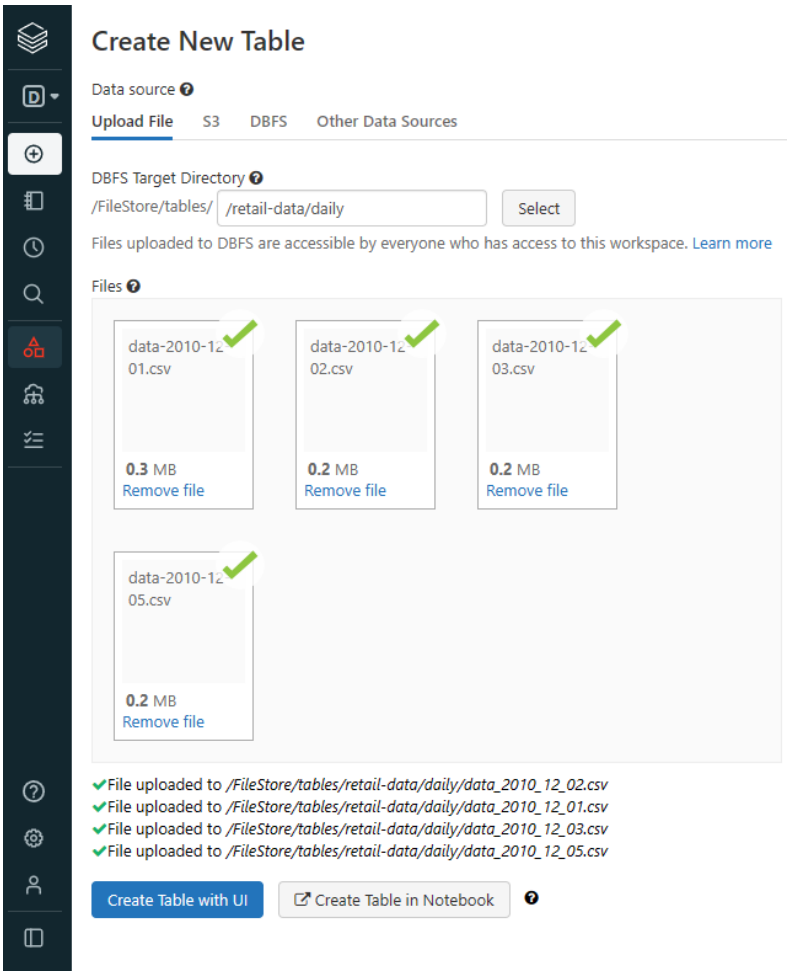
- **There will be 2 cases:**

1- The first case that the **entire cluster** will be down 24 days in the year.
   - which mean that the **Downtime** will be $\left(\frac{24}{365}\right) * \mathbf{100} = \mathbf{6.575}\%$
   - So, the **Uptime** will be $(1 - \mathbf{Downtime}) * \mathbf{100} = \mathbf{93.424657}\%$
   - Which mean that the **Availability = 1 nine.**

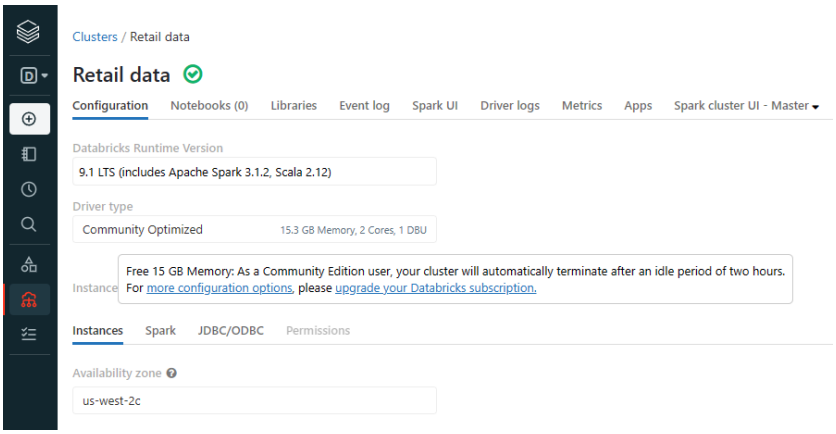
2- The second case that **each node in the cluster** is expected to have up to 24 days Downtime in the year.
   - which mean that the **Downtime** will be $\left(\frac{24}{365}\right) * \left(\frac{24}{365}\right) * \left(\frac{24}{365}\right) * \left(\frac{24}{365}\right) * \left(\frac{24}{365}\right) * \mathbf{100} = \mathbf{0.000122911}\%$
   - So, the **Uptime** will be $(1 - \mathbf{Downtime}) * \mathbf{100}$ **= 99.99987%**
   - Which mean that the **Availability = 5 nines.**

Q2)

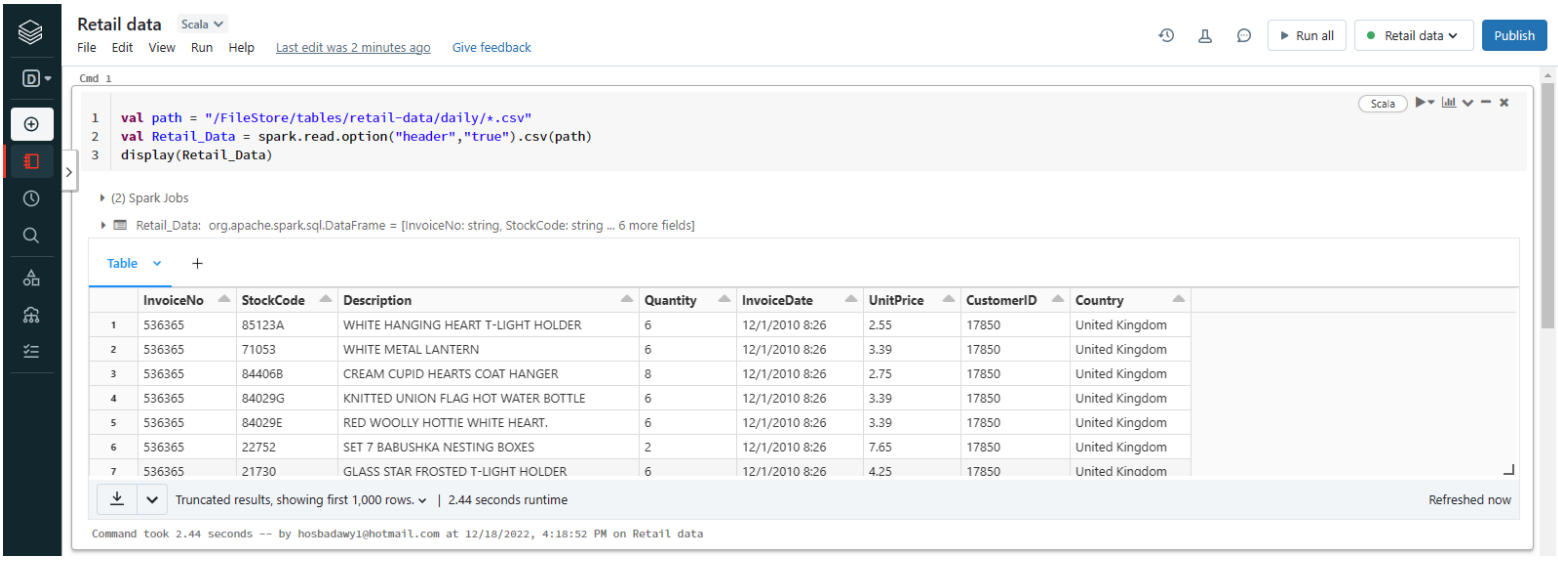
- Upload the csv files into my account:



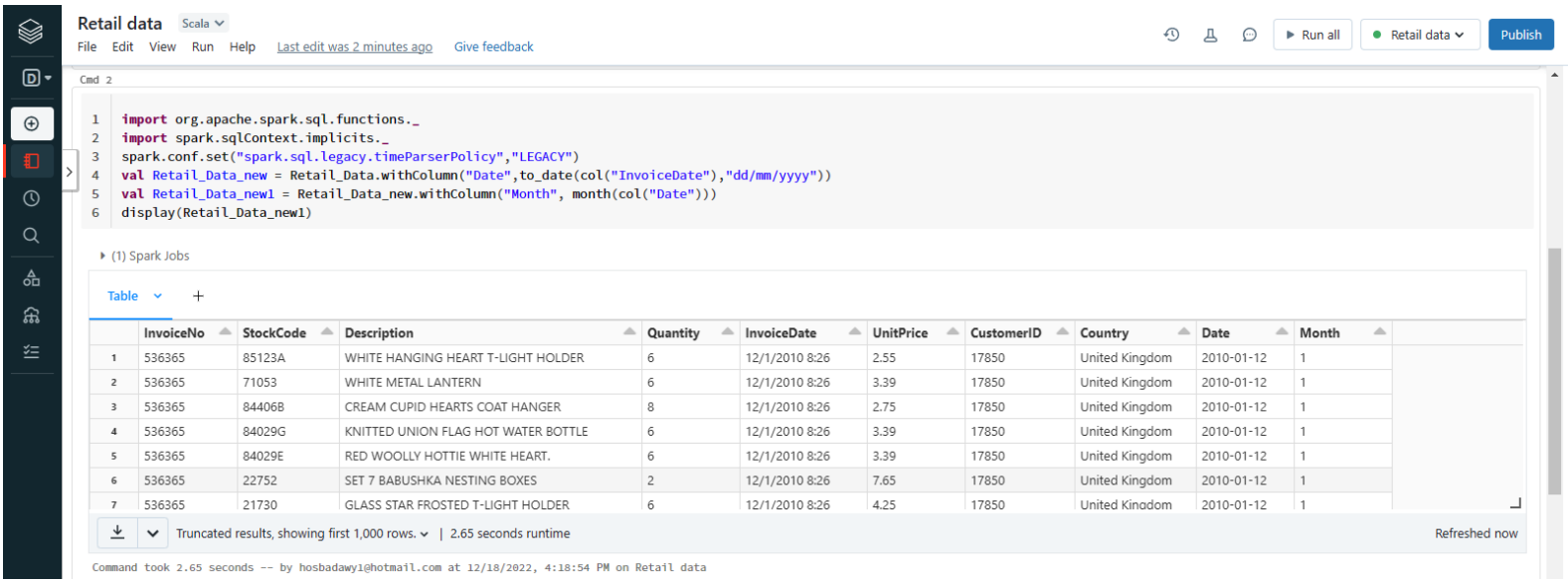
- Creating Cluster.



- Read the data:



- Create the Month column to be able to aggregate



- Aggregation:

Q3)

1- **Calculate the number of nodes.**

- **The disk size = 512GB,** and this disk has a used storage which is **112GB, so HDFS** will take the remaining empty storage space which is 400GB.
- **HDFS = disk size – the used storage = 512GB - 112GB = 400GB**
- **number of nodes = ((the size of data needs to be procced (in giga byte)) * replication factor) / available size of HDFS.**
- **number of nodes = ((300*1024) *3)/400 = 2304 node.**
- It's required to be **2304 node** within the Hadoop YARN cluster to store the entire 300 TB of data in HDFS.

2- Calculate the number partitions and nodes.

- Log files need to be procced = **640GB**
- memory allocation per node **= 64GB memory - 14GB memory = 50GB**
- Num of nodes = **log files / memory allocation per node** = **640GB / 50GB = 12.8 = 13 node**
- Number of partitions = **number of containers per node * number of partitions on each container * number of nodes**
- Number of partitions = **4 * 1 (assume that there is 1 partition in each container) * 13 = 52 partitions**

Q4)

1- Calculate the number of sensors

- The Number of sensors **= meters / 1.5 Km for each sensor**
- The Number of sensors = **6000/ 1.5 = 4000 sensors**

2- Calculate the number of events and partitions

- As it's mentioned in the question the sensor receives an event each 30 seconds, which mean that the senor receives **2 events per minute.**
- **The Total number of events for all sensors per minute = 4000 * 2 = 8000 events per minute**
- **Then Total number of partitions = 8000/500 = 16 partitions**

3- Calculate the number of events hubs and namespaces.

- I will assume that we will use basic tier according to Microsoft tiers which has 10 event hubs per namespace, and 32 partitions per event hub.
- Then Number of event hubs = **number of partitions / partitions per event hub = 16 / 32 = 0.5 = 1 event hub**
- Then Number of namespaces = **number of event hubs / event hubs per namespace = 1/10 = 0.1 = 1 namespace**

4- ASA Job queries:

- **ASA job query for below 800 psi (leak alert)**

```
%sql
with input_data as(
 select *
 from [event_hub]
 TIMESTAMP by process_date
),

PSI_800_data as(
 select *
 from [input_data]
 where PressureReading < 800
)

select COUNT(*), SerialNumber
into
 [output_data]
from
 [PSI_800_data]
group by SerialNumber, SlidingWindow(minute,5)
having COUNT (*) >= 3
```

- **ASA job query for above 1200 psi (obstruction or blockage downstream)**

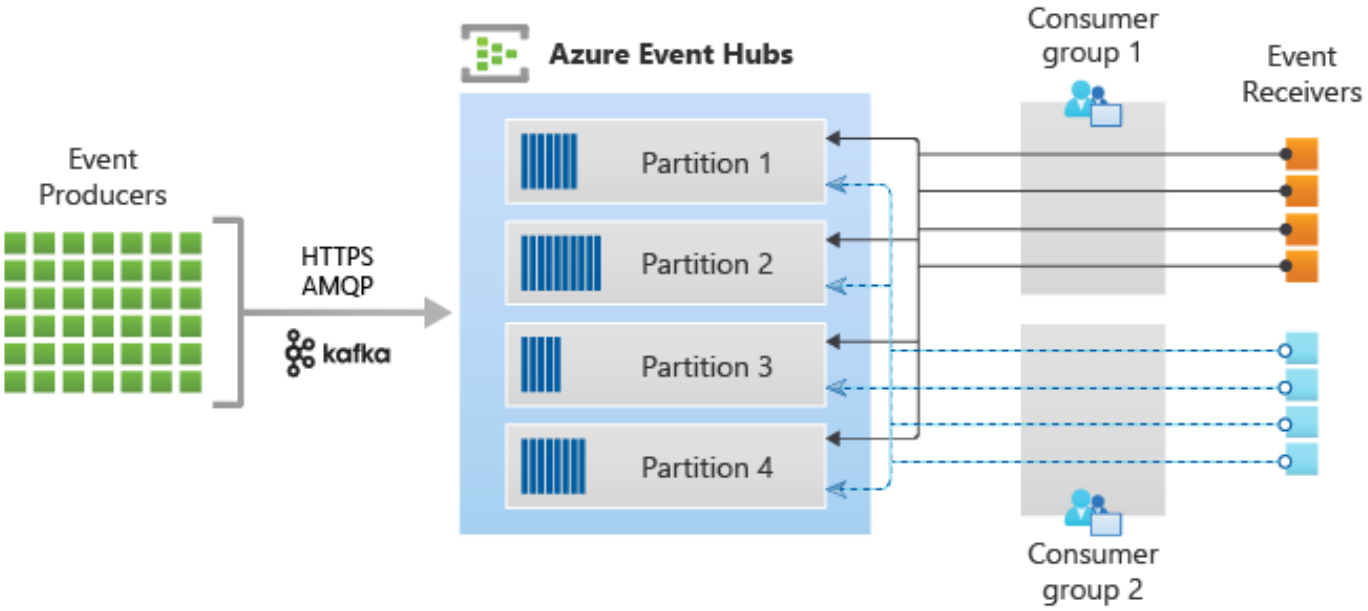
```
WITH input_data AS(
 SELECT *
 FROM [event_hub]
 TIMESTAMP BY process_date
),

PSI_1200_data AS(
 SELECT *
 FROM [input_data]
 WHERE PressureReading > 1200
)

SELECT COUNT(*), SerialNumber
INTO
 [output_data]
FROM
 [PSI_1200_data]
GROUP BY SerialNumber, SlidingWindow(minute,5)
HAVING COUNT (*) >= 3
```

5- Number of consumers groups, their roles, and what to do to support receivers:

- There must be **two consumers for each adverse event**, because there is 2 type of events, one for leakage and the other one for blockage.
- The consumer groups are responsible for allowing multiple consuming applications to have their own view of the event stream and read it independently at their own pace and offsets.
- Distribute the events across all partitions, and when publishing events on a regular basis, use the AMQP protocol whenever possible.



Preferred Event Hubs stream processing architecture.

Reference: [Overview of features - Azure Event Hubs - Azure Event Hubs | Microsoft Learn]