Structured Streaming profiling through SparkLens

Sparklens is very useful for tuning Spark Applications and also for EMR Cluster size estimations. Currently it only supports batch applications but can be extended to Spark Structured streaming.

At Intuit, we started using SparkLens for batch Jobs and working on structured streaming feature support (we are planning to contribute back to Sparklens)

As structured streaming applications will be running continusouly, so we cannot directly use current *com.qubole.sparklens.QuboleJobListener*, because streaming application will be running continusly.

Each streaming Application consists of multiple batches based on Triggers and each batch might consists of `m` number of Jobs launched for each Trigger(Batch).

So, If we use current *com.qubole.sparklens.QuboleJobListener*, then it will provide job stats for huge number of jobs after terminating the streaming query, but users/developers will be interested in getting these reports continusly for every few batches.

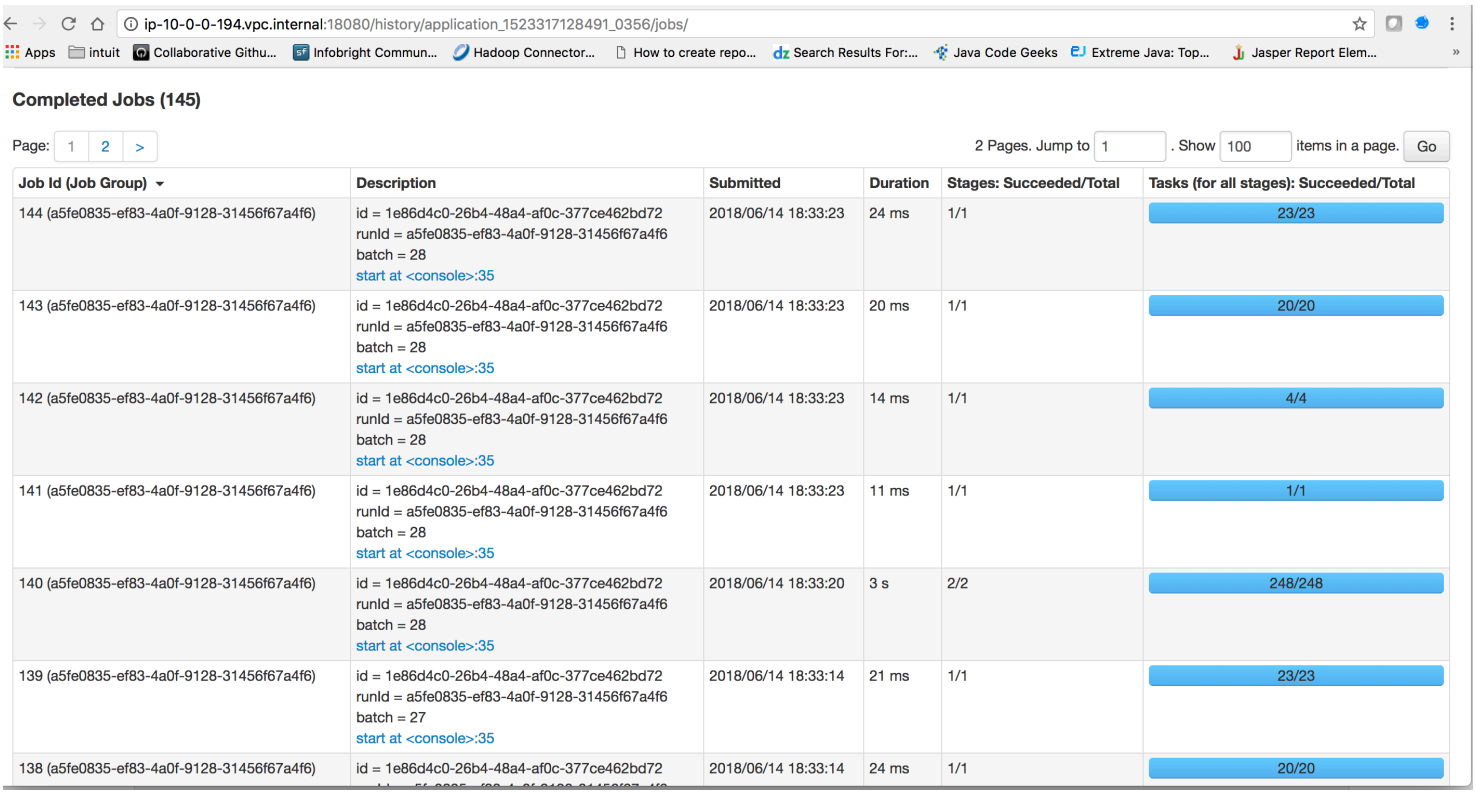
Expectations from Structured streaming profiling

1. Keep providing these job/stage stats for every `k` batches. (Including batch wise stats)
2. Analysis of cluster usage for each batch compare to previous batch
3. InputRate vs ProcessingRate for each batch
4. If there is any possibility, get the above stats for each Job group including Job group name instead of just Job ID. So that users can identify and tune performance of high latency block of code.

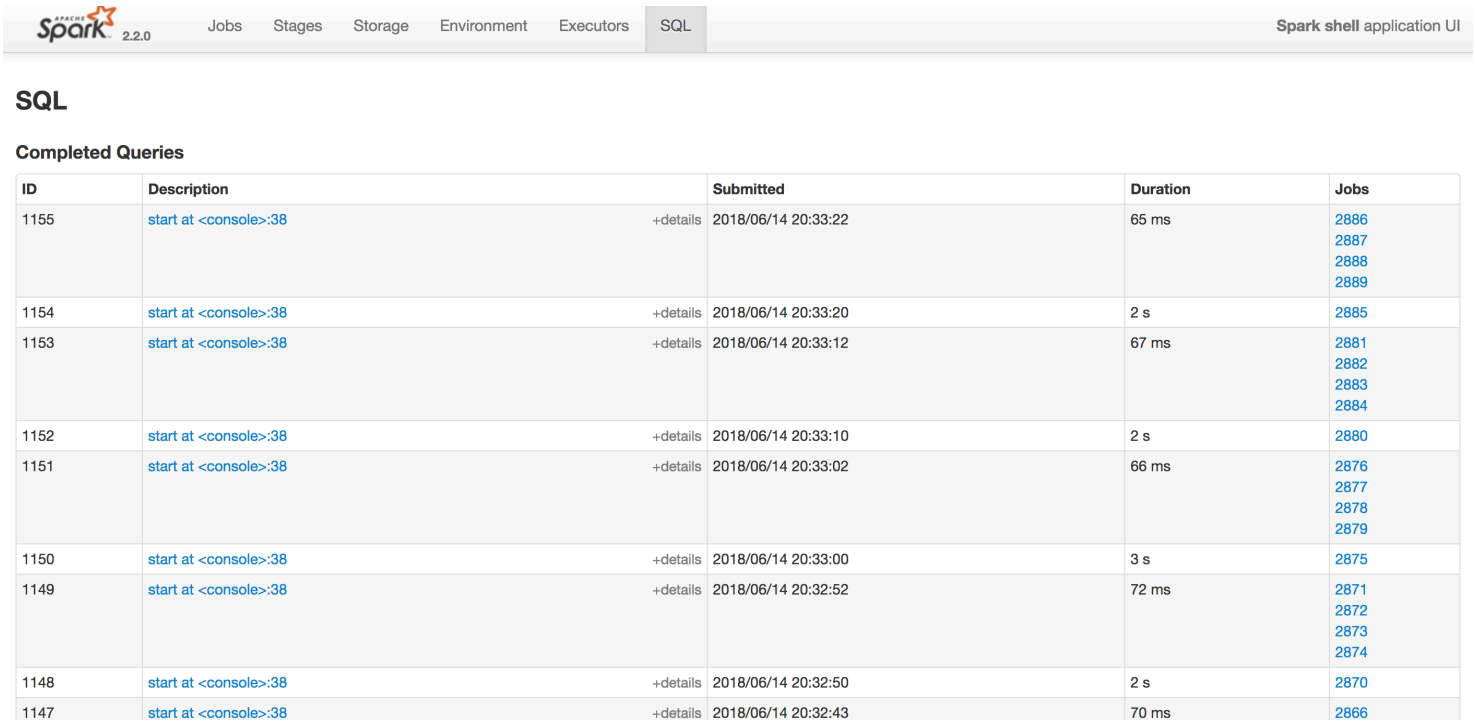
For Structured streaming Spark Inbuilt *org.apache.spark.sql.streaming.StreamingQueryListener* provides below Events

1. QueryStartedEvent
2. QueryProgressEvent
3. QueryTerminatedEvent

If we observe the Structured streaming Web UI, we can see each batch can have multiple Jobs as below.



And it's corresponding SQL Web UI is as below



The structured streaming sparklens proposed output looks like below.

```
Application ID
```

```
-     Batch 1
  -- Job 1
      --- Stage 0
      --- Stage 1
      :
      :
      --- Stage n
  -- Job 2
      --- Stage 0
      --- Stage 1
      :
      :
      --- Stage n

      :
      :

  -- Job n
      --- Stage 0
      --- Stage 1
      :
      :
      --- Stage n

  [ All SparkLens Metrics for above Jobs ]

  -- InputRate
  -- ProcessRate
  -- RelativeProcessingRate
  -- RelativeInputRate
  -- Other Stats


:
:
:

 -    Batch k
  -- Job 1
      --- Stage 0
      --- Stage 1
      :
      :
      --- Stage n
  -- Job 2
```

```
        --- Stage 0
        --- Stage 1
        :
        :
        --- Stage n

        :
        :

    -- Job n
        --- Stage 0
        --- Stage 1
        :
        :
        --- Stage n

    [ All SparkLens Metrics for above Jobs ]

    -- InputRate
    -- ProcessRate
    -- RelativeProcessingRate
    -- RelativeInputRate
    -- Other Stats


    [Cumulative metrics for K batches ]

    :
    :
    :
```