These are the time taken to read data from epsilon oracle database using a glue job . We tried with different combination to tune this job .

Test Case1:

* Not specified any partition column, num partitions

Job Name : pre-demo-job-1

|  |  |
| --- | --- |
| Job Run Id | jr\_6eeafacea527071632509aa518f051af9699d03d2e44f4d2cbaa5f5d33ef2c62 |
| Worker Type | G1.X |
| **Number of workers** | **40** |
| FETCHSIZE | 1000 |
| **Duration** | **2 Hr 52 Min** |
| Run status | Aborted |

Test Case2:

Job Name : pre-demo-job-1

|  |  |
| --- | --- |
| Job Run Id | jr\_2ac8d6575a77031165dc81a259b7c59177acca52e03c80f14a1b965a67146001 |
| Worker Type | G1.X |
| **Number of workers** | **40** |
| **partitionColumn** | **TXN\_ITEM\_ID** |
| **lowerBound** | **1255534593** |
| **upperBound** | **1829324174** |
| **numPartitions** | **40** |
| FETCHSIZE | 1000 |
| **Duration** | **7 Mins** |
| Run status | Succeeded |

Test Case3:

Job Name : pre-demo-job-1

|  |  |
| --- | --- |
| Job Run Id | jr\_75281a3fcb666af36da28bef393df7bf9b950035281098c11e927b0fef7a8416 |
| Worker Type | G1.X |
| **Number of workers** | **40** |
| **partitionColumn** | **TXN\_ITEM\_ID** |
| **lowerBound** | **1255534593** |
| **upperBound** | **1829324174** |
| **numPartitions** | **80** |
| FETCHSIZE | 1000 |
| **Duration** | **4 Mins** |
| Run status | Succeeded |

Test Case4 :

Job name : [pre\_spark\_jdbc\_tune\_pk](https://console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=jobs)

|  |  |
| --- | --- |
| Job Run Id | jr\_11631093be17bdd857929a8433b36013fa8e1aa09834bdc5b6339a0bed01b98b |
| Worker Type | G1.X |
| **Number of workers** | **40** |
| **partitionColumn** | **TXN\_ITEM\_ID** |
| **lowerBound** | **1255534593** |
| **upperBound** | **1829324174** |
| **numPartitions** | **40** |
| **FETCHSIZE** | **1000** |
| Duration |  |
| Run status | **Failed (Due to unavailability of private ip address in the subnet**  **Executors failed ) – as there was no available ips in the subnet** |

Test Case5:

Job name : [pre\_spark\_jdbc\_tune\_pk](https://console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=jobs)

|  |  |
| --- | --- |
| Job Run Id | [jr\_144baf04b3bd5b0790b5fc9dcc9760e6e2ef2227e8f1b32ffc3a70270a6ed6e7](https://console.aws.amazon.com/glue/home?region=us-east-1#jobRun:jobName=pre_spark_jdbc_tune_pk;jobRunId=jr_144baf04b3bd5b0790b5fc9dcc9760e6e2ef2227e8f1b32ffc3a70270a6ed6e7) |
| Worker Type | G1.X |
| **Number of workers** | **10** |
| **dbtable** | "(**SELECT \* FROM DM\_OWNER.LOFT\_F\_TXN\_ITEM\_VW**) subq" |
| **partitionColumn** | **TXN\_ITEM\_ID** |
| **lowerBound** | **1255534593** |
| **upperBound** | **1829324174** |
| **numPartitions** | **40** |
| **FETCHSIZE** | **1000** |
| **Duration** | **7 Mins** |
| Run status | Succeeded |

Test Case6:

Job name : [pre\_spark\_jdbc\_tune\_pk](https://console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=jobs)

|  |  |
| --- | --- |
| Job Run Id | jr\_3f633f4219f344d7cb298eeccc8dcd4840749c65d58a02f261c05443ac118570 |
| Worker Type | G1.X |
| **Number of workers** | **10** |
| **dbtable** | **"(SELECT TXN\_ITEM\_ID,TXN\_ID,TXN\_NBR,MODIFIED\_DT FROM "+ tableName +" where MODIFIED\_DT>=CURRENT\_DATE-4) subq"** |
| **partitionColumn** | **TXN\_ITEM\_ID** |
| **lowerBound** | **1255534593** |
| **upperBound** | **1829324174** |
| **numPartitions** | **40** |
| **FETCHSIZE** | **1000** |
| **Duration** | **7 Mins** |
| Run status | Succeeded |

Test Case7:

Job name : [pre\_spark\_jdbc\_tune\_pk](https://console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=jobs)

|  |  |
| --- | --- |
| Job Run Id | jr\_89f85b2046988a04009c254633b0cad23b2e0f2e109c614b028899f30bcf396d |
| Worker Type | G1.X |
| **Number of workers** | **10** |
| **dbtable** | **"(SELECT TXN\_ITEM\_ID,TXN\_ID,TXN\_NBR,MODIFIED\_DT FROM "+ tableName +" where MODIFIED\_DT>=CURRENT\_DATE-4) subq"** |
| **partitionColumn** | **TXN\_ITEM\_ID** |
| **lowerBound** | **1255534593** |
| **upperBound** | **1829324174** |
| **numPartitions** | **40** |
| FETCHSIZE | **10000** |
| **Duration** | **2 Mins** |
| Run status | Succeeded |

<https://spark.apache.org/docs/2.4.0/sql-data-sources-jdbc.html>

We are getting job failure intermittently due to availability of IP address . When one developer is running the job it may not cause an issue but when multiple developers will start working , this job failures will be more frequent . The only limitation at this current situation is , it will reduce the productivity .