**7th Dec 2022** **YARN**

* It is a Hadoop Tool.
* We can monitor the job while it is running, checking the logs files of the job and progress of the job.
* We have different clusters and each cluster we have different Yarns.
* When a job triggers in Appworx, then it will generate an Application ID in Yarn from which we can monitor whether the job is running or not.
* We have different of states
* Accepted – when the script is triggered but server is not getting the required resources to run the job
* Running – show the running jobs
* Finished – show the finished jobs
* Failed – show the failed jobs
* Killed – if we killed manually, it will show here
* In Schedular option, we can see the usage of the queue.
* Queues are nothing but we just dividing the jobs into different lines.
* Queue helps us to maintain the memory usage and don’t let big jobs to acquire more memory and slow down the other jobs.
* Queue is like waiting line.
* Queue is divided on the basis of user, from which user job is triggering.
* Developer has divided the queue.
* User indicates the application which queue it belongs to.
* On the basis of the jobs, we are resizing the Nodes, which helps to reduce the cost.
* We have 3 Application Type
* Spark – If the job is running in Pyspark technology, then it comes under Spark Application.
* Tez – It is Python Jobs
* MapReduce – It has SQOOP Jobs
* Application Priority – If we want any jobs to be finished earlier, then we set it.
* We usually don’t give application priority because in Appworx they have schedule the job in such a way that it will complete in order.
* Start Time – The time when the job is triggered in Appworx.
* Launch Time – The time when the job is triggered in Server Level.
* Running Containers – It shows how many AWS Containers (Storage Block in AWS) the job is using.
* By Clicking on the Application Master, we can directly see the progress of a job.
* In Spark, we can see the logs of the running job by clicking on the Execute Option.
* Mainly, we check after triggering the job in Appworx, whether the Application ID is generated or not, whether it is progressing or not, check the status of the job whether it is running or accepted state.
* Mostly, we check the progress of the Long Running Jobs.
* Check the resize of the clusters.

Prod Alert – Mirror Layer

Sanity Validation - Consumption

Reconciliation Alert – QDI related Alerts

Sbdt Alert – If any tables fail, we’ll get the alerts regarding which table got failed and the Error.

Hive Lock Alert – If any table gets lock for more than 30 mins, then we’ll get this alert.

**9th Dec 2022**

**QDI (Qlik Data Intregation)**

* QDI stands for Qlik Data Intregation
* QDI is a ETL Tool, used to pull the data from source to our environment.
* QDI will connects to transactional logs instead of database and get the data but for that we need to make sure that at the source end, Supplementary logs should be enabled.
* Supplementary log -
* There are two types of tasks.
* CDC and Full Load task
* CDC will capture the changes from source.
* Full Load Task – Whenever we want to reload the data or capture the data of 2-3 days.
* 27/7 CDC task should be in running state.
* We can see all warnings and failure logs in View Logs
* View logs will help to monitor all the logs whether it is capturing data from source or not.
* In QDI we have to see load type it means how the table was loading so we have 3 types of loading the data into the table. They are...

It has three+1 types

1. Full Load – Connects to the source, truncate all the data and update.
2. Incremental – Update to the latest data.
3. conditional – based on given condition it loading the table
4. Append-it means directly adding the new data to existing one.

**12/12/2022:**

1.QDI we have 5 channel related to CDC Those are

\*odw,\*rmd,\*e-service,\*proficy,\*proficygrr

\* These are the main five channel we are monitor in qdi and a play button we saw on channel’s that means that channel will be in running state and square symbol means that that channel will be in stopped state.

\*Want to create a new channel means we have the option--” +**new task" when** click on this

We get few fields to fill and click on load type and click on ok then channel will created

And that channel will be in stopped state.

\*After select the channel we have to see the warnings and error occurred while running

And we have an option to monitor in this option we saw running tables and want to reload the table means “select the table –click on reload –go to run option we have 3 options in those 3 we click on resume running and it will start running

And we have designer tab in that we need to filter for that task we can do that in designer

But if we need to run the task, we need to be in monitor tab …

\*QDI has full load task as well we just search full and click on enter then we get related full

Load task.