**Overwatch MultiWorkspace Deployment Guide**

Overwatch multi workspace deployment is driven by a config CSV .Customers need to fill up the csv file with all the required relevant details to initiate the deployment.

As an example customers can refer to the dummy CSV file provided by Databricks Overwatch Team.

**CSV Column Descriptions**

* **Workspace\_name**:
  + **Is Required**: True
  + **Description**: URL of the workspace
* **Workspace\_id**:
  + **Is Required**: True
  + **Description**: ID of the workspace(execute in scale dbutils.notebook.getContext.tags("orgId") to get the workspaceID)
* **Workspace\_url**:
  + **Is Required**: True
  + **Description**: URL of the workspace
* **Api\_url**:
  + **Is Required**: True
  + **Rules**:Should have a valid URL.
  + **Description**:

API URL for the Workspace (execute in scala dbutils.notebook.getContext().apiUrl.get to get the API URL for the workspace.NOTE: Workspace\_URL and API\_URL can be different for a workspace)

* **Cloud**:
  + **Is Required**: True
  + **Rules**:Should Be either Azure or AWS
  + **Description**: Cloud provider( Azure or AWS)
* **Primordial\_date**:
  + **Is Required**: True
  + **Rules**:Should be a valid date in yyyy-MM-dd format and should be less than or equal to current date
  + **Description** The date from which Overwatch will capture the details.(The format should be yyyy-MM-dd ex:2022-05-20)
* **Eetl\_storage\_prefix** :
  + **Is Required**: True
  + **Rules**:ETL Storage Prefix should be common across the workspaces.
  + **Description**: The mount location for Overwatch(It should be common for all)
* **Etl\_database\_name**:
  + **Is Required**: True
  + **Rules**:Workspaces should have a common ETL Database Name.
  + **Description**:The name of the ETL data base for Overwatch(It should be common for all)
* **Consumer\_database\_name**:
  + **Is Required**: True
  + **Rules**:Workspaces should have a common Consumer Database Name.
  + **Description**:The name of the Consumer database(It should be common for all)
* **Secret\_scope** :
  + **Is Required**: True
  + **Rules**:Secrete scope should not be empty.
  + **Description**: Secrate scope name(This should be created on the workspace on which Overwatch job will execute)
* **Secret\_key\_dbpat**:
  + **Is Required**: True
  + **Rules**:PAT key should not be empty.
  + **Description**: This will contain the PAT token of the workspace(The key should be present in the secret\_scope)
* **Auditlogprefix\_source\_aws**:
  + **Is Required**: True for AWS
  + **Rules**:Folder with Primordial date should be present inside AuditLogPrefix.
  + **Description**:Location of auditlog (Only for AWS) , ignore this field for Azure.
* **Eh\_name**:
  + **Is Required**: True for Azure
  + **Rules**:Data will be consumed from the Event hub
  + **Description**: Event hub name (Only for Azure ,The event hub will contain the audit logs of the workspace) ,ignore this field for AWS.
* **Eh\_scope\_key**:
  + **Is Required**: True for Azure
  + **Rules**:Should have a valid key to connect with the event hub.
  + **Description**: Key to connect with event hub(Only for Azure ,this key should be present in the secret scope),ignore this field for AWS.
* **Interactive\_dbu\_price**:
  + **Is Required**: True
  + **Description**:Price of the interactive dbu.
* **automated\_dbu\_price**:
  + **Is Required**: True
  + **Description**:Price of the automated dbu.
* **sql\_compute\_dbu\_price**:
  + **Is Required**: True
  + **Description**: Price of the automated dbu.
* **jobs\_light\_dbu\_price**:
  + **Is Required**: True
  + **Description**: Price of the automated dbu.
* **Max\_days**:
  + **Is Required**: True
  + **Rules**:Max Days should be a number.
  + **Description**:Incase of very old primordial date the max\_days will be use to batch the data(Recommendation: 30)
* **Excluded\_scopes**:
  + **Is Required**: True
  + **Rules**:Excluded scope can be audit:sparkEvents:jobs:clusters:clusterEvents:notebooks:pools:accounts:dbsql
  + **Description**:

Scopes that should not be performed in the run for that workspace.You can choose between audit,sparkEvents,jobs,clusters,clusterEvents,notebooks,pools,accounts,dbsql

(scopes should be separated by “:” .Note:Please don't fill this column if you want to run all the scopes)

* **Active**:
  + **Is Required**: True
  + **Rules**:Should be either true/false.
  + **Description**:If active this config row will be considered for deployment(Its a boolean value true means active and false means inactive)

**Advance Options**

* **proxy\_host**:
  + **Is Required**: False
  + **Rules**:Should be a valid url.
  + **Description**:Proxy url for the workspace.
* **proxy\_port**:
  + **Is Required**: False (True if proxy\_host is present)
  + **Rules**:Should be a valid port.
  + **Description**:Proxy port for the workspace.
* **proxy\_user\_name**:
  + **Is Required**: False
  + **Rules**:Should be a valid proxy user name.
  + **Description**:proxy user name for the workspace.
* **proxy\_password\_scope**:
  + **Is Required**: False (True if proxy\_user\_name is present)
  + **Rules**:Scope should be present in the workspace on which overwatch job will execute.
  + **Description**:Scope which contains the proxy password.
* **proxy\_password\_key**:
  + **Is Required**: False (True if proxy\_user\_name is present)
  + **Rules**:proxy\_password\_key should be present in proxy\_password\_scope.
  + **Description**:key which contains proxy password.
* **success\_batch\_size**:
  + **Is Required**: False (default value is 200)
  + **Description**:Overwatch makes multiple api calls and stores the interim results to a temp location success\_batch\_size indicates the size of the buffer on filling of which the result will be written to a temp location.This is used to boost the performance.
* **error\_batch\_size**:
  + **Is Required**: False (default value is 500)
  + **Description**:Overwatch makes multiple api calls and stores the interim results to a temp location error\_batch\_size indicates the size of the buffer which contains the error incase of the api call failure on filling of which the result will be written to a temp location.This is used to boost the performance.
* **enable\_unsafe\_SSL**:
  + **Is Required**: False
  + **Rules**:Max Days should be either true or false.
  + **Description**: Enables unsafe ssl.
* **thread\_pool\_size**:
  + **Is Required**: False(default value is 4)
  + **Rules**:Should be a number.
  + **Description**: Overwatch makes multiple api calls in parallel thread\_pool\_size indicates the level of parallelism.
* **api\_waiting\_time**:
  + **Is Required**: False(default value 300000ms)
  + **Rules**:Should be a number.
  + **Description**: Overwatch makes multiple async api calls in parallel.api\_waiting\_time signifies the waiting time incase of no response received from the api call.

**Command Descriptions**

**Providing the Config Csv path**

val configCsvPath = "dbfs:/FileStore/sriram/databricks\_azure.csv"

**Initializing the MultiWorkspace Object**

val multiWorkspaceDeploymentTest = com.databricks.labs.overwatch.MultiWorkspaceDeployment(configCsvPath,"<Temp Location>")

**//**Temp location is used for validation purposes only and it gets cleaned after each run.

**Running the Validation**

multiWorkspaceDeploymentTest.validate(4)

//4 indicates the number of threads that will be running in parallel to complete the task.

**Performing the deployment**

multiWorkspaceDeploymentTest.deploy(2,"Bronze,Silver,Gold")

//2 indicates the number of threads that will be running in parallel to complete the task.

//"Bronze,Silver,Gold" indicates the zone which you want to run.It can be either Bronze,Silver,Gold or all of it.