Table of Contents

[**Main Parts to run Databricks notebook from Control-M:** 2](#_Toc143850665)

[**Databricks job:** 2](#_Toc143850666)

[**AZ Login:** 3](#_Toc143850667)

[**Create YAML File:** 3](#_Toc143850668)

[**Copy on-prem source files to Azure Cloud:** 3](#_Toc143850669)

[**Azpipr app:** 3](#_Toc143850670)

[**Control-M job creation:** 4](#_Toc143850671)

[**Sample Code for Examples:** 4](#_Toc143850672)

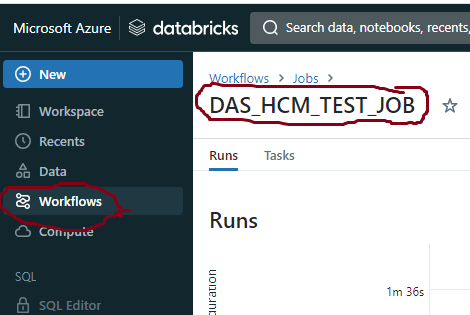
**Control-M Orchestration for Databricks**

# **Main Parts to run Databricks notebook from Control-M:**

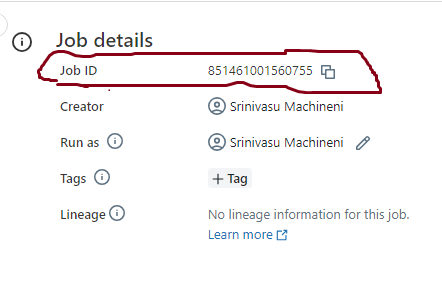
## **Databricks job:**

Let’s assume we have a databricks notebook with script(s) written in their cells and a job is created under <workflows>.

The below job: DAS\_HCM\_TEST\_JOB is created for POC to show case running databricks notebook from Control-M Job.



When the job is created, it assigns an ID to the job as you see below.



Link to the databricks notebook @ DEV : [https://adb-7704615097133693.13.azuredatabricks.net/?o=7704615097133693#notebook/1336572539334954/command/1336572539334955](https://adb-7704615097133693.13.azuredatabricks.net/?o=7704615097133693%23notebook/1336572539334954/command/1336572539334955)

(When the below job is executed , it will run the above databricks notebook)

Link to the job: <https://adb-7704615097133693.13.azuredatabricks.net/?o=7704615097133693#job/851461001560755>

## **AZ Login:**

You need to login Azure environment in the shell wrapper script by providing username and password.

The <az login> command is part of the az CLI (Command-Line Interface) for Microsoft Azure, which is the official Command-Line Interface provided by Microsoft for managing Azure resources and services. The <az login> command is used to authenticate and log in to the Azure cloud platform.

Ex:

az login -u $service\_account -p $service\_account\_password

## **Create YAML File:**

Create yaml file with the information such as job\_ID and any other parameters that are needed to pass on to the databricks notebook.

create\_yaml() {

input\_params="parameters: {\"job\_id\":\"$databricks\_job\_id\"}"

echo "jobs:" > $runtime\_yml\_file

echo " job1:" >> $runtime\_yml\_file

echo " pipeline: ${job\_name} " >> $runtime\_yml\_file

echo " $input\_params" >> $runtime\_yml\_file

}

In the above create\_yaml function:

$databricks\_job\_id => 851461001560755

$job\_name => DAS\_HCM\_TEST\_JOB

$runtime\_yml\_file => /path\_on\_azure\_edge\_server/<databricks\_job\_name>.yaml

## **Copy on-prem source files to Azure Cloud:**

If you receive source files on Azure VM for your application, make sure you copy them from Azure VM to Azure Cloud for databricks notebook to access them as below.

## Function that copies files from Azure VM to azure Cloud.

copy\_files\_2\_azure() {

# it will loop through the files in the directory to copy each file to azure cloud one at a time.

for FILE in ${sourced}/\*

do

az storage blob upload --account-name app0000672156pdevnc307 --container-name hcmconsumption-dev --auth-mode login --file ${FILE} --name /hcm/carelon\_poc/scenario-1/csv/$(basename "${FILE}")

done

}

In the above function: copy\_files\_2\_azure,

$source\_dir => directory name on Azure VM in which source files are available.

${FILE} => Represents path/file\_name that are in the $source\_dir location.

--account-name => Azure Storage account name

--container-name => Container name of the Azure Storage account

--name => Folder/sub-folder/file\_name in the Azure Storage account’s Container.

## **Azpipr app:**

Azpipr is a script developed by HCSC team to connect to databricks from Azure VM and is made available on Azure VMs. You can invoke this app by simply calling as azpipr

# This below function will call azpipr to run the databricks notebook in the specific environment (DEV/TEST/PROD).

Call\_azpipr() {

azpipr\_cmd="azpipr run -w "$databricks\_host" -f "$runtime\_yml\_file" --scope "AZURE\_DATABRICKS" --auth-method CLI"

$azpipr\_cmd

}

Here,

$databricks\_host => "adb-7704615097133693.13" ( it is DEV databricks env)

$runtime\_yml\_file => /path\_on\_azure\_VM/yaml\_file\_name ( this was created above under create yaml file section)

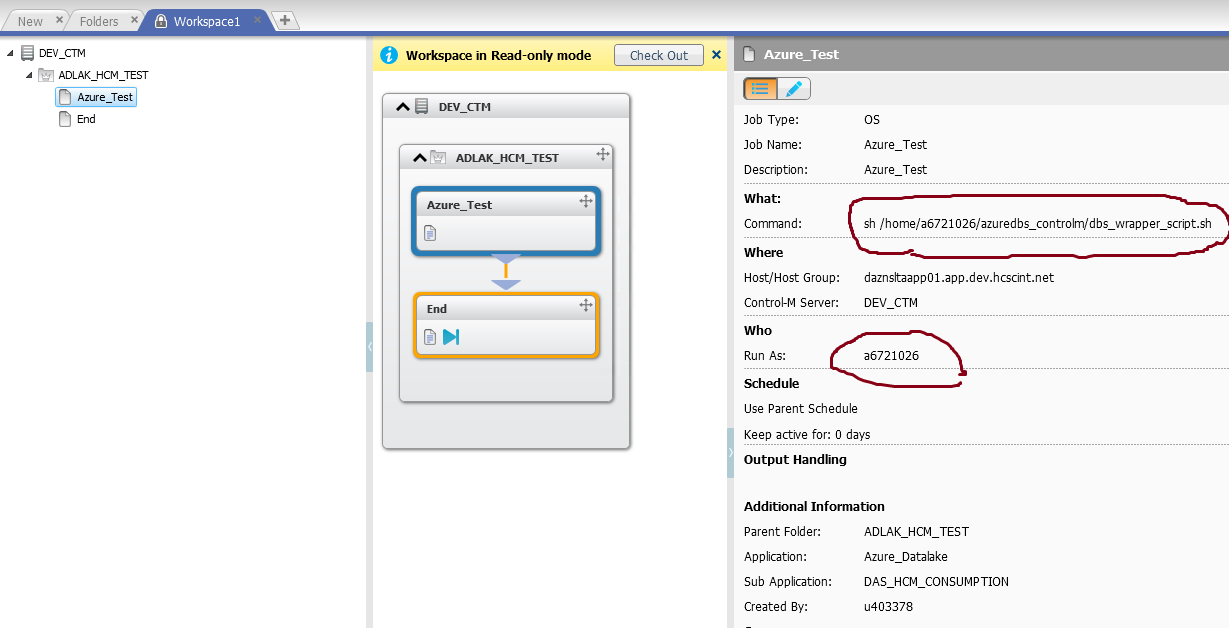
At a minimum, we need the above code in the shell wrapper script on Azure VM to be able to execute the databricks notebook from Control-M

## **Control-M job creation:**

When you create a job in control-m, you will mention the shell wrapper script to the command.

You will select NPID: a6721026 to Run As @ Dev environment.

You will select NPID: a6721021 to Run As @ Test environment.

****

# **Sample Code for Examples:**

