**Follow the Steps for AWS Parquet files Ingestion Process:**

1. Login to Snowflake, open a new worksheet, and select SYSADMIN role. Run the following SQL to create a project-specific data warehouse for the new project (Note: in the SQL below, replace XXXX with the project name (i.e. abbreviation), with spaces replaced by underscores):

CREATE WAREHOUSE DP4BI\_LOAD\_WH\_DEV\_XXXX WITH

WAREHOUSE\_SIZE = 'XSMALL'

WAREHOUSE\_TYPE = 'STANDARD'

AUTO\_SUSPEND = 60

AUTO\_RESUME = TRUE

MIN\_CLUSTER\_COUNT = 1

MAX\_CLUSTER\_COUNT = 1

SCALING\_POLICY = 'STANDARD'

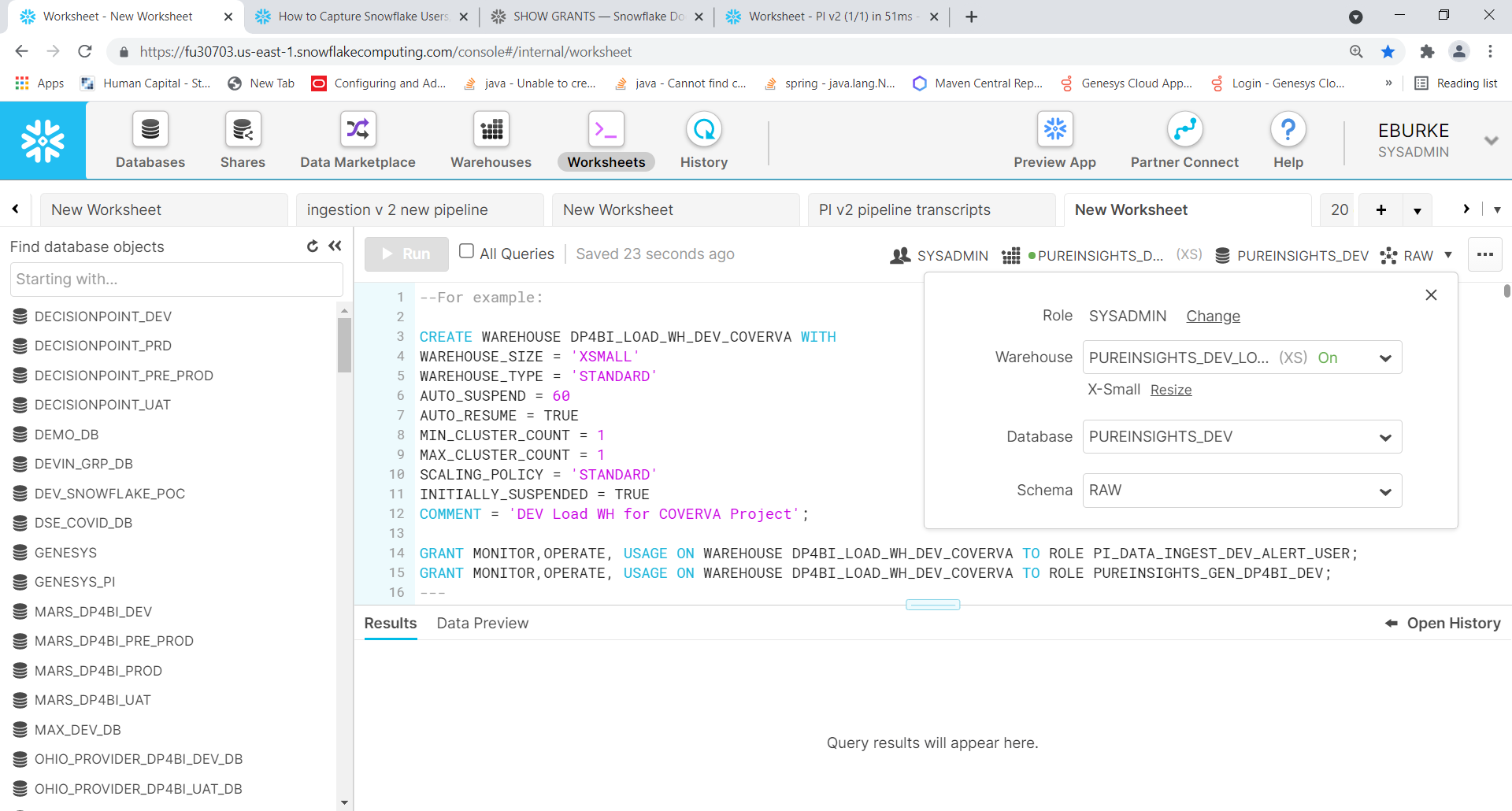
INITIALLY\_SUSPENDED = TRUE

COMMENT = 'DEV Load WH for XXXX Project';

GRANT MONITOR,OPERATE, USAGE ON WAREHOUSE DP4BI\_LOAD\_WH\_DEV\_XXXX TO ROLE PI\_DATA\_INGEST\_DEV\_ALERT\_USER;

GRANT MONITOR,OPERATE, USAGE ON WAREHOUSE DP4BI\_LOAD\_WH\_DEV\_XXXX TO ROLE PUREINSIGHTS\_GEN\_DP4BI\_DEV;

For example:



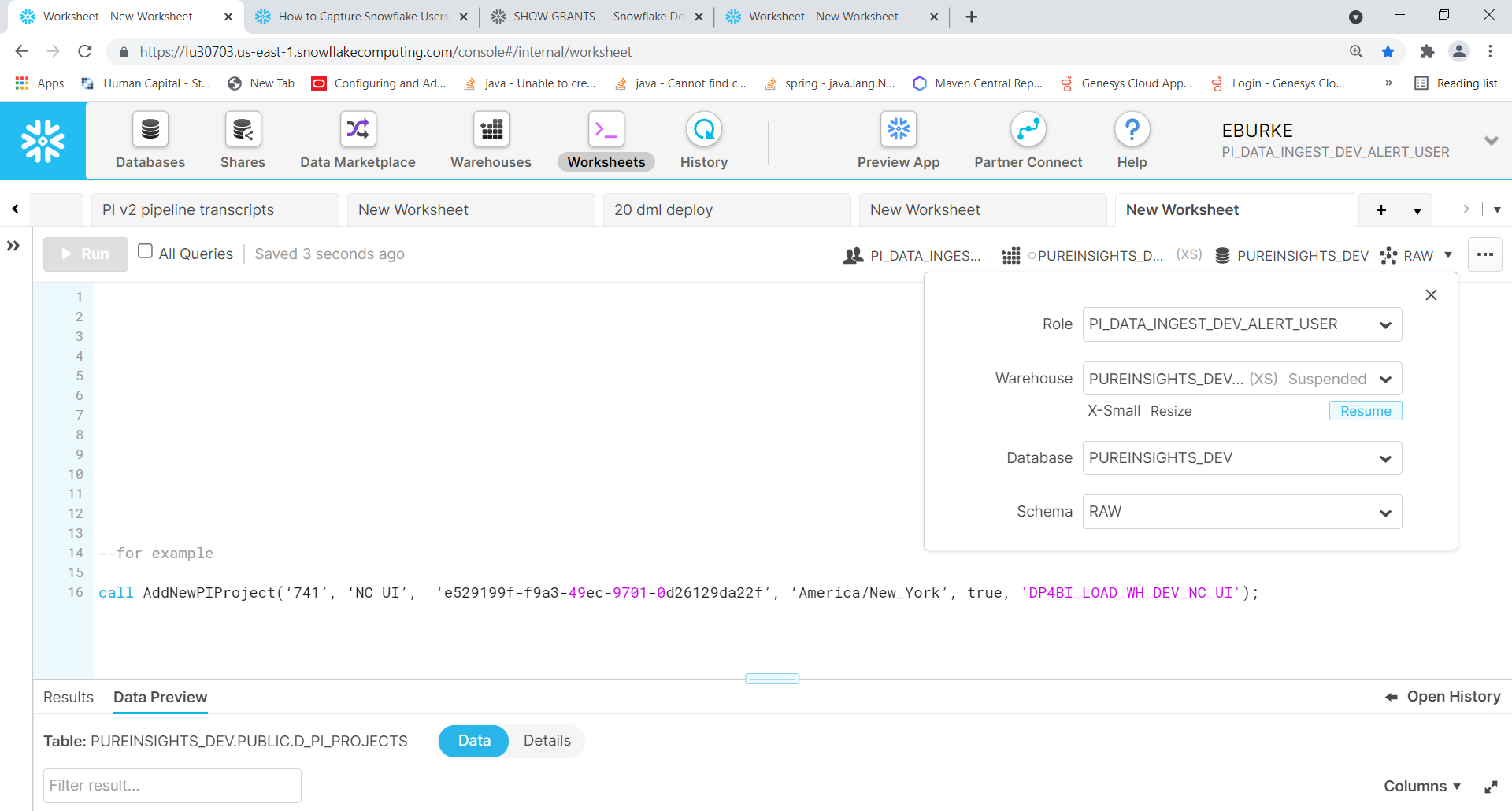
1. Log in to Snowflake, open a new worksheet, and select the following: Role: PI\_Data\_Ingest\_DEV\_Alert\_User, Warehouse: PUREINSIGHTS\_DEV\_LOAD\_DAILY\_WH,

Database: PUREINSIGHTS\_DEV, SCHEMA: RAW Run the following stored procedure to add the new project to the configuration tables, to schedule the project for daily ingestion, and to run the initial ingestion.

AddNewPIProject(projectid VARCHAR, projectname VARCHAR, AWSFOLDERNAME VARCHAR, PROJECTTIMEZONE VARCHAR,ACTIVE BOOLEAN ,INGEST\_WH VARCHAR)

For example:

call AddNewPIProject(‘741’, ‘NC UI’,  ‘e529199f-f9a3-49ec-9701-0d26129da22f’, ‘America/New\_York’, true, 'DP4BI\_LOAD\_WH\_DEV\_NC\_UI');



This will add the configuration information for the new project into the public.d\_pi\_projects table and into the public.d\_pi\_project\_unavailable\_tables table (based on current availability of data in S3 for this project). It will also schedule the project for daily ingestion and immediately perform the initial ingestion of the data. Note: logs for the configuration table insertion are found in the raw.pi\_data\_check\_log table; logs for the data ingestion are found in the raw.ingest\_pi\_data\_det\_log table. You can also check the status of the initial data ingestion by selecting ROLE: SYSADMIN and running the following SQL:

select \* from public.admin\_pi\_ingestion\_status\_by\_project\_vw where status\_date = current\_date() order by projectid;

1. Check for missing data and mis-formatted data headers in the parquet files/raw tables by running each data view in PUBLIC for the new project. (For this, select the following: Role: SYSADMIN, Warehouse: PUREINSIGHTS\_DEV\_LOAD\_DAILY\_WH,

Database: PUREINSIGHTS\_DEV, SCHEMA: PUBLIC). Look for columns that have only nulls in them.  If they correspond to the unavailable tables inserted into d\_pi\_project\_unavailable tables in step 2 above, this is not an issue.  However, if they correspond to available tables, further research will be required.  You can compare to the data in the corresponding MYSQL database table to see if the data is missing there as well.  If necessary, you can look at the RAW column in the raw table to see if the data headers in the json string that represents the data has changed from other projects.  If there is an issue, open a support ticket with Pure Insights by emailing [support@mypureinsights.com](mailto:support@mypureinsights.com). Copy all the relevant folks on our team and also Arnold.

1. Check for duplicate records in the raw tables by running the following stored procedure (For this select the following: Role: PI\_Data\_Ingest\_DEV\_Alert\_User, Warehouse: PUREINSIGHTS\_DEV\_LOAD\_DAILY\_WH, Database: PUREINSIGHTS\_DEV, SCHEMA: RAW):

              call checkAllTablesForDups();

Results will be written to the log table raw.pi\_data\_check\_log. Note: this stored procedure checks all projects for duplicates.  If it is flagging duplicates for projects other than the one you are working on, let the team know (Eric and Rama, for now).  If it flags duplicates for the project you are working on, check the MYSQL database tables to see if the duplicates also exist there. If the duplicates exist in the raw tables but not in MYSQL, check first with the team (for now, Eric or Rama) and they will indicate how to proceed (likely, it will be to open a support ticket with PI).  If the duplicates exist in both environments, open a support ticket with Pure Insights ([support@mypureinsights.com](mailto:support@mypureinsights.com)) , copying the relevant folks on our team and Arnold.

1. For deployment to UAT and PRD, you only need to repeat steps 1 and  2 in these environments, logging in as the equivalent UAT or PRD user and selecting the equivalent roles, warehouses, and databases (and raw for the schema).
2. If for any reason you need to manually reingest the data for a project, use the following stored procedure (For this, select the following: Role: PI\_Data\_Ingest\_DEV\_Alert\_User, Warehouse: PUREINSIGHTS\_DEV\_LOAD\_DAILY\_WH, Database: PUREINSIGHTS\_DEV, SCHEMA: RAW):

Call ingestUningestedPIData('XXX',true,false,true,false)–replae 9999 with the projected for your new project.