Contents

[Daily PI Parquet File Monitoring. 1](#_Toc81471000)

[PI Data Ingestion Successful 1](#_Toc81471001)

[PI Data Ingestion Not Completed on Time for Some Projects (Case One) 2](#_Toc81471002)

[PI Data Ingestion Not Completed on Time for Some Projects (Case Two) 3](#_Toc81471003)

[PI Data Ingestion Errors 3](#_Toc81471004)

# Daily PI Parquet File Monitoring.

Check your email at 6:30 am EST.  You should receive an ingestion status email for UAT and for PRD.  These emails will each contain one of three possible subjects/messages.

1. Subject: PUREINSIGHTS\_PRD: PI Data Ingestion Successful -- this email indicates that all projects configured as active in the relevant environment (UAT/PRD) have ingested successfully and on time (prior to 6:30 am EST). The email will contain a list of projects and ingestion completion times, as well as a list of warnings (you can ignore the warnings unless any of these are unexpected).
2. Subject: PUREINSIGHTS\_PRD: PI Data Ingestion Not Completed on Time for Some Projects – this email indicates that one or more active projects did not complete ingestion by 6:30 am EST. The email will list the projects that have not completed by 6:30 am.
3. Subject: PUREINSIGHTS\_PRD: PI Data Ingestion Errors – this email indicates that there were errors or failures in the ingestion for at least some active projects. This email will contain a list of ingestion processes that failed to complete (if there were any), a list of handled errors (if there were any), a list of projects without a success indicator in the log (if there are any), a list of tables without a success indicator in the log (if there are any), a list of stored procedures without a success indicator in the log (if there are any), as well as a list of projects that did complete ingestion successfully.

## PI Data Ingestion Successful

If the only email you receive for each environment is the email titled “PI Data Ingestion Successful” then do the following:

1. If the information in the email looks good, you are done.
2. If the information in the email seems unusual or incomplete in any way, you can confirm by running the “by project” ingestion status view in the relevant environment (UAT/PRD). Login to Snowflake and select the PUREINSIGHTS\_UAT or PURINSIGHTS\_PRD database and the PUBLIC schema. Run the following:

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

1. Verify that for all active projects in that environment, there is an entry with a status of “COMPLETED” or “COMPLETED WITH WARNINGS”, that the status\_time is less than or equal to 6:30 am EST, that the tables\_ingested count is equal to the tables\_available count and that the stored\_procedures\_completed count is equal to the stored\_procedures\_available count.
2. If you discover any problems, go ahead and troubleshoot (using the resources listed at the end of this email)

## PI Data Ingestion Not Completed on Time for Some Projects (Case One)

If you receive the email titled “PI Data Ingestion Not Completed on Time for Some Projects” and you have not yet received the follow-up ingestion email (indicating success or errors) that is sent once the delayed projects do finally ingest, then do the following:

1. Research the reason for the delay for the projects listed. The most common reason is that the PI data is late from Pure Insights, that it has not been placed in our S3 bucket in time (the data for all projects is supposed to be available in our S3 bucket by approximately 6:00 am EST to allow time for the ingestion processes to complete by 6:30 am).  The most direct way to check on this is to run the sp checkunavailabletables and compare the results with what is in the project\_unavailable\_tables table for each project that is delayed. Login to Snowflake and select the “PI\_Data\_Ingest\_UAT\_Alert\_User” role or “PI\_Data\_Ingest\_PRD\_Alert\_User” role, select the PUREINSIGHTS\_UAT or PUREINSIGHTS\_PRD database and the RAW schema, and run the following:

call raw.checkunavailabletables('XXX')  – replace XXX with projectid, run from RAW schema, logged in as the alert ingest user for that env.

select \* from public.d\_pi\_project\_unavailable\_tables where projectid = ‘XXX’  -- replace XXX with projectid

 The tables that are returned by the sp checkunavailabletables but are not in the unavailable tables table for that project are the tables that are still missing data at the time that you check.

1. If it is the case that the delay is due to data being not yet available for certain tables for these projects (as indicated in step 1), then open a support ticket with PI describing which projects and which tables are still missing. Open the support ticket by sending an email describing which tables for which projects are still missing as of such and such a time. Send the email to [support@mypureinsights.com](mailto:support@mypureinsights.com) and copy [soapaxbiproduct@maximus.com](mailto:soapaxbiproduct@maximus.com), [DP4BI@maximus.com](mailto:DP4BI@maximus.com) and any other relevant stakeholders. Also send a separate email to the above list (minus PI support) and let them know that there is a delay and describe what is happening and indicate that you will keep them updated).
2. Continue to monitor your emails until Kevin from PI informs you that the data is now available and you receive the follow-up ingestion status email.  If you receive the “PI Data Ingestion Successful” message, then write a follow-up email to the above recipients letting them know that the data has now been ingested and any affected cubes or reports can be refreshed if needed.
3. The ingestion scripts will keep looking for the data and will automatically ingest it when it becomes available as long as the data becomes available within 4 hours of the original deadline.  Kevin is usually able to make the missing data available within this window, in which case item 3 will apply.
4. However, if the data is not made available within this time frame, then the ingestion script is configured to time out and generate an error. In this case, you will need to communicate back and forth with Kevin until he communicates that the data is available and then you will need to rerun the ingestion script manually for the projects that were delayed. . Login to Snowflake and select the “PI\_Data\_Ingest\_UAT\_Alert\_User” role or “PI\_Data\_Ingest\_PRD\_Alert\_User” role, select the PUREINSIGHTS\_UAT or PUREINSIGHTS\_PRD database and the RAW schema, and run the following:

call raw. ingestUningestedPIData(‘XXX’, true, false, true, false); -- replace ‘XXX’  with projectid,  run logged in as alert ingest user for the appropriate env, run from RAW schema.

1. Check to make sure the ingestions were successful by running the “by project” ingestion status view in the appropriate environment. Login to Snowflake and select the PUREINSIGHTS\_UAT or PUREINSIGHTS\_PRD database and the PUBLIC schema. Run the following:

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

Then send out your follow-up email to the relevant stakeholders letting them know ingestion has completed and any affected cubes or reports can be refreshed.

1. If, after the delayed projects are ingested (either automatically as in step 3 above or manually as in step 5 above), you receive the “PI Data Ingestion Errors” email message or you discover errors by running the ingestion status view, then use the logs and views listed below to research.
2. Reach out to discuss remediation steps.

## PI Data Ingestion Not Completed on Time for Some Projects (Case Two)

If you receive the email titled “PI Data Ingestion Not Completed on Time for Some Projects” and you have also already received the follow-up ingestion email (indicating success or errors) that is sent once the delayed projects do finally ingest, then do the following:

1. If the follow-up message is a success message, then open a support ticket with PI (as in step 2 above) to make sure the problem is resolved before the next day’s ingestion.  In the support ticket, let Kevin know which projects were late.  If he needs more information, check the appropriate sub-folders in the S3 buckets to see the datetimestamp for parquet files within the table folders for the relevant projects. This will tell you when the data became available for the relevant tables. If the datetimestamp is significantly later than 6 am, you can let Kevin know that this project/table was an issue.
2. If the follow-up message is an error message, then use the logs and views listed below to research. Reach out to discuss remediation steps.

## 

## PI Data Ingestion Errors

If you receive the email titled “PI Data Ingestion Errors”, then use the logs and views listed below to research. Reach out to discuss remediation steps.  Keep Aurelio in the loop.

Logs and Views:

Main ingestion log:

select \* from raw.ingest\_pi\_data\_det\_log where to\_date(to\_varchar(ts)) = current\_date() order by ts desc;

Data Check Log (if you run any of the “Check…” admin sps:

select \* from raw.pi\_data\_check\_log order by ts desc;

By project ingestion status view:

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

By table ingestion status view:

      select \* from public.admin\_pi\_ingestion\_status\_by\_table\_vw where projectid = 'XXX' order by table\_name asc;