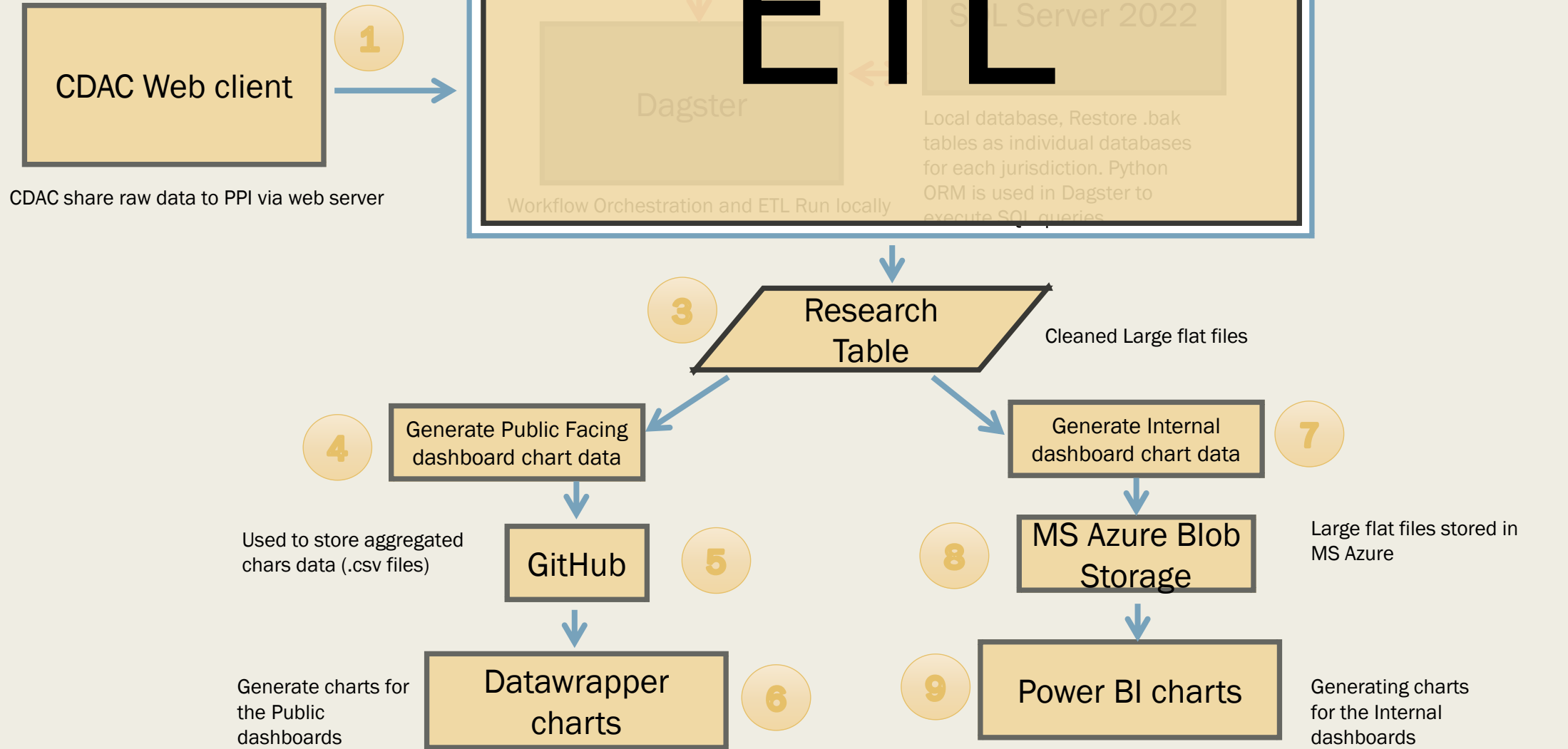


PUBLIC AND INTERNAL DASHBOARDS DATA PREPARATION PROCESS



06/11/2024

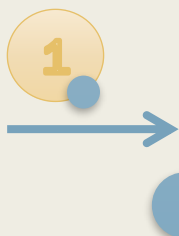
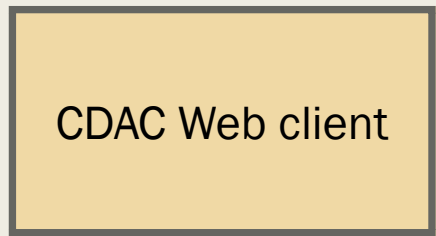
Dashboard Data Preparation Workflow



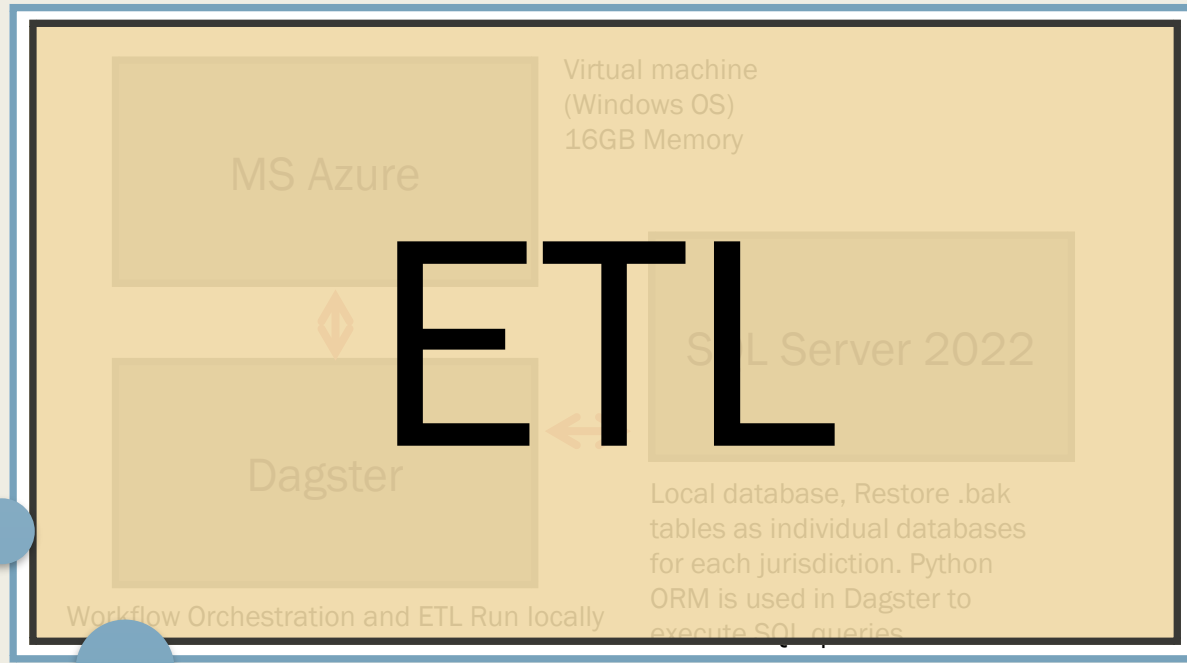
Dashboard Data Preparation Workflow

2

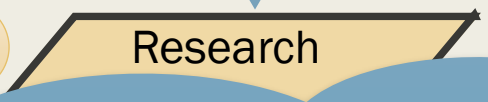
PPI clean the data and generate research table



CDAC share raw data to PPI via web server



3



4

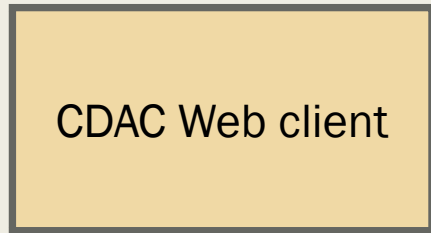
1. Data Collection by CDAC: CDAC will gather the necessary data from its databases and make it accessible via the CDAC web client for PPI.

Generate charts for the Public dashboards

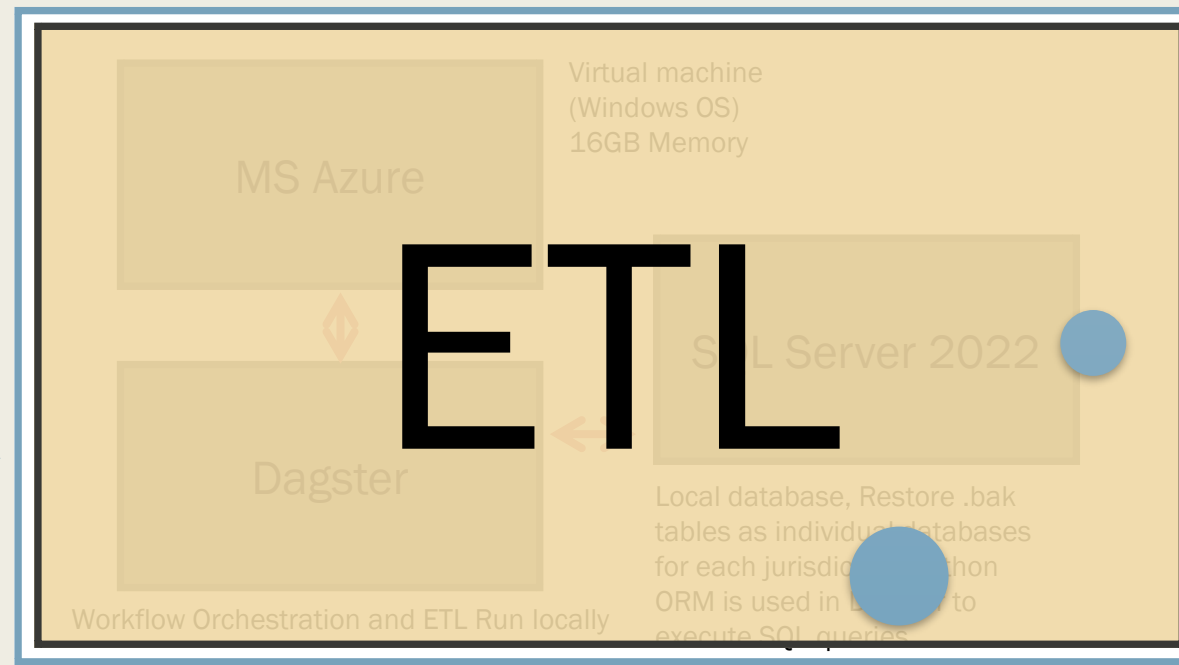


Generating charts for the Internal dashboards

Dashboard Data Preparation Workflow



CDAC share raw data to PPI via web server



PPI clean the data and generate research table



2. Data Extraction and Transformation by PPI: PPI will retrieve the data from the CDAC web client and perform ETL (Extract, Transform, Load) operations. This involves extracting raw data, transforming it into a usable format, and loading it into a cleaned data set.

Generate charts for the Public dashboards

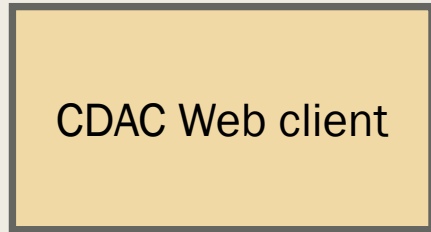


Generating charts for the Internal dashboards

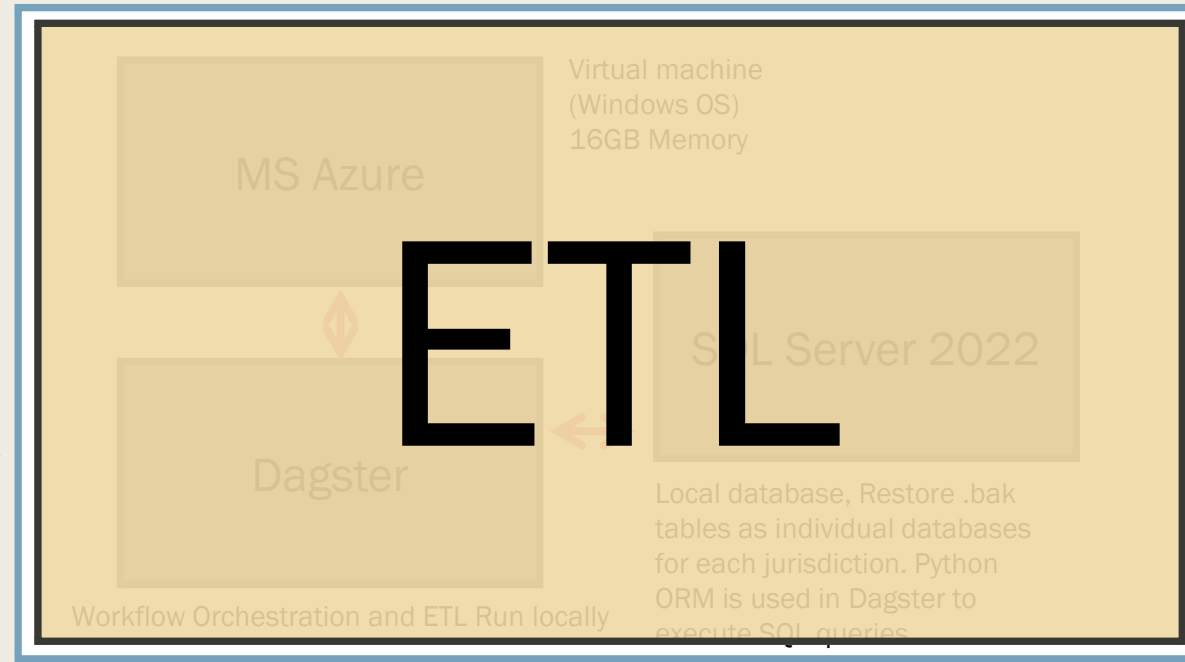
Dashboard Data Preparation Workflow

2

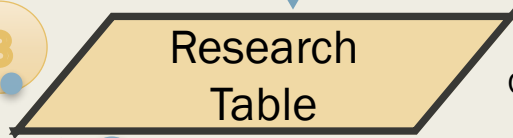
PPI clean the data and generate research table



CDAC share raw data to PPI via web server

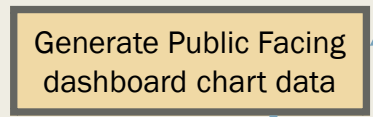


3

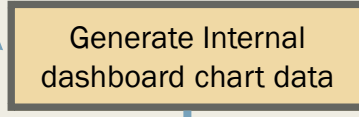


Cleaned Large flat files

4



7



Used to store aggregated chars data (.csv files)



Generate charts for the Public dashboards

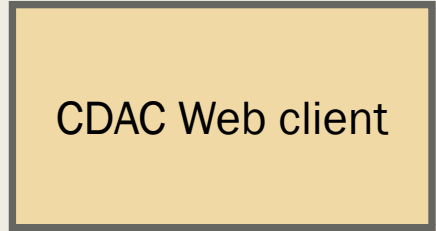


3. Generation of Research Table: The cleaned data will be compiled into a flat file named "research table" (in CSV format) for each district.

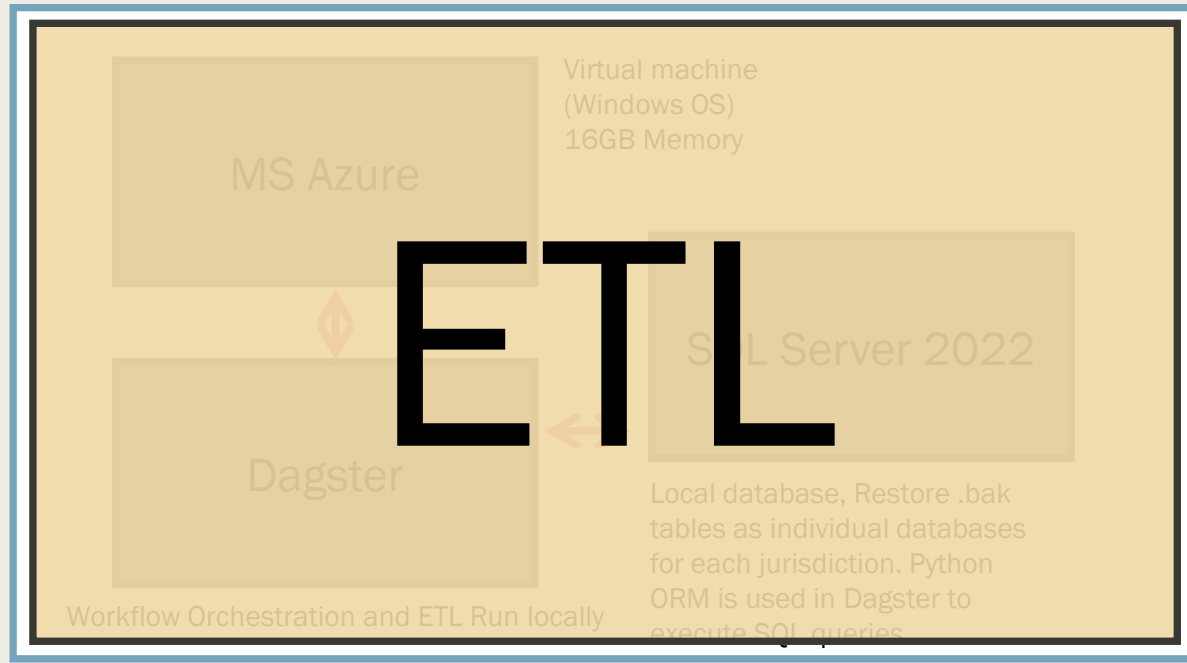
Dashboard Data Preparation Workflow

2

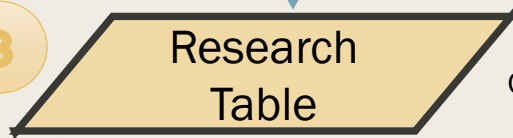
PPI clean the data and generate research table



CDAC share raw data to PPI via web server

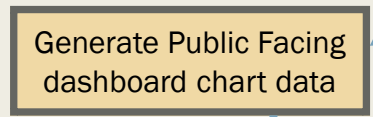


3

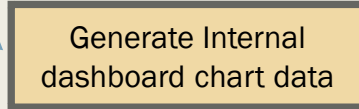


Cleaned Large flat files

4



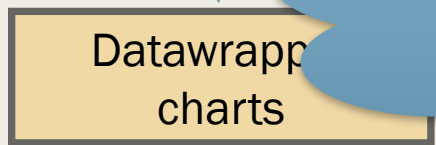
7



Used to store aggregated chars data (.csv files)

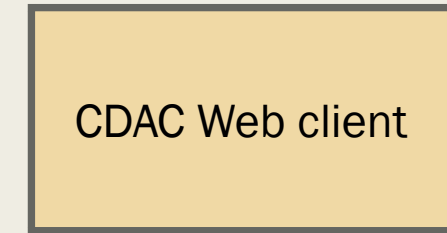


Generate charts for the Public dashboards



4. Aggregation for Public Dashboard: For the public dashboard charts, the research table will be aggregated, and data files will be generated for each chart.

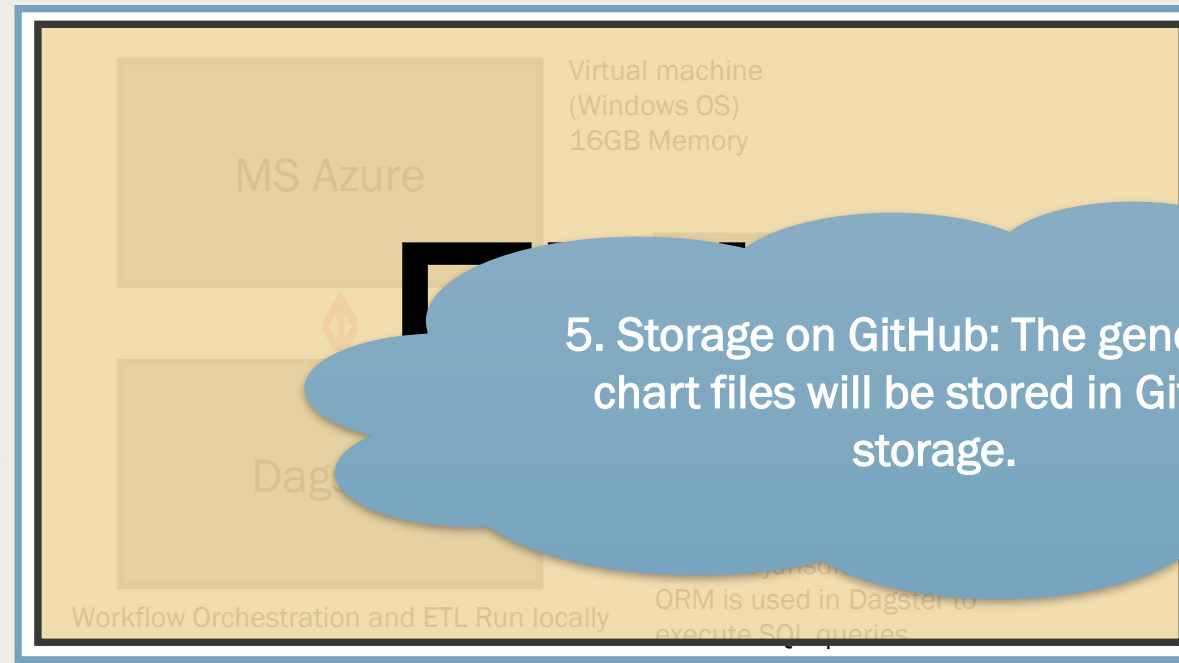
Dashboard Data Preparation Workflow



1

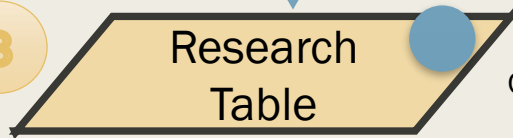


CDAC share raw data to PPI via web server



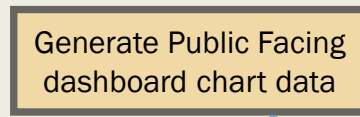
2

3

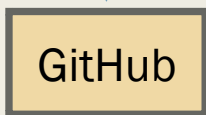


Cleaned Large flat files

4

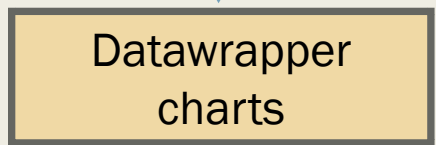


Used to store aggregated chars data (.csv files)



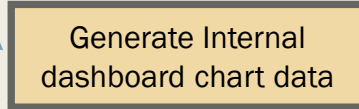
5

Generate charts for the Public dashboards



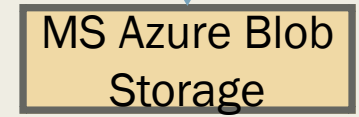
6

7



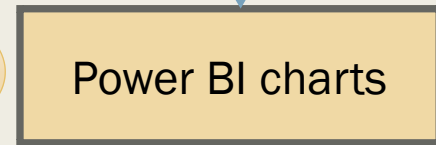
Large flat files stored in MS Azure

8

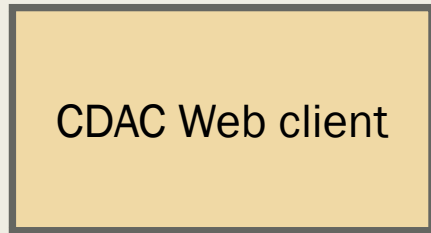


9

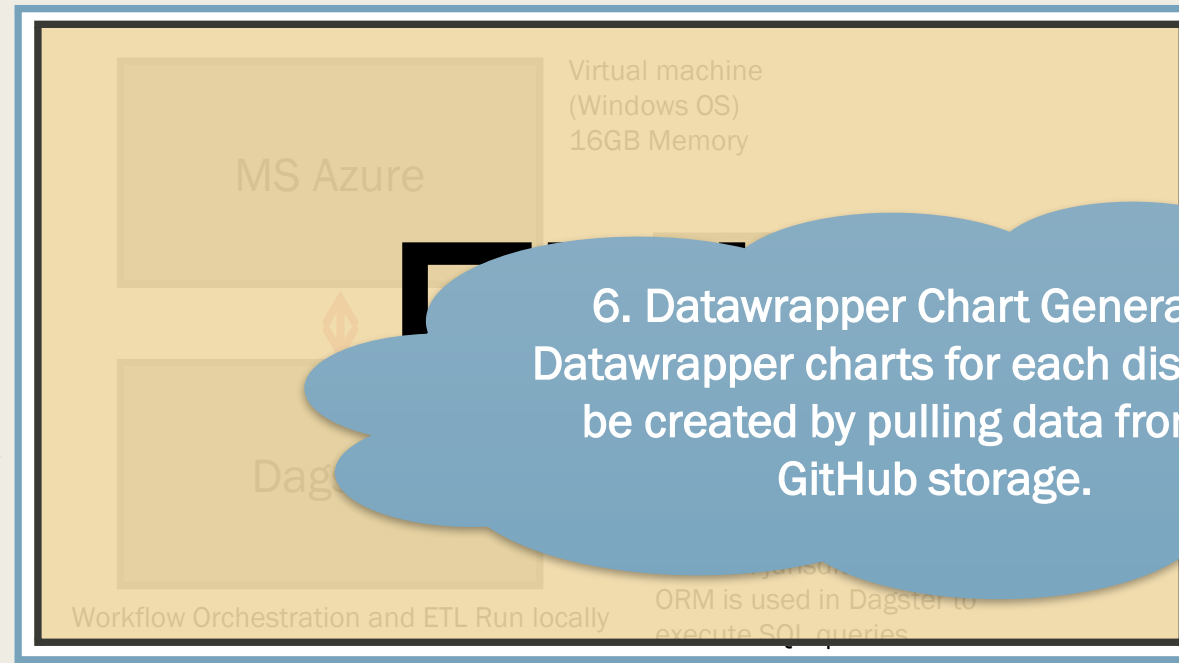
Generating charts for the Internal dashboards



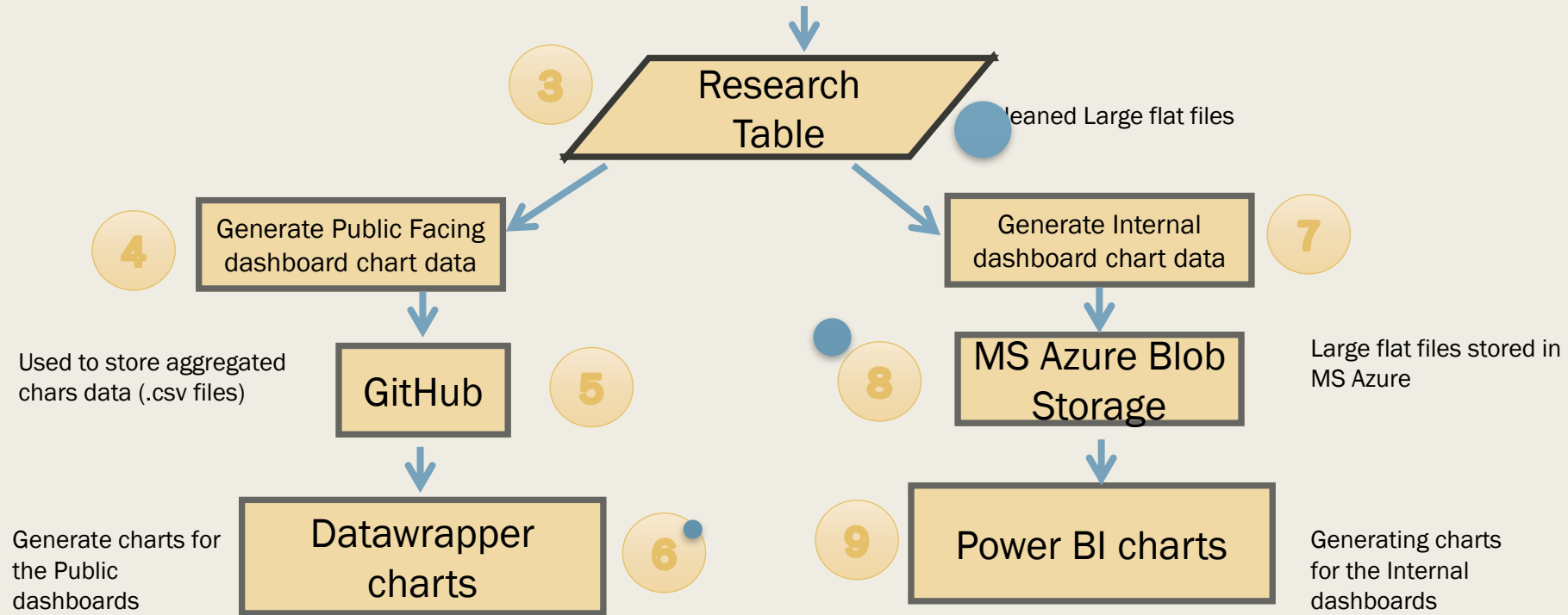
Dashboard Data Preparation Workflow



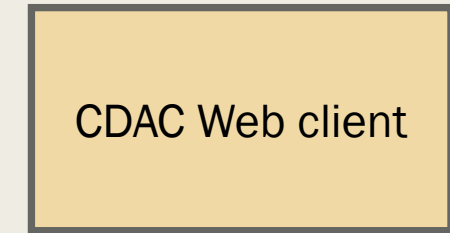
CDAC share raw data to PPI via web server



6. Datawrapper Chart Generation: Datawrapper charts for each district will be created by pulling data from the GitHub storage.



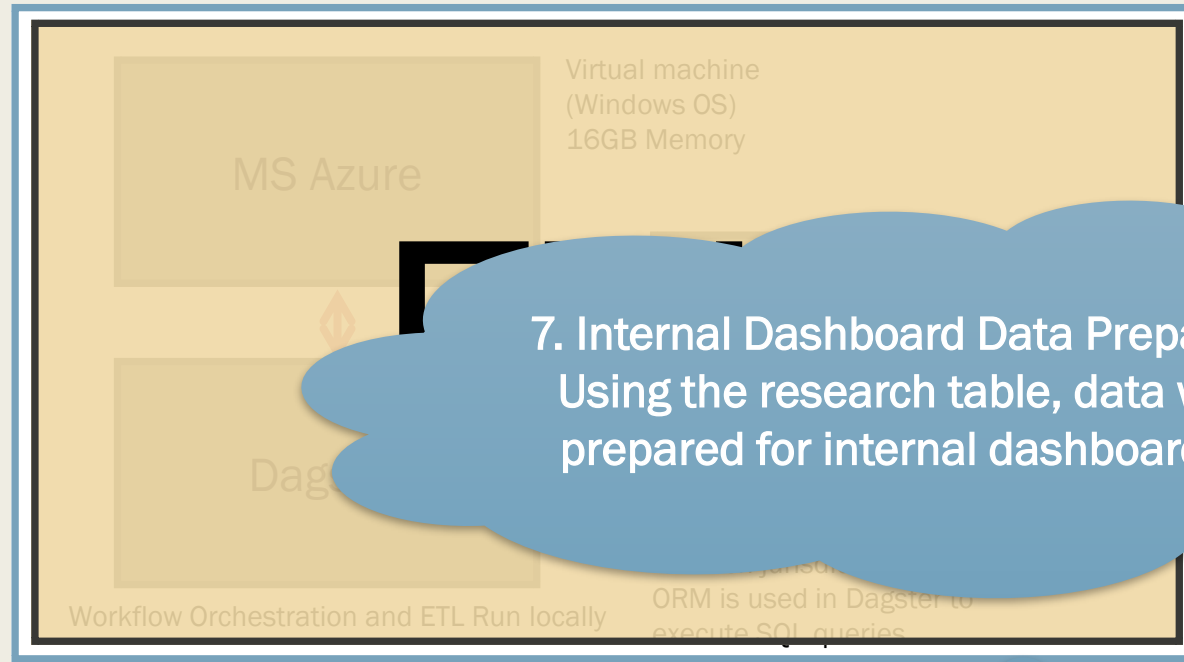
Dashboard Data Preparation Workflow



1



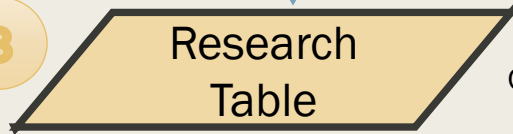
CDAC share raw data to PPI via web server



2

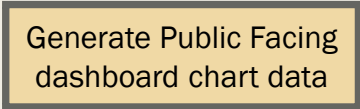
7. Internal Dashboard Data Preparation: Using the research table, data will be prepared for internal dashboard use.

3

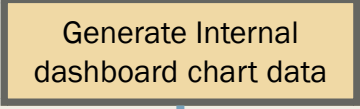


Cleaned Large flat files

4



7

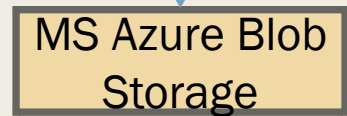


Used to store aggregated chars data (.csv files)



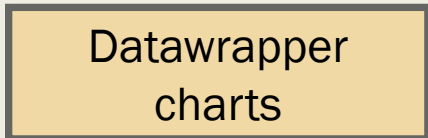
5

8



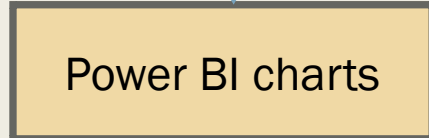
Large flat files stored in MS Azure

Generate charts for the Public dashboards



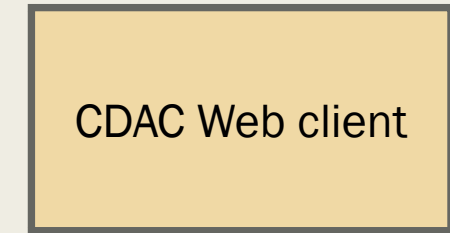
6

9



Generating charts for the Internal dashboards

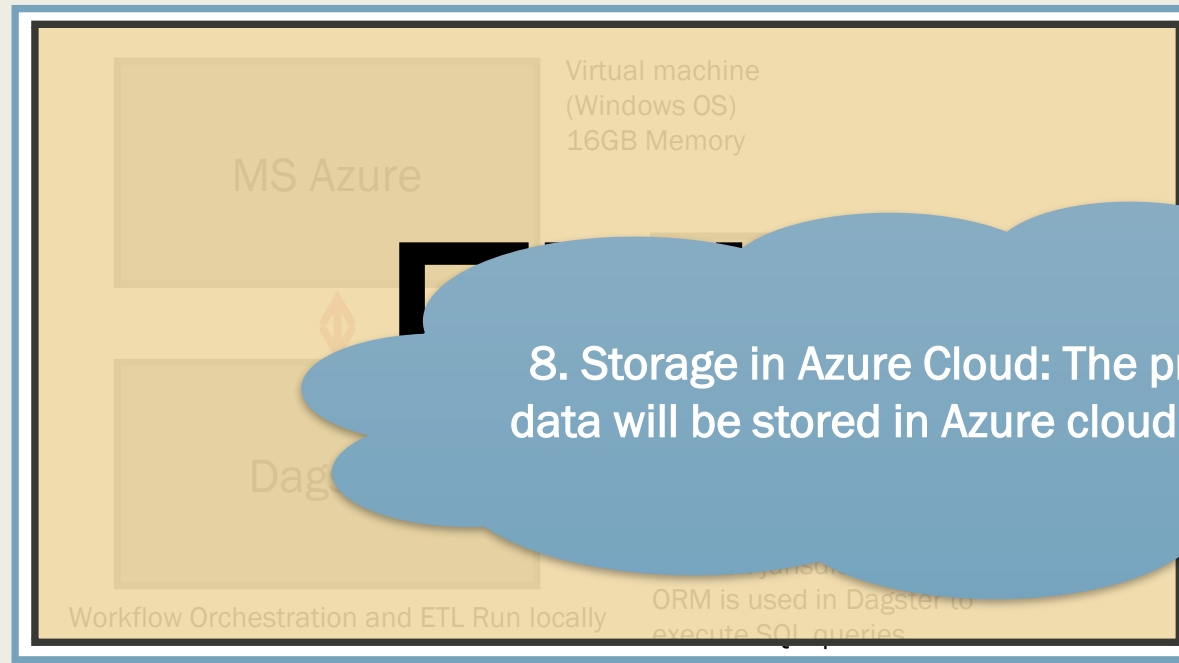
Dashboard Data Preparation Workflow



1

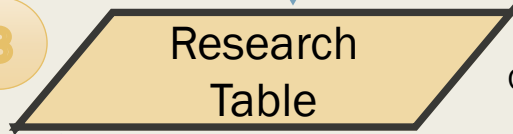


CDAC share raw data to PPI via web server



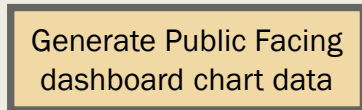
2

3

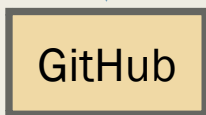


Cleaned Large flat files

4

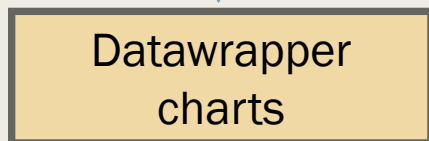


Used to store aggregated chars data (.csv files)

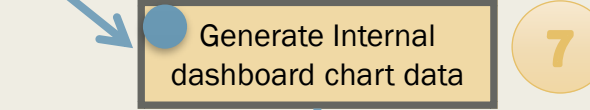


5

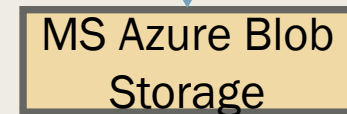
Generate charts for the Public dashboards



6

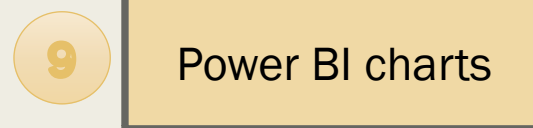


7



Large flat files stored in MS Azure

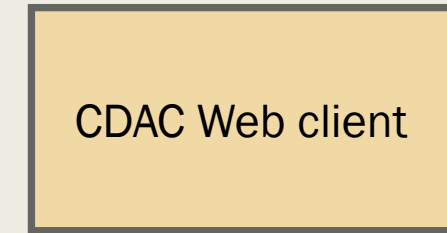
8



Generating charts for the Internal dashboards

9

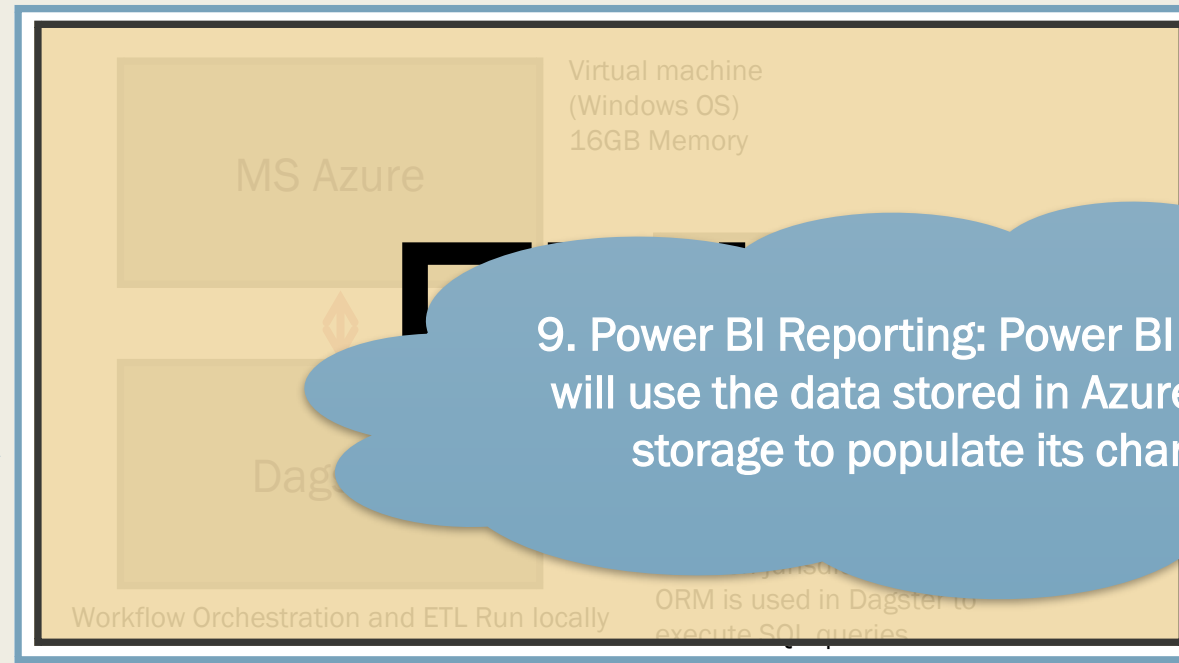
Dashboard Data Preparation Workflow



1



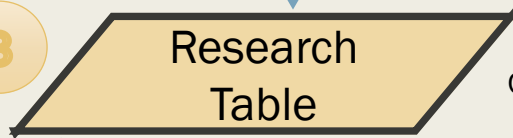
CDAC share raw data to PPI via web server



2

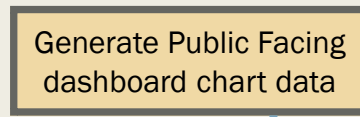
9. Power BI Reporting: Power BI reports will use the data stored in Azure cloud storage to populate its charts.

3

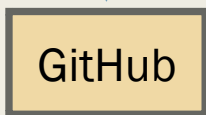


Clear large flat files

4

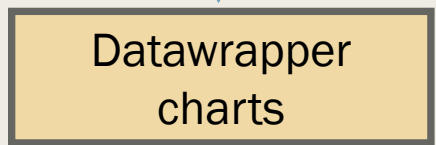


Used to store aggregated chars data (.csv files)



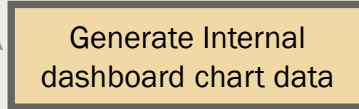
5

Generate charts for the Public dashboards



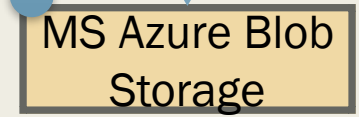
6

7

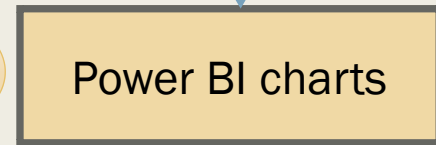


Large flat files stored in MS Azure

8



9



Generating charts for the Internal dashboards

Dashboard Data Preparation Workflow

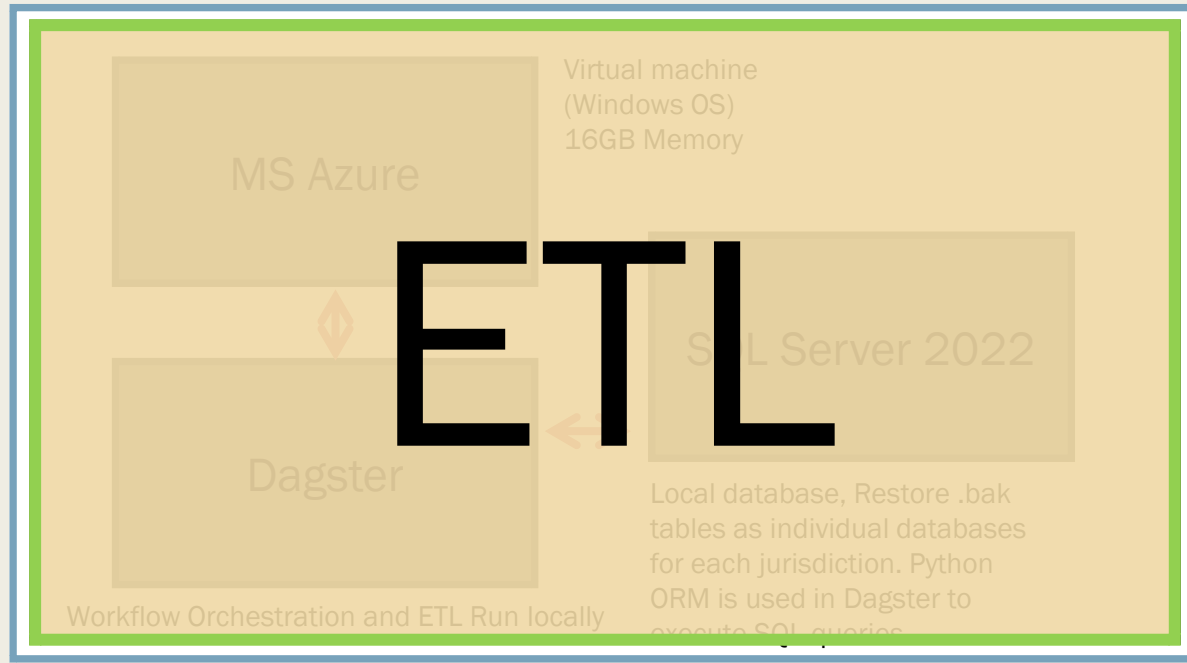
2

PPI clean the data and generate research table

CDAC Web client

1

CDAC share raw data to PPI via web server



3

Research Table

Cleaned Large flat files

4

Generate Public Facing dashboard chart data

7

Generate Internal dashboard chart data

Used to store aggregated chars data (.csv files)

GitHub

5

8

MS Azure Blob Storage

Large flat files stored in MS Azure

Generate charts for the Public dashboards

Datawrapper charts

6

9

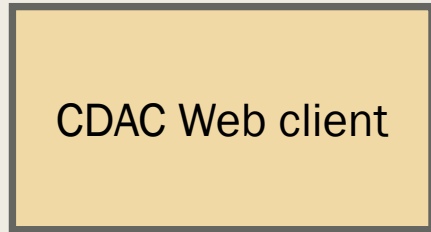
Power BI charts

Generating charts for the Internal dashboards

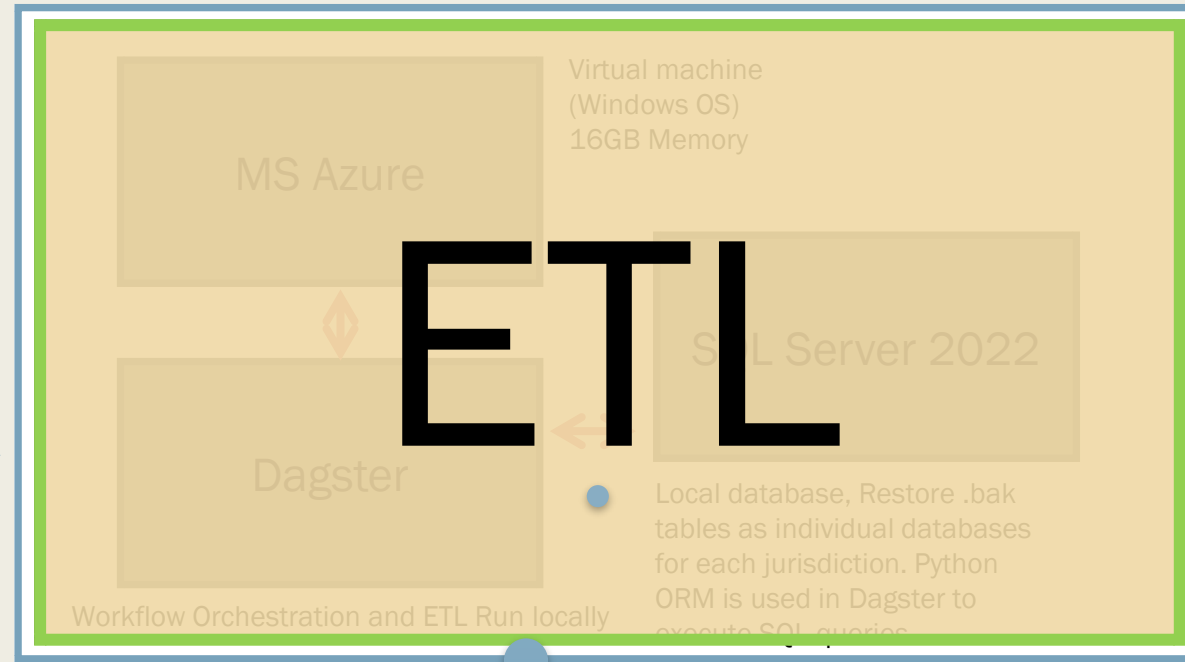
Dashboard Data Preparation Workflow

2

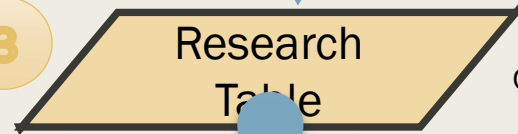
PPI clean the data and generate research table



CDAC share raw data to PPI via web server

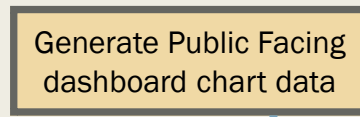


3



Cleaned Large flat files

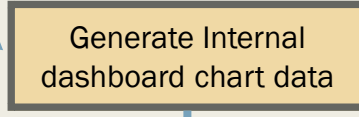
4



Used to store and generate public facing dashboard chart data (.csv)

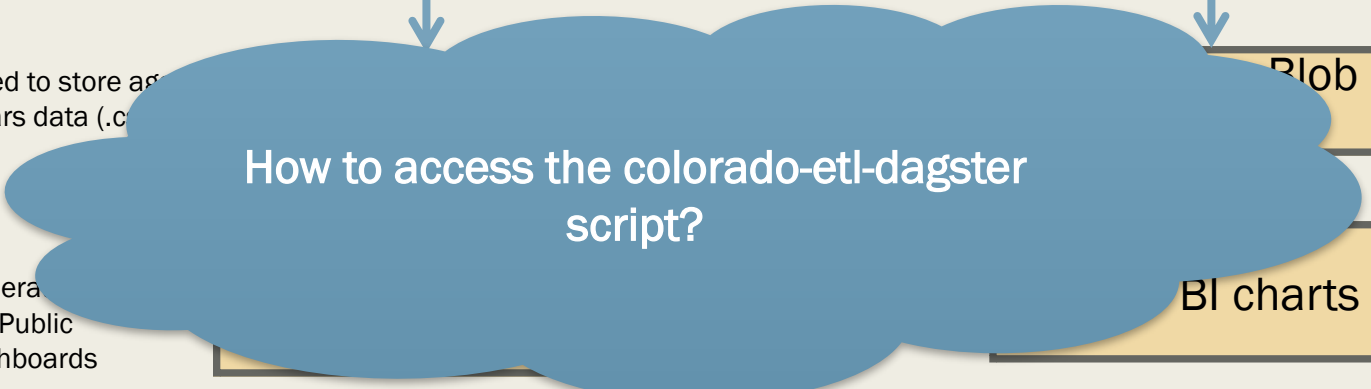
Generate the Public dashboards

7



Large flat files stored in MS Azure Blob

Generating charts for the Internal dashboards



Step-by-step guide to be a collaborator on colorado-etl-dagster repository

■ 1. Create a GitHub Account

a. Sign Up

- - Go to <https://github.com/>.
- - Click on "Sign up" in the upper-right corner.
- - Fill in the required details: username, email address, and password.
- - Complete the verification process and follow the on-screen instructions to finish creating your account.

Step-by-step guide to be a collaborator on colorado-etl-dagster repository

■ 2. Accept an Invitation to Be a Collaborator

a. Receive Invitation

The repository owner (PPI/Branden) will send you an invitation to collaborate on their repository. This will usually arrive via email or as a notification on GitHub.

b. Accept Invitation via Email

- Open the invitation email you received from GitHub.
- Click the "View Invitation" button in the email.
- This will take you to the GitHub website. Click "Join" to accept the invitation.

c. Accept Invitation via GitHub

If you didn't receive an email, you can also accept the invitation directly on GitHub:

- *Log in to your GitHub account.*
- *Click on your profile picture in the upper-right corner to open the menu, and select "Your repositories."*
- *Click on the "Invitations" tab to see pending invitations.*
- *Click "Accept" next to the relevant invitation.*

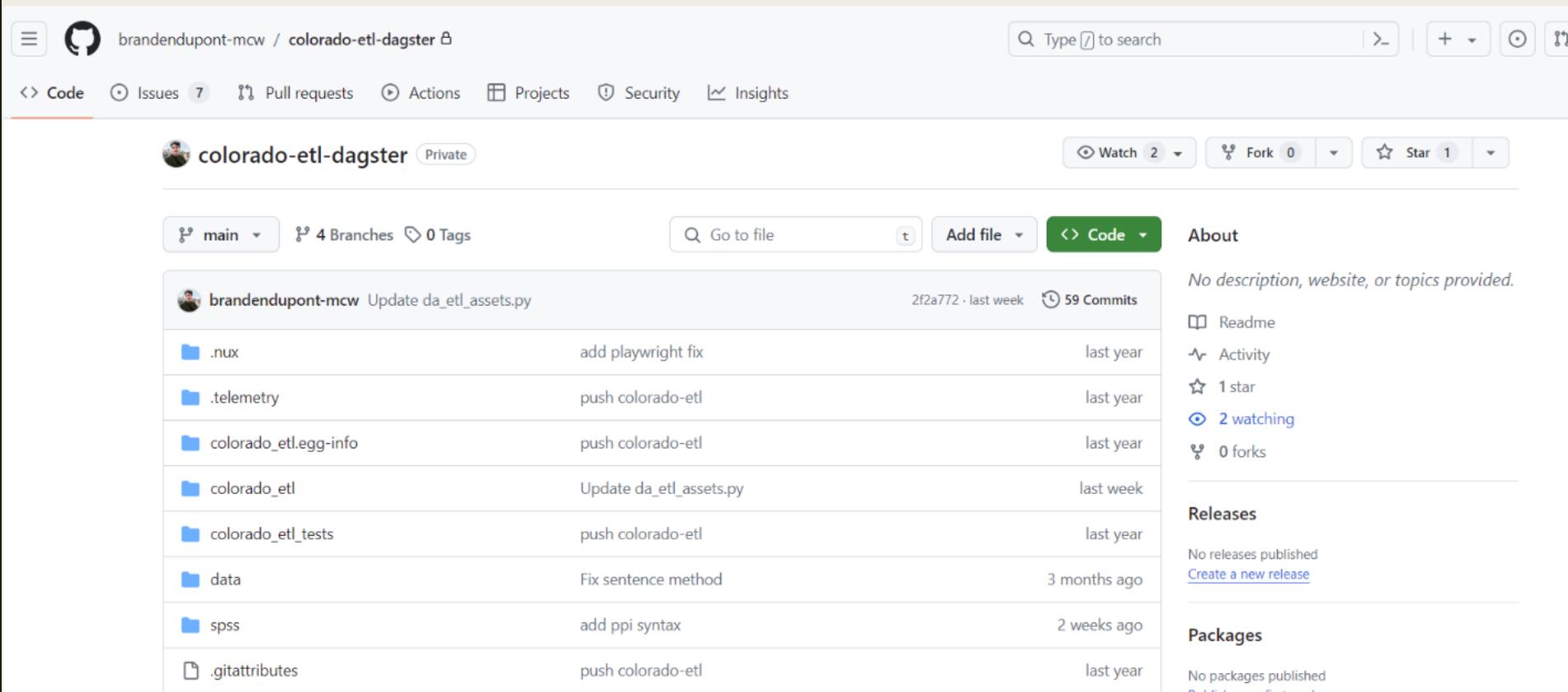
Step-by-step guide to be a collaborator on colorado-etl-dagster repository

■ 3. Access the Repository Online

a. Go to the Repository

- After accepting the invitation, navigate to the repository by going to

<https://github.com/brandendupont-mcw/colorado-etl-dagster>



The screenshot shows the GitHub interface for the repository 'colorado-etl-dagster' by user 'brandendupont-mcw'. The repository is private and has 1 star, 0 forks, and 2 watchers. The main branch is 'main' with 4 branches and 0 tags. The repository contains several directories and files, with the most recent commit being 'Update da_etl_assets.py' by 'brandendupont-mcw' last week, with 59 commits. The 'About' section indicates no description, website, or topics are provided. The 'Releases' section shows no releases published, and the 'Packages' section shows no packages published.

File/Directory	Commit Message	Commit Date
.nix	add playwright fix	last year
.telemetry	push colorado-etl	last year
colorado_etl.egg-info	push colorado-etl	last year
colorado_etl	Update da_etl_assets.py	last week
colorado_etl_tests	push colorado-etl	last year
data	Fix sentence method	3 months ago
spss	add ppi syntax	2 weeks ago
.gitattributes	push colorado-etl	last year

Step-by-step guide to be a collaborator on colorado-etl-dagster repository

■ 4. Interact with the Repository

a. Browse Files

On the main page of the repository, you'll see a list of files and directories. Navigate to the 'spss' folder by clicking on it to explore its contents. Here, you'll discover the SPSS scripts associated with each table received from CDAC, utilized in the ETL process.

b. Download the Repository

If you want to download the entire repository as a ZIP file:

I. Go to the repository's main page on GitHub.

II. Click the "Code" button.

III. In the dropdown menu, click "Download ZIP."

IV. The ZIP file will be downloaded to your computer, and you can unzip it to access the files.

brandendupont-mcw / colorado-etl-dagster

Search: Type to search

Code Issues 7 Pull requests Actions Projects Security Insights

colorado-etl-dagster Private Watch 2 Fork 0 Star 1

main 4 Branches 0 Tags

brandendupont-mcw Update da_etl_assets.py 2f2a772 · last week 59 Commits

File/Directory	Commit Message	Time
.nux	add playwright fix	last year
.telemetry	push colorado-etl	last year
colorado_etl.egg-info	push colorado-etl	last year
colorado_etl	Update da_etl_assets.py	last week
colorado_etl_tests	push colorado-etl	last year
data	Fix sentence method	3 months ago
spss	add ppi syntax	2 weeks ago
.gitattributes	push colorado-etl	last year

Releases: No releases published. Create a new release

Packages: No packages published

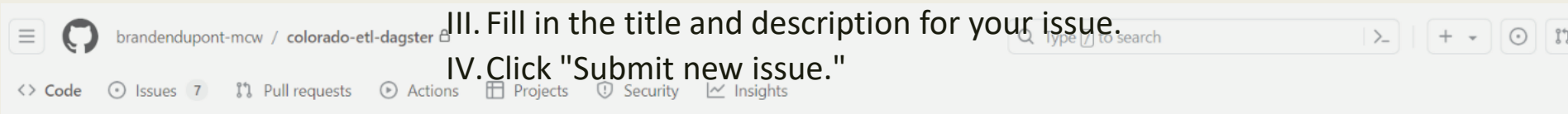
Step-by-step guide to be a collaborator on colorado-etl-dagster repository

■ 4. Interact with the Repository (cont...)

c. Create Issues

- If you encounter a bug or have a suggestion:

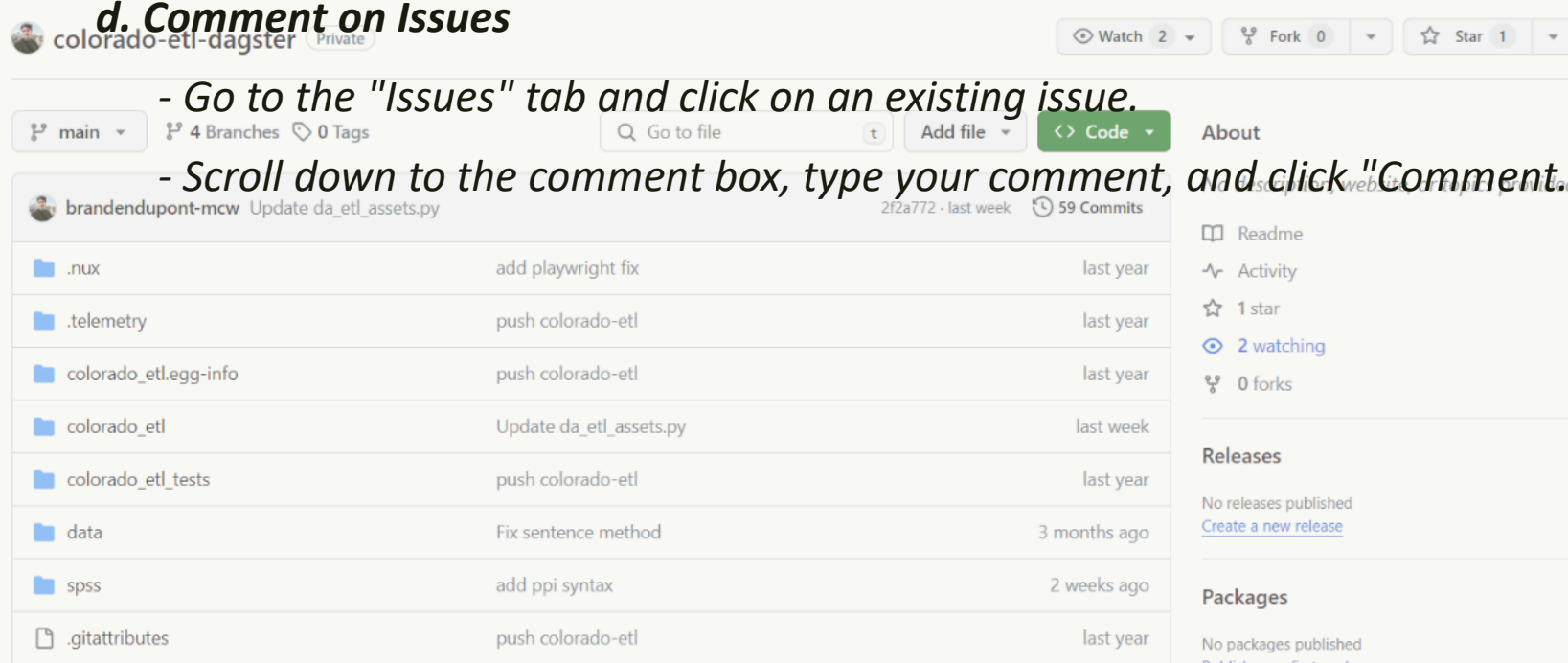
- I. Click on the "Issues" tab.
- II. Click the "New issue" button.
- III. Fill in the title and description for your issue.
- IV. Click "Submit new issue."



d. Comment on Issues

- Go to the "Issues" tab and click on an existing issue.

- Scroll down to the comment box, type your comment, and click "Comment."



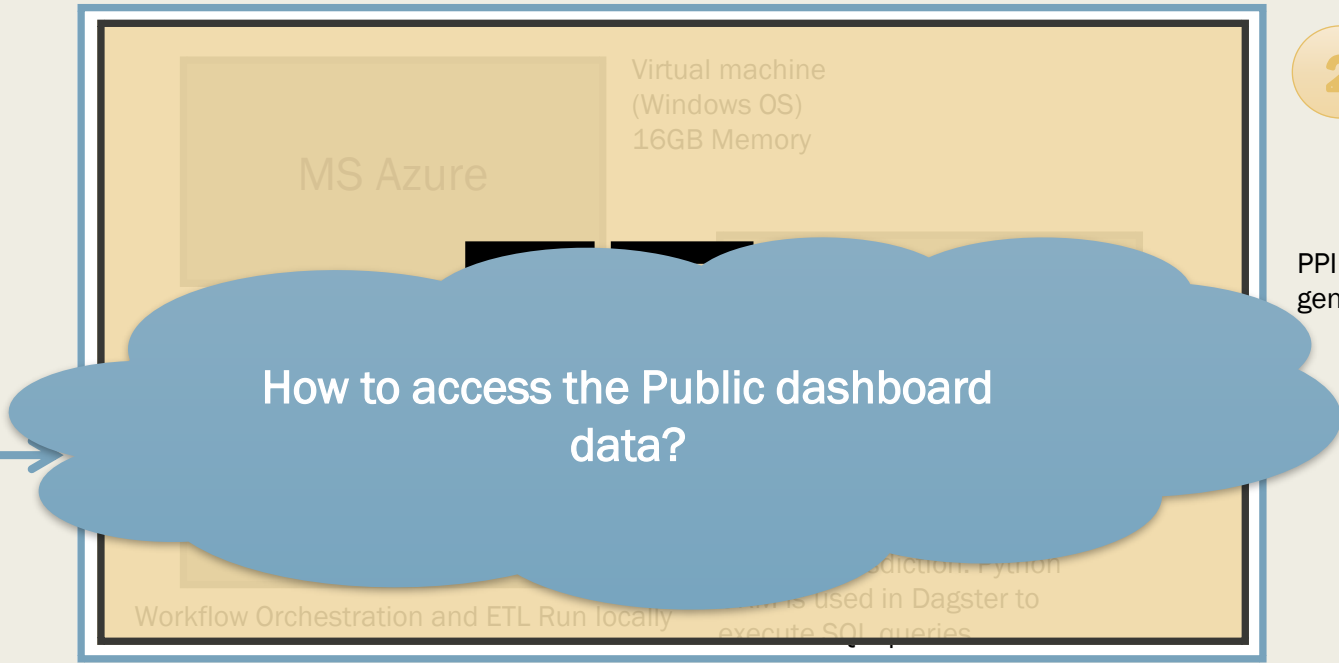
Dashboard Data Preparation Workflow

2

PPI clean the data and generate research table

1
CDAC Web client

CDAC share raw data to PPI via web server



3

Search Table

Cleaned Large flat files

4

Generate Public Facing dashboard chart data

Used to store aggregated chars data (.csv files)

5
GitHub

Generate charts for the Public dashboards

6
Datawrapper charts

7

Generate Internal dashboard chart data

Large flat files stored in MS Azure

8

MS Azure Blob Storage

9

Power BI charts

Generating charts for the Internal dashboards

Access Public dashboard data

- <https://github.com/dstemenluc/CCJ.Colorado>

(We do not need to login to GitHub account to access the public dashboard charts data)

The screenshot displays the GitHub repository page for **CCJ.Colorado**, which is public. The repository is owned by **dstemenluc** and contains 29 commits. The main content area shows a list of folders (JD1 to JD20) with their respective commit messages and dates. The right sidebar contains 'About' information, including activity, stars, watching, and forks, as well as 'Releases' and 'Packages' sections.

Folder	Commit Message	Commit Date
JD1	Deleting 2017 and 2018 data	2 months ago
JD11	New JD data updates	2 months ago
JD12	Quarter 1 2024 update	2 months ago
JD13	New JD data updates	2 months ago
JD17	JD 17 data story	last month
JD18	Deleting 2017 and 2018 data	2 months ago
JD19	New JD data updates	2 months ago
JD2	Deleting 2017 and 2018 data	2 months ago
JD20	Deleting 2017 and 2018 data	2 months ago

Dashboard Data Preparation Workflow

2

PPI clean the data and generate research table

1
CDAC Web client

CDAC share raw data to PPI via web server

How to access the data files used in Power BI report?

Virtual machine (Windows OS) 16GB Memory

MS Azure

Workflow Orchestration and ETL Run locally

execute SQL queries

3

Research Table

Cleaned Large flat files

4

Generate Public Facing dashboard chart data

Used to store aggregated chars data (.csv files)

5
GitHub

Generate charts for the Public dashboards

6
Datawrapper charts

7

Generate Internal dashboard chart data

Large flat files stored in MS Azure

8
MS Azure Blob Storage

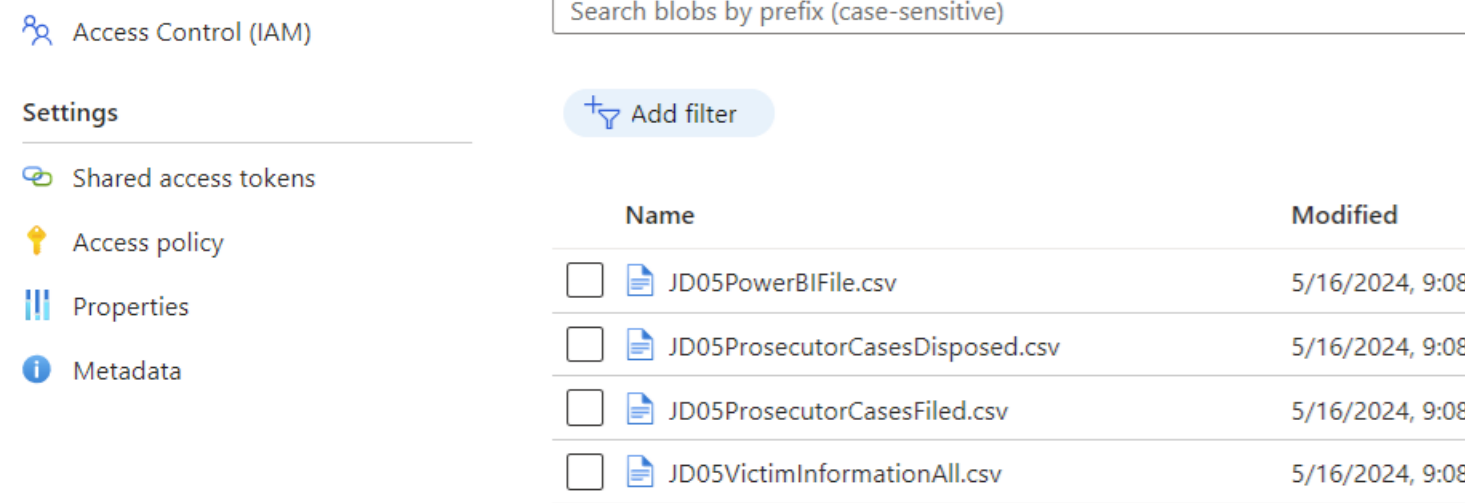
9

Power BI charts





Generating charts for the Internal dashboards

Access Power BI report data files

- CDAC can provide you with access to the data files used in preparing the Power BI report. Please inform us if you're interested in gaining this access. CDAC will offer necessary instructions and support.



The screenshot shows the Azure Storage Explorer interface. On the left, there is a sidebar with navigation options: Access Control (IAM), Settings, Shared access tokens, Access policy, Properties, and Metadata. The main area displays a search bar with the text "Search blobs by prefix (case-sensitive)" and an "Add filter" button. Below the search bar is a table listing four CSV files. Each row includes a checkbox, a document icon, the file name, and the modification date and time.

	Name	Modified
<input type="checkbox"/>	 JD05PowerBIFile.csv	5/16/2024, 9:08
<input type="checkbox"/>	 JD05ProsecutorCasesDisposed.csv	5/16/2024, 9:08
<input type="checkbox"/>	 JD05ProsecutorCasesFiled.csv	5/16/2024, 9:08
<input type="checkbox"/>	 JD05VictimInformationAll.csv	5/16/2024, 9:08