Pip Technical Guidelines

R.Andres Castenada Ronak Shah

2025-08-03

Table of contents

Preface

This is a Quarto book. In this book we intend to write technical information about PIP projects.

1 Introduction

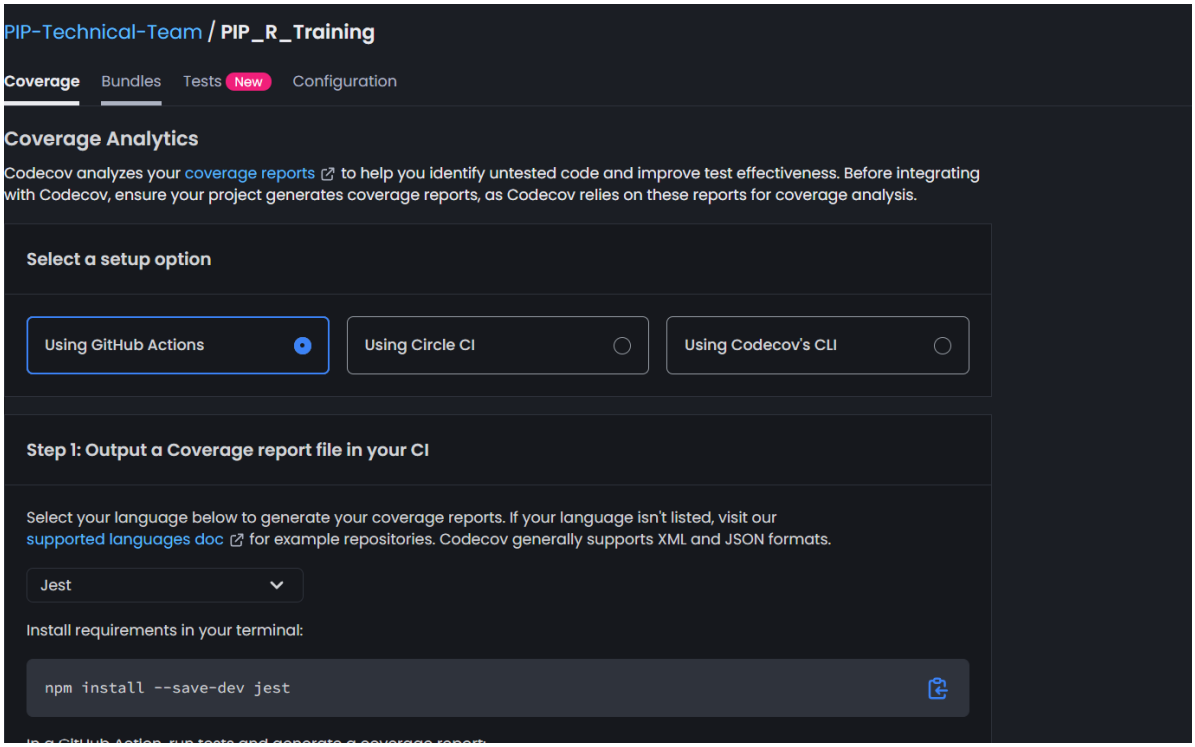
The purpose of this book is to gather all the technical knowledge specific to PIP in one place.

2 Add code coverage badge to your GitHub Repository.

In this article we will learn how to add code coverage badge to your GitHub repository.

2.1 Codecov.io

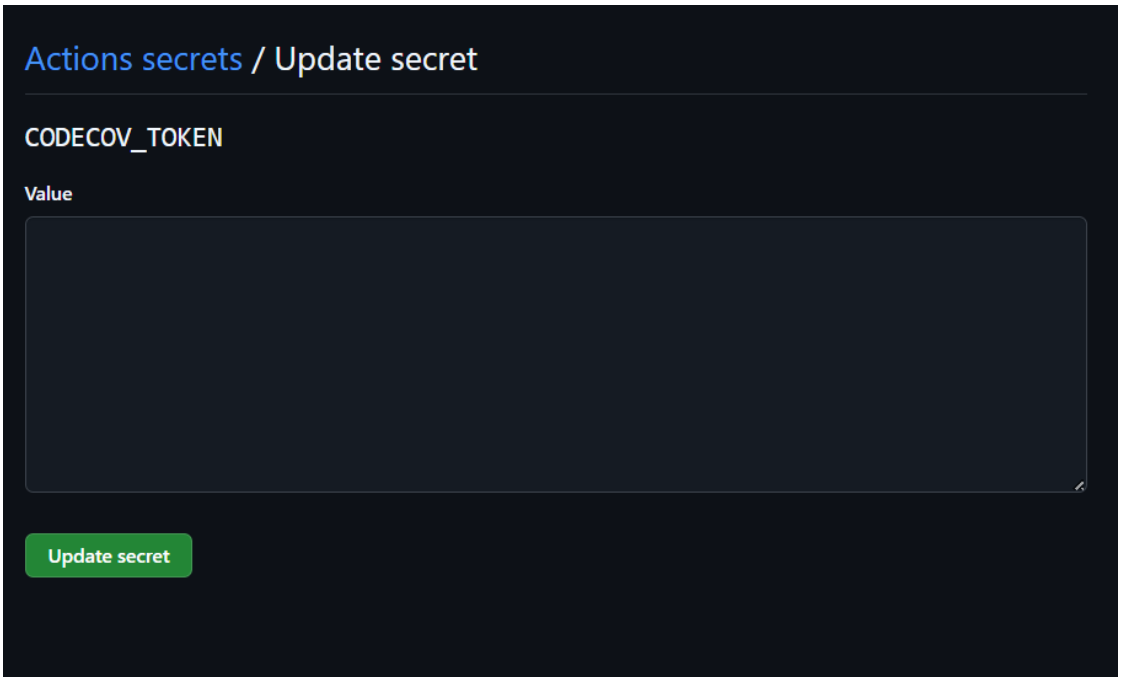
- Create an account at https://about.codecov.io/ , sign up with your GitHub account if you don't have an account already. **Codecov** is a popular tool for measuring and visualizing **code coverage** in software projects. It integrates with GitHub, GitLab, Bitbucket, and other CI/CD systems to provide insights into how much of your code is tested by your test suite.
- You can sync your private Github repositories on codecov platform to get started. If you want to add code coverage badge to a repository which is part of an organization (like PIP-Technical-Team, GPID-WB etc) then you need to be an admin of that organization. Admin role is needed because to sync the communication between Codecov.io with GitHub we need to generate a token which can only be done by admins.
- Once your repo is synced with codecov and you can see it there click on Configure to start the process. As an example it should give you this screen



- If you scroll below it will ask you to generate a repository secret, click on that to get a unique token for your repository and copy it.
- You can ignore rest of the steps mentioned on that page since those are very generic language agnostic steps and since we want to setup this for R packages, we have a better option which I will share below.

2.2 GitHub

- Now, moving to GitHub go to your repository. Click on Settings -> Secrets and Variables -> Actions -> Repository Secrets add the new token with name `CODECOV_TOKEN` and copy the token value which was generated in the previous step.



- Next, we are going to setup GitHub Action to run and calculate code coverage after every push. The calculated coverage report would be uploaded on codecov.io and would be visible on their dashboard.
- Additionally, I also added a possibility to run `R CMD CHECK` after every push. `R CMD check` is a tool that runs a series of automated checks on an R package to ensure it's correctly structured, documented, and error-free. It helps catch issues in code, tests, and documentation before sharing or submitting to CRAN. So it is like an additional validation that we have on our code.
- The new workflow file looks like below

```yaml
name: R-CMD-check and Codecov

on:
  push:
    branches: [master]
  pull_request:
    branches: [master]

jobs:
  R-CMD-check:
    runs-on: ubuntu-latest

    steps:
```

```
      - name: Checkout repository
        uses: actions/checkout@v4

      - name: Set up R
        uses: r-lib/actions/setup-r@v2

      - name: Set up pandoc
        uses: r-lib/actions/setup-pandoc@v2

      - name: Install dependencies
        run: |
          install.packages(c("remotes", "rcmdcheck", "covr"))
          remotes::install_deps(dependencies = TRUE)
        shell: Rscript {0}

      - name: Run R CMD check
        run: |
          rcmdcheck::rcmdcheck(args = "--no-manual", error_on = "warning")
        shell: Rscript {0}

      - name: Run test coverage
        run: |
          covr::codecov()
        shell: Rscript {0}
        env:
          CODECOV_TOKEN: ${{ secrets.CODECOV_TOKEN }}
```

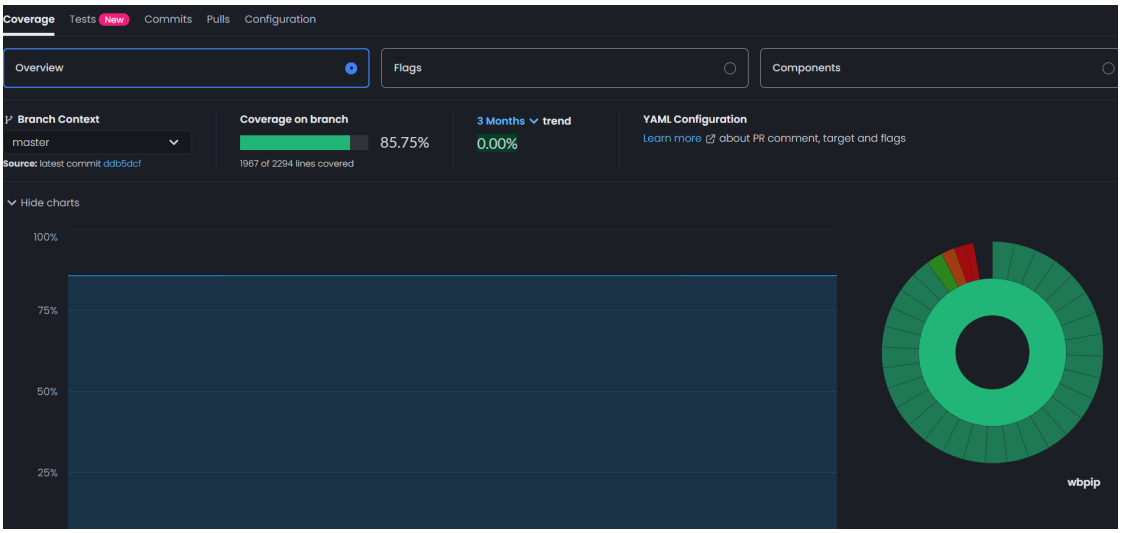
- This file is self explanatory but briefly, it checks out the repository that we want to run our action on, sets up R to run `R CMD CHECK` and finally generate code coverage report and upload it to codecov.io .
- One tip that I can share is to check if this workflow file works on your local branch before running on `master` branch. To do that you should temporarily enable the workflow file to run on your local branch. This can be done as below -

```
on:
  push:
    branches: [master, your-branch]
```

  where `your-branch` is the name of the local branch that you want to run the workflow for. Once you have verified that everything works as expected in the local branch, you can remove `your-branch` from the list again.

- Once the workflow runs successfully the dashboard on codecov.io should be updated and

you should see something like this



- Every time a push or PR is made to `master` the dashboard will be updated with latest data.

3 Setting up Github Actions for Auto Deployment of Quarto book

3.1 Introduction

One of the best parts of using Quarto for websites, blogs, or reports is how easily it integrates with GitHub Pages. With a simple GitHub Actions workflow, you can automatically render and publish your site every time you update your repository. In this post we are going to learn how we have enabled auto deployment for this quarto book.

3.2 Workflow

This is the workflow that we are using in Github Actions . Let's look at it one by one.

```yaml
on:
  workflow_dispatch:
  push:
    branches: main

name: Quarto Publish

jobs:
  build-deploy:
    runs-on: ubuntu-latest
    permissions:
      contents: write
    steps:
      - name: Check out repository
        uses: actions/checkout@v4

      - name: Set up Quarto
        uses: quarto-dev/quarto-actions/setup@v2
        env:
          GH_TOKEN: ${{ secrets.GITHUB_TOKEN }}
```

```
        with:
          tinytex: true
      - name: Render and Publish
        uses: quarto-dev/quarto-actions/publish@v2
        with:
          target: gh-pages
        env:
          GITHUB_TOKEN: ${{ secrets.GITHUB_TOKEN }}
```

3.2.1 Triggering the Workflow

```
on:
  workflow_dispatch:
  push:
    branches: main
```

This tells GitHub Actions when to run the workflow. There are two triggers here:

- **`push to main`** – Any time you commit or merge changes into the `main` branch, the workflow runs
- **`workflow_dispatch`** – Allows you to manually trigger the workflow from the GitHub Actions tab in your repository. This is useful when you want to force a rebuild and republish without committing new changes.

3.2.2 Naming the workflow

```
name: Quarto Publish
```

This gives the workflow a friendly name that will appear in the Actions tab.

3.2.3 Defining the job

```
jobs:
  build-deploy:
    runs-on: ubuntu-latest
    permissions:
      contents: write
```

Here we're defining a single job called `build-deploy`.

- **`runs-on: ubuntu-latest`** – The job will run inside an Ubuntu-based virtual machine provided by GitHub.
- **`permissions: contents: write`** – The workflow needs permission to write to the repository (required for publishing to the `gh-pages` branch).

3.2.4 The Steps

3.2.4.1 1. Check out the repository

```yaml
- name: Check out repository
  uses: actions/checkout@v4
```

This makes your repository's files available in the workflow environment so Quarto can render your project.

3.2.4.2 2. Set up Quarto

```yaml
- name: Set up Quarto
  uses: quarto-dev/quarto-actions/setup@v2
  env:
    GH_TOKEN: ${{ secrets.GITHUB_TOKEN }}
  with:
    tinytex: true
```

This installs Quarto in the workflow environment. The `tinytex: true` option ensures LaTeX support is available for rendering PDFs. The `GH_TOKEN` is github token repository secret that is added in Repo settings -> Secrets and Variables -> Actions . It is used for authentication when publishing.

3.2.4.3 3. Render and Publish

```yaml
- name: Render and Publish
  uses: quarto-dev/quarto-actions/publish@v2
  with:
    target: gh-pages
  env:
    GITHUB_TOKEN: ${{ secrets.GITHUB_TOKEN }}
```

This step does two things:

1. **Renders** your Quarto project (turns `.qmd` files into HTML, PDF, or other output formats).
2. **Publishes** the output to the `gh-pages` branch, which GitHub Pages uses to serve your site. The `target: gh-pages` option ensures everything is pushed to the right branch.

3.3 Don't ignore the `.gitignore` file

Make sure that your `.gitignore` file excludes `_book`, `_site` folders. These are the folders where Quarto renders HTML/PDF files when testing them locally. These files should not be tracked since Github Actions will build them with our auto deployment process.

3.4 Conclusion :

With this workflow in place, the Quarto book will automatically rebuild and deploy whenever a push is made to the `main` branch or whenever we manually trigger the workflow. No more running commands locally or remembering to push generated files.

This is a clean, reproducible, and automated way to publish your Quarto projects using GitHub Pages. As a side not `usethis` package has a lot of good utility functions that helps you to set up similar workflow. You may explore using them. A good starting point is `usethis::use_github_action("render-quarto")`.

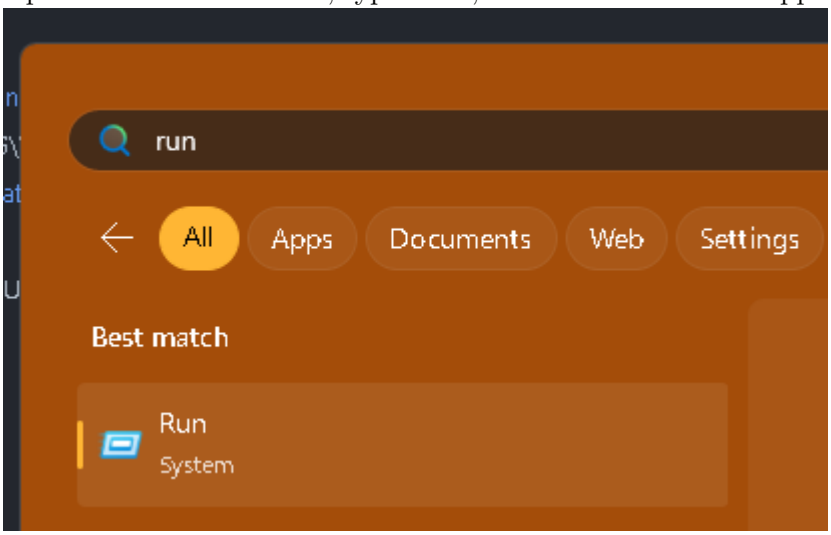
For reference the Quarto book is published here.

4 Improve Efficiency of a WB Laptop

There are a few things you can do to improve the efficiency of a WB laptop. These tips won't make your laptop *fast*, but they will help optimize performance and make it feel more responsive.

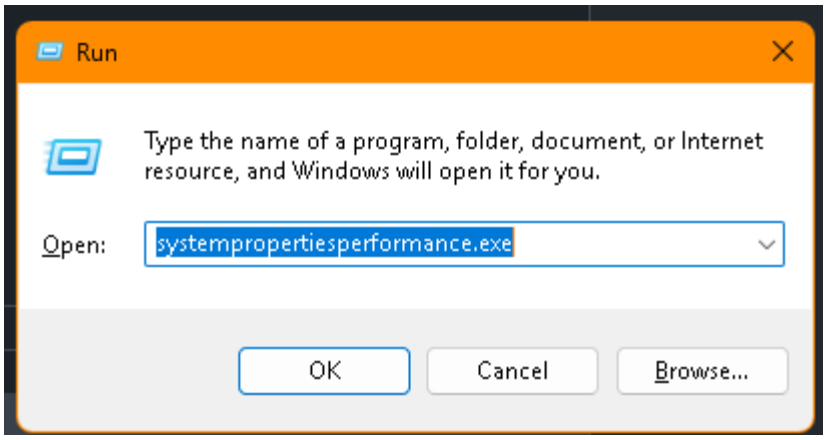
4.1 Modify System Properties for Performance

1. Open the Windows menu, type **Run**, and click on the *Run* app.



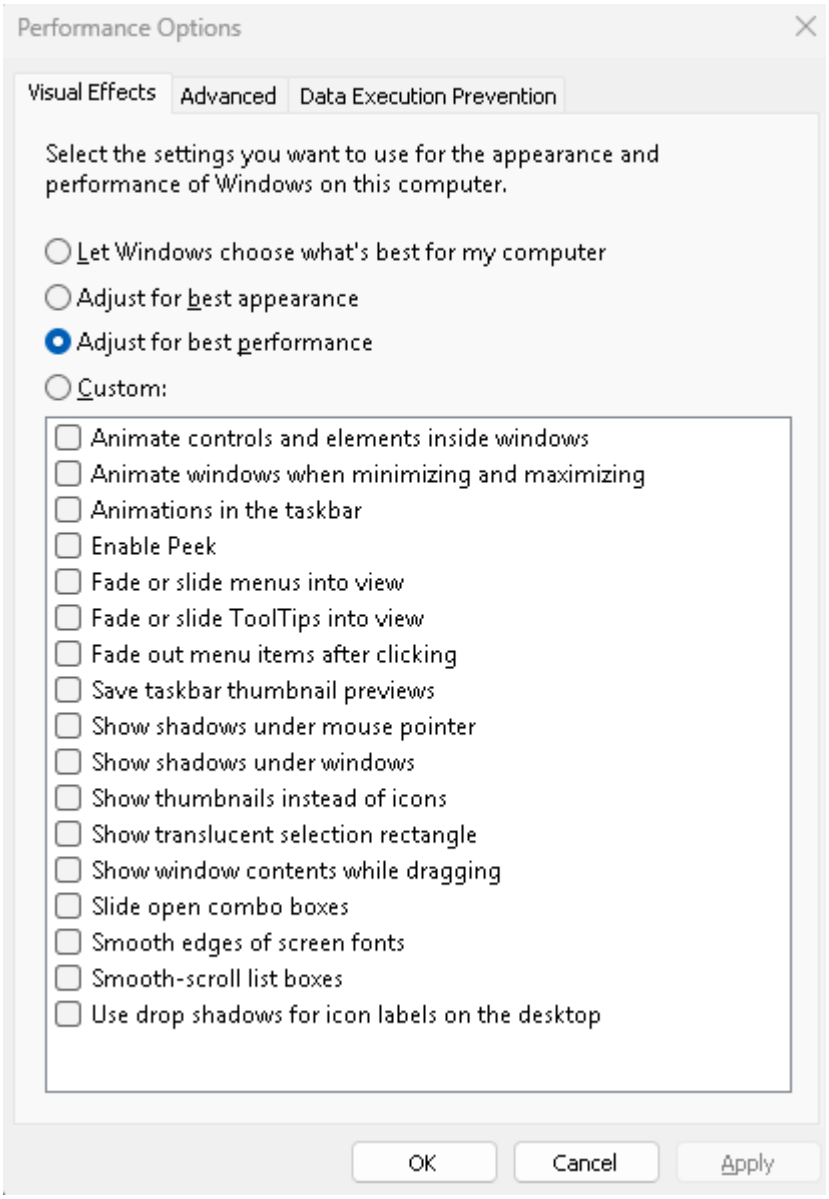
2. In the Run app, type the following command and hit **Enter**:

```
systempropertiesperformance.exe
```



3. In the *Performance Options* window, select the **Visual Effects** tab.

   - Click on **Adjust for best performance**.
   - Make sure all the checkboxes are unchecked.
   - Click **Apply** and then **OK**.

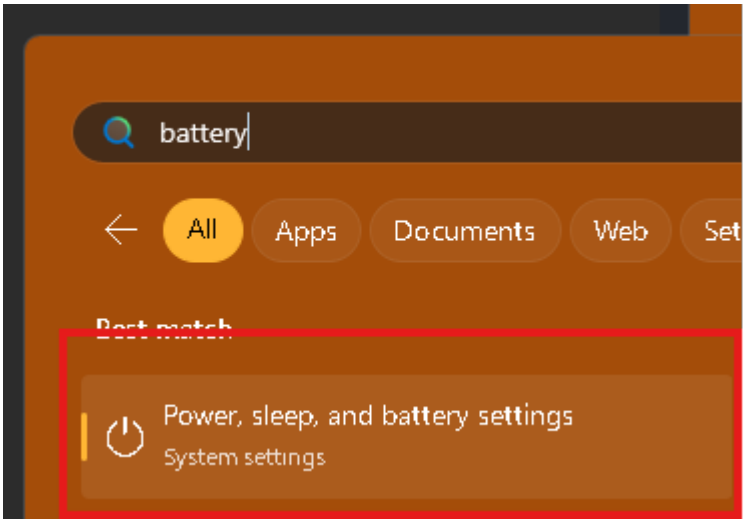


Warning

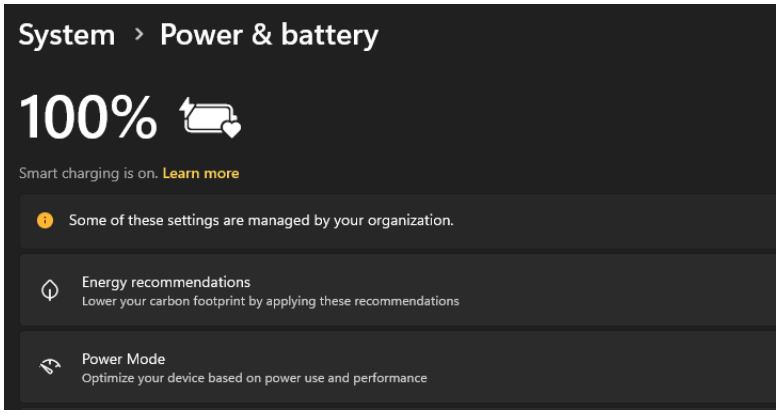
Note: This may affect the appearance of your system, since many visual effects will be disabled.

4.2 Battery Settings

1. Open **Battery Settings**.



2. Set the battery mode to **Best performance**.

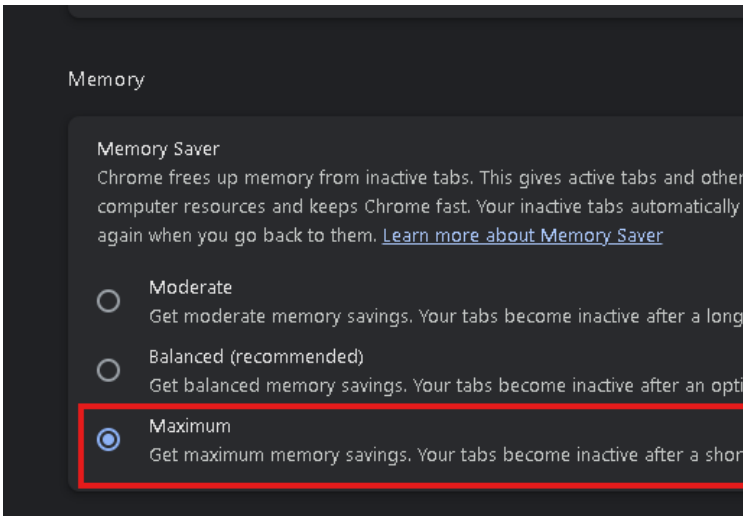


4.3 Google Chrome Settings

1. In Google Chrome, type the following in the address bar:

```
chrome://settings/performance
```

2. Under the **Memory** section, select **Maximum**.



5 How to interact with the Virtual Machine (VM) that hosts PIP

This guide explains how to connect to, inspect, and maintain the Linux Virtual Machines (VMs) hosting the PIP (Poverty & Inequality Platform) APIs. It covers user access, Docker container management, data storage, debugging, and safe cleanup procedures. This document is written for users with limited Linux/Docker experience who need to operate, inspect, or troubleshoot the PIP infrastructure safely.

5.1 Access and Session Management

5.1.1 Login to the VM

1. Go to PrivX.
2. Use **SSO login** (World Bank Authenticator App).
3. In the **Connections** tab, select the VM you want:
   - **Development:** `Linux-wzlxdpip01.worldbank.org`
   - **QA:** `Linux-wzlxqpip01.worldbank.org`
   - **Production:** `Linux-wzlxppip01.worldbank.org`
4. Once connected, switch to the `srvpip` user:

```
sudo su - srvpip
```

Shortcut	Action	Notes

5.1.2 Copy, Paste, and Keyboard Shortcuts

Shortcut	Action	Notes
Ctrl + C	Stop the current running process	e.g. stop an infinite loop or hanging R process
Ctrl + D	Log out of the current shell	Ends your session as `srvpip`
Shift + Ctrl + V	Paste from clipboard	Works in most PrivX terminals
Shift + Insert	Paste (alternative)	Often easier in remote shells

Notice that **Ctrl + C** does **not** copy text in terminal. You only need to highlight with your mouse the section you want to copy. Once you stopped highlighting, the text is copied to your clipboard automatically.

5.2 Understanding the Environment

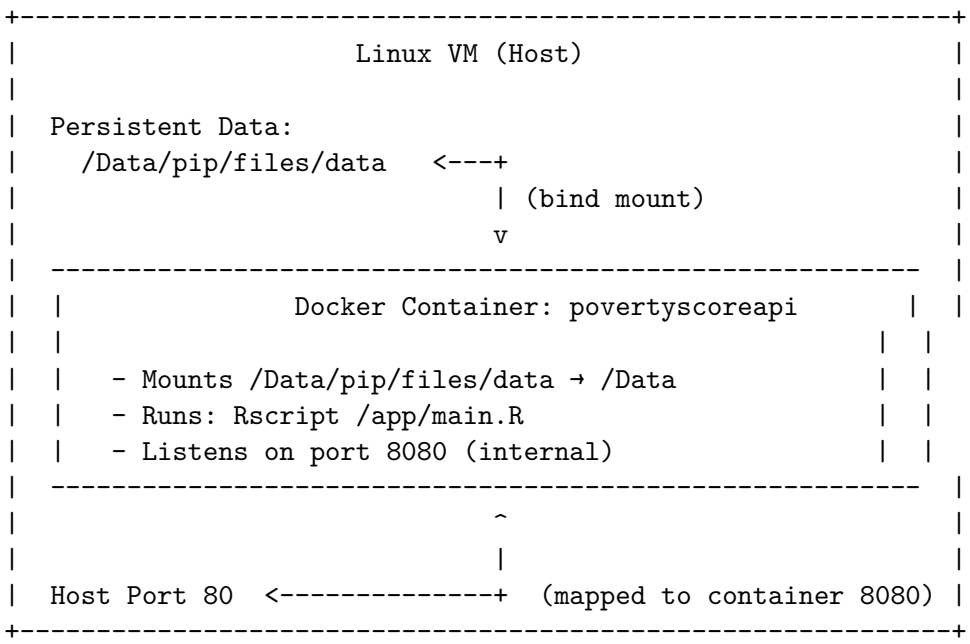
5.2.1 Users and Permissions

- **Your account:** initial login via PrivX (not root)
- **`srvpip` user:** used to manage Docker containers
- **`root` user:** required for system-level commands; use only with `sudo`

Common prefixes:

- `docker` → run Docker as normal user (often fails without permissions)
- `sudo docker` → correct way when managing containers
- `sudo su - srvpip` → switch to service account context

5.2.2 Architecture Overview



Legend: - Data written to /Data inside the container is actually stored on the host at /Data/pip/files/data. - The API runs inside the container on port 8080, but is accessible from outside the VM on port 80.

5.3 Essential Docker Commands (Cheat Sheet)

Task	Command	Notes
List all containers	`sudo docker ps -a`	Shows running & stopped
View logs	`sudo docker logs -f povertyscoreapi`	Add `--tail 200` to see last lines
Enter container shell	`sudo docker exec -it povertyscoreapi /bin/bash`	For interactive debugging

Task	Command	Notes
Check mounted volumes	`sudo docker inspect povertyscoreapi | grep Mounts -A 5`	See `/Data` source
Delete stopped containers	`sudo docker container prune`	Cleans unused ones
Delete old images	`sudo docker image prune -a`	Be cautious — removes all unused
Check Docker service status	`sudo systemctl status docker`	Confirms Docker is active

5.4 Managing Data

5.4.1 Data on the VM (Persistent)

- Main data location:

```
cd /Data/pip/files/data
```

- Inspect space usage:

```
df -h | grep Data
du -h --max-depth=1 | sort -h
```

- Preview files safely:

```
ls -lah
```

- View folder sizes:

```
du -sh *
```

5.4.1.1 Find large or old files

```
find /Data/pip/files/data -type f -mtime +365 | head -n 20
```

5.4.2 Data Inside Containers (Ephemeral)

Check where containers mount volumes:

```
sudo docker inspect povertyscoreapi | grep Mounts -A 5
```

You'll see:

```
"Mounts": [
  {
    "Type": "bind",
    "Source": "/Data/pip/files/data",
    "Destination": "/Data"
  }
]
```

That means:

- Inside container → `/Data`
- On host → `/Data/pip/files/data`

5.4.2.1 Access via container shell:

```
sudo docker exec -it povertyscoreapi ls -lh /Data
```

5.4.3 Safe Deletion and Cleanup

Preview first:

```
ls -d /Data/pip/files/data/project_*
```

Delete specific folders:

```
rm -rf /Data/pip/files/data/folder1 /Data/pip/files/data/folder2
```

Delete all contents but keep parent folder:

```
rm -rf /Data/pip/files/data/*
```

Move to trash instead of deleting:

```
mkdir -p /Data/pip/files/trash
mv /Data/pip/files/data/folder1 /Data/pip/files/trash/
```

Caution: `rm -rf` is irreversible — always check with `ls` first.

5.5 Diagnosing API or Container Issues

5.5.1 Check Container State

```
sudo docker ps -a
```

Get detailed exit info:

```
sudo docker inspect povertyscoreapi --format='ExitCode={{.State.ExitCode}} OOMKilled={{.State
```

- `ExitCode=0` → normal exit
- `ExitCode=137` → killed (likely out of memory)

5.5.2 View Logs

```
sudo docker logs povertyscoreapi | head -n 20   # startup logs
sudo docker logs --tail 200 povertyscoreapi     # recent logs
```

5.5.3 Reproduce Interactively

Start an interactive debug session:

```
sudo docker run --rm -it \
  --name povertyscoreapi-debug \
  -p 8080 \
  -v /Data/pip/files/data:/Data \
  itsesippscoreregistryprod.azurecr.io/povertyscoreapi:latest \
  /bin/bash
```

Inside:

```
Rscript /app/main.R
```

Press `Ctrl + C` to stop and `exit` to leave.

5.5.4 Restart or Rebuild Containers

```
sudo docker restart povertyscoreapi
```

Or rebuild completely:

```
sudo docker stop povertyscoreapi
sudo docker rm povertyscoreapi
sudo docker pull itsesippscoreregistryprod.azurecr.io/povertyscoreapi:latest
sudo docker run -d --name povertyscoreapi \
  -p 80:8080 \
  -v /Data/pip/files/data:/Data \
  itsesippscoreregistryprod.azurecr.io/povertyscoreapi:latest
```

5.6 Testing API Endpoints

5.6.1 From Inside the Container

```
curl http://localhost:80/api/v1/health-check
```

5.6.2 Loop through common endpoints

```
for ep in health-check pkgs-version data-signature gh-hash; do
  echo "---- Testing $ep ----"
  curl -s http://localhost:80/api/v1/$ep
  echo
done
```

5.6.3 From Outside (VM host or browser)

```
curl http://wzlxdpip01.worldbank.org/api/v1/health-check
```

5.7 Troubleshooting Common Errors

Symptom	Likely Cause	Fix
`curl: (7) Failed connect`	Container not running or wrong port	`sudo docker ps -a`
`Exited (137)`	Out of memory (OOMKilled)	Increase VM memory or limit API load
`authentication required`	Not logged in to Azure Container Registry	`sudo az acr login --name itsesippscoreregistryprod`
`permission denied` on `/Data`	Wrong user ownership	`sudo chown -R srvpip /Data`
No logs shown	Container failed before startup	Run `sudo docker inspect povertyscoreapi`

Symptom	Likely Cause	Fix
Port 8080 works but 80 doesn't	Port not mapped	Run container with `-p 80:8080`

5.8 Understanding Ports (80 vs 8080)

- Inside Docker, the API runs on **8080**:

```
pipapi::start_api(port = 8080, host = "0.0.0.0")
```

- Outside Docker (host or browser), you can map any port using:

```
-p 80:8080
```

  This means:

  `host_port:container_port`

  So users access:

  `http://wzlxdpip01.worldbank.org/api/v1/...`

  even though internally it listens on 8080.

5.9 Monitoring Resource Usage

Watch real-time CPU and memory:

```
sudo docker stats povertyscoreapi
```

Or globally:

```
top -u srvpip
```

5.10 Appendix

5.10.1 Check Disk Usage Quickly

```
df -h | grep Data
```

5.10.2 Check Docker Service Logs

```
sudo journalctl -u docker --since "2025-09-23 17:00"
```

5.10.3 Check R version inside container

```
sudo docker exec -it povertyscoreapi Rscript -e "R.version.string"
```

6 Summary

References