PowerPoint Cleanup Tool - Linux/Bash/WSL Guide

Remove unused images, slide masters, and layouts from PowerPoint (.pptx) files to reduce file size.

Overview

PowerPoint files often contain unused content from:

- Copied slides that brought their own masters/layouts
- Deleted slides that left behind images
- Multiple themes merged together

This tool identifies and removes that bloat using Bash scripts (WSL on Windows, or native Linux/Mac).

Install

WSL (Windows Subsystem for Linux)

WSL lets you run Linux commands and bash scripts on Windows. Install from PowerShell (Admin):

wsl --install

Restart your computer, then open "Ubuntu" from Start menu to complete setup.
See Microsoft WSL Documentation for details.

Required Tools

In WSL/Linux terminal, ensure zip and unzip are installed:

sudo apt update

sudo apt install zip unzip

Python (for the cleanup script)

Python is usually pre-installed in Ubuntu. Verify:

python3 --version

If not installed:

sudo apt install python3

Usage

Method 1: One-Command Cleanup (Recommended)

The bash wrapper script handles everything automatically:

./cleanup.sh input.pptx output.pptx

This automatically:

1. Unzips the presentation
2. Removes unused images and layouts
3. Re-zips to output file
4. Reports size savings

If the script won't run, make it executable first:

chmod +x cleanup.sh

Method 2: Step-by-Step

Step 1: Unzip the PowerPoint

PowerPoint files are ZIP archives. Choose one method:

Option A: Manual (File Manager)

1. Make a copy of your .pptx file (always keep a backup!)
2. Rename the copy from .pptx to .zip
3. Right-click the .zip file → "Extract Here" or "Extract to..."
4. Extract to a folder (e.g., presentation)

Option B: Command Line

unzip presentation.pptx -d presentation/

Step 2: Analyze

python3 pptx_cleanup.py ./presentation/

This produces a report showing:

- Active vs unused masters
- Active vs unused layouts
- Active vs unused images
- Total space that can be reclaimed

Step 3: Remove Unused Content

 # Safe: Remove only unused images

python3 pptx_cleanup.py ./presentation/ --remove-images

 # Moderate: Remove unused layouts (updates [Content_Types].xml)

python3 pptx_cleanup.py ./presentation/ --remove-layouts

 # Advanced: Remove unused masters (also updates XML files)

python3 pptx_cleanup.py ./presentation/ --remove-masters

 # Combined: Remove images and layouts together

python3 pptx_cleanup.py ./presentation/ --remove-images --remove-layouts

Step 4: Re-zip as PowerPoint

Choose one method:

Option A: Manual (File Manager)

1. Open the presentation folder
2. Select ALL contents inside (Ctrl+A)
3. Right-click → "Compress" or "Create Archive"
4. Save as a .zip file outside the presentation folder
5. Rename from .zip to .pptx

IMPORTANT: Zip the CONTENTS of the folder, not the folder itself! Zipping the parent folder will cause PowerPoint to fail to open the file.

Option B: Command Line

cd presentation

zip -r ../cleaned.pptx .

cd ..

Note: The cd into the folder and using . ensures you zip the contents, not the folder itself.

Step 5: Test

Open the cleaned .pptx in PowerPoint and verify all slides display correctly.

Method 3: Using Generated Scripts

After running the analyzer, bash scripts are generated in the presentation folder:

cd presentation/

 # Run any or all of these:

bash remove_unused_images.sh

bash remove_unused_layouts.sh

bash remove_unused_masters.sh

Or review unused_components.txt and delete files manually.

Files

File	Description
pptx_cleanup.py	Main Python script - analyzes and cleans
cleanup.sh	Bash wrapper for one-command cleanup
README_WSL.md	This documentation (Bash/WSL)
README_PY.md	Documentation for Python/Windows users

Generated Files (in presentation folder)

File	Description
remove_unused_images.sh	Script to remove unused images
remove_unused_masters.sh	Script to remove unused masters
remove_unused_layouts.sh	Script to remove unused layouts
unused_components.txt	List of all unused components
backup_YYYYMMDD_HHMMSS/	Backup of XML files (when using --remove-masters)

What Gets Removed

Safe to Remove (--remove-images)

- Images in ppt/media/ not referenced by any active slide, master, or layout
- No XML editing required

Moderate (--remove-layouts)

- Slide layouts not used by any active slide or master
- Updates [Content_Types].xml
- Creates backup before changes
- Often the biggest source of bloat in merged presentations

Advanced (--remove-masters)

- Slide masters not used by any active slide
- Updates presentation.xml and [Content_Types].xml
- Creates backup before changes

How It Works

PowerPoint files are ZIP archives containing XML and media:

presentation.pptx (renamed .zip)

├── [Content_Types].xml # File type declarations

├── ppt/

│ ├── presentation.xml # Slide and master references

│ ├── slides/ # Active slides

│ ├── slideMasters/ # Master templates

│ ├── slideLayouts/ # Layout templates

│ └── media/ # Images and media

The tool:

1. Parses presentation.xml to find active slides
2. Traces slide → layout → master chain
3. Identifies which media is referenced by active components
4. Marks unreferenced content for removal

Troubleshooting

"presentation.xml not found" Ensure you're pointing to the unzipped folder, not the .pptx file.

Script won't run Make it executable: chmod +x cleanup.sh

PowerPoint won't open cleaned file

- Make sure you zipped the CONTENTS of the folder, not the folder itself
- Restore from backup and try removing only images first

"zip" or "unzip" not found Install them: sudo apt update && sudo apt install zip unzip

apt install fails with 404 error Run sudo apt update first to refresh package lists.

Tips

- Always keep a backup of your original file

- Start by removing only images (safest)
- Test the cleaned file before deleting the original
- Large presentations with many merged themes benefit most
- Use the one-command cleanup.sh for quick processing