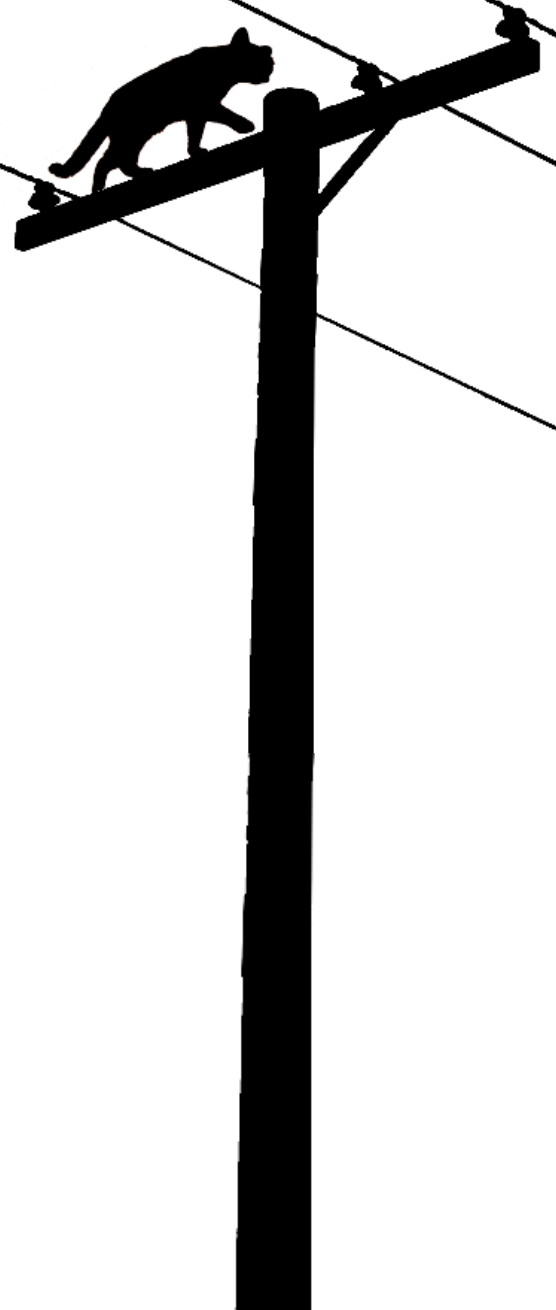


Public Utility Data Liberation

2025 Annual Report

Produced by Catalyst Cooperative
February 2026



Powered by
Catalyst Cooperative
catalyst.coop
hello@catalyst.coop

Made possible by our
generous Sustainers



Highlights

2025 was the first year of the [PUDL Sustainer’s Program](#). We designed the program to ensure the ongoing impact and reliability of [PUDL](#): data and software are kept up to date, bugs are fixed, and everything gets tested and released on a regular schedule. At the same time, we sought to foster closer relationships with the organizations that rely heavily on PUDL data for their work, giving them an opportunity to see under the hood and weigh in on important decisions.

Thanks to support from our inaugural sustainers, [RMI](#) and [GridLab](#), we were able to launch the program and keep PUDL up and running!

24,000 unique IPs Accessing PUDL data from cloud storage buckets	577 users Registered for the PUDL Data Viewer	70+ schools Accessing PUDL data
20 archives Of new data added to Zenodo	14 citations In academic journals, pre-prints, and theses (including Applied Energy , The Electricity Journal , and Environmental Research: Energy)	14 new tables Including imputed FERC 714 and EIA 930 data, SEC 10-K outputs tracking parent-subsidary company relationships, and AEO projected energy use data
10 conferences Attendance and presentation at USAEE, INFORMS, EPRI, TWEEDS and others!	8 terabytes Of PUDL data distributed through S3 and Zenodo	4 interviews In the MIT Technology Review , the New York Times , Salon and Les Echos

PUDL Sustainer Projects

Funding from PUDL Sustainers keeps our existing data up to date. Most PUDL data is published on an annual basis with some updating monthly or sub-annually. For annual data, we've committed to pulling new data into our nightly-builds within two weeks of its initial publication. For monthly data we run a quarterly update to capture new information. For extensive detail about all of the PUDL updates - see PUDL [release notes](#).

Annual Data Updates

We archived, integrated, and published the most recent year of data (2024) for our existing PUDL data sources:

FERC 1	EIA 860 Early & Final Release	EIA 923 Early & Final Release
FERC 714	EIA 861 Early & Final Release	EIA-EPA Crosswalk

NREL ATB was not published as expected in 2025 - we were not able to confirm why, but the likely explanation is priority and budget changes at NREL. We are cautiously optimistic and budgeted time to integrate it in 2026.

We archived the most recent year of data (2024) for some of our extracted, un-published datasets:

PHMSA Natural Gas	EIA 176	EIA AEO 2025 projections
--------------------------	----------------	------------------------------------

These datasets are partially integrated into PUDL, and these data updates keep these sources up to date to support future improvements. Thanks to funding from GridLab and Future of Heat Initiative and support from open-source contributors we were also able to transform and publish tables from these sources (see Other PUDL Projects section below).

We added the most recent year of data (2024) to our [SQLite databases for](#):

FERC 1	FERC 2	FERC 6	FERC 60	FERC 714
---------------	---------------	---------------	----------------	-----------------



Quarterly Data Updates

We continued our cadence of updating these sources on a quarterly basis, ingesting fresh quarterly data in February, May, August, and November:

EIA 860M	EIA 923	EIA 930	EPA CEMS
----------	---------	---------	----------

Thanks to funding from Madrone Advisors, we were able to start integrating EIA 860M data on a monthly basis (see Other PUDL Projects section below).

Infrastructure Upkeep

As the scale of data processed by PUDL increases, we periodically encounter new performance bottlenecks in the data pipeline that need to be addressed to ensure that we and volunteer contributors can continue to develop PUDL efficiently. The PUDL Sustainers program helped keep PUDL software tested and up to date, funding vital infrastructure updates that will reduce the long-term maintenance burden.

Keeping PUDL's infrastructure maintained includes the following:

- Running nightly builds of the PUDL pipeline.
- Publishing versioned releases of PUDL data.
- Ensuring that PUDL uses the most recent versions of the more than 650 open source software packages it depends on.
- Keeping example notebooks functional.
- Triaging and addressing user-reported bugs.
- Fixing failures in the nightly builds and usage metrics pipeline.
- Managing the flakiness in the Zenodo archive API.
- Updating to a new version of Python.

Special Projects

Outside of the projects that we planned for in 2025, we decided to spend about 10% of the total PUDL budget over 2025 on three projects that emerged during the year. We decided to invest time in these three projects because they had some external funding, but needed additional last-mile support to fully realize their benefits.

Pipeline Performance Improvements

We supplemented external funding to wrap up a round of improvements to the PUDL data processing pipeline - making it much faster and less resource intensive.

EQR Archiver Improvements

While scoping the integration of FERC EQR we captured the 100 GB dataset with a simple script. We decided to automate the process to ensure we'd always have fresh data down the road.

PHMSA Natural Gas

In order to increase access to 50 years of PHMSA Gas data we had extracted, we spent the extra time to transform and [publish eight tables](#) about gas distribution infrastructure.

Budget Breakdown

PUDL started 2025 underfunded for the year, causing us to proceed with caution. We had raised enough funds for the first quarter, but the second, third, and fourth, were below our desired threshold. To preserve access to resources throughout the year, we limited our work to the bare essentials and incurred a bit of technical debt. To the extent possible, we also supplemented PUDL updates and code maintenance with grant funding where appropriate.

By mid 2025, we secured funding for the rest of the year. Our frugality early on left us with surplus funds that we are carrying into 2026 and will use to fill funding gaps as necessary.

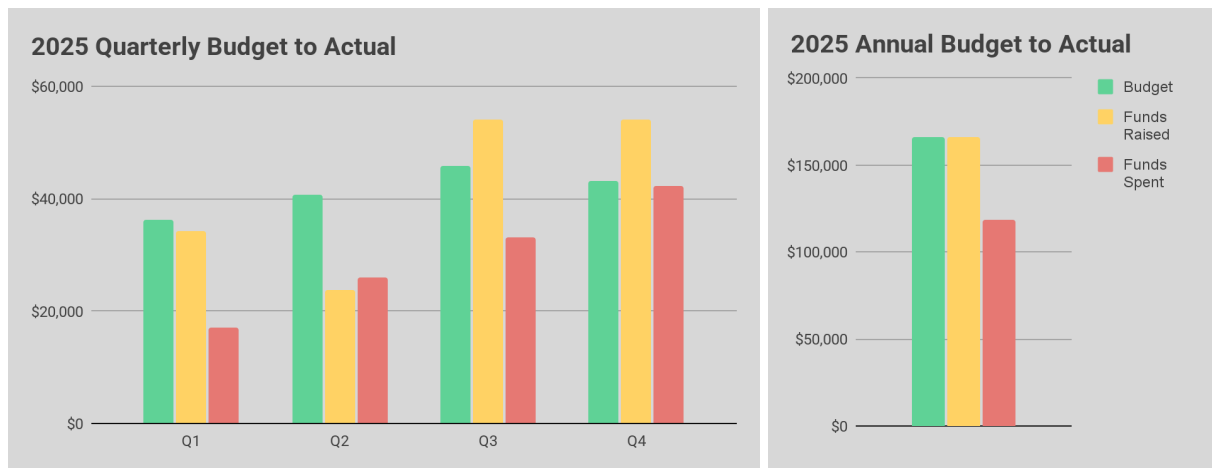


Fig. 1: 2025 PUDL Sustainer Budget broken down by quarter (left) and year (right)

Other PUDL Projects

Beyond the work that was directly funded by the PUDL Sustainers in 2025, we integrated several new datasets into PUDL and made improvements to how users interact with PUDL. This work was supported by grants, open-source contributions, and client funding.

Open-source contributors enabled the following data integration:

EIA-176

Developers at SwitchBox began integration of EIA Form 176 (Annual Report of Natural and Supplemental Gas Supply and Disposition) into PUDL. This data describes the origins, suppliers, and disposition of natural gas on a yearly and state by state basis.

EIA-923 Schedule 8

With help from contributors, we were able to begin integrating the tables in EIA-923 Schedule 8 Annual Environmental Information, which are some of the last EIA-923 tables which haven't been integrated into PUDL. These tables provide information on how plants are storing and disposing of byproducts from power generation.



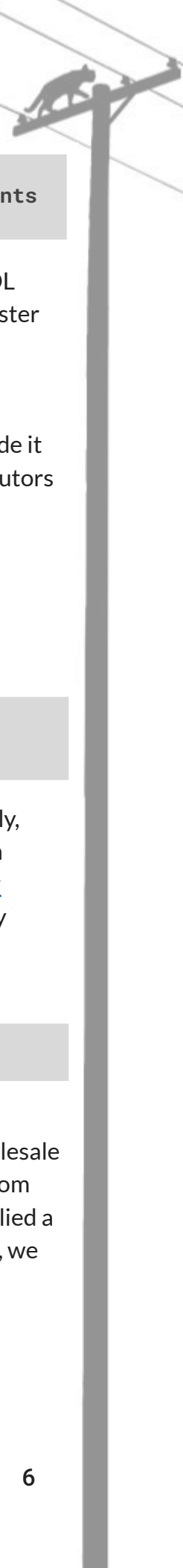
Utilizing funds from our NSF Pathways to Enable Open Source Ecosystems (POSE) grant we made the following improvements:

PUDL Data Viewer

Phased out our previous data viewing UI, Datasette, and launched our new data viewing tool, the [PUDL Data Viewer](#). The PUDL Data Viewer is an interactive data exploration portal for users to search for, preview and download PUDL data. By the end of 2025, nearly 600 users had created an account.

Data Validation

Overhauled our data validation infrastructure by adopting a tool called *dbt*. This enables more comprehensive, maintainable tests that catch upstream data changes and pipeline bugs, and makes our data quality checks more robust and scalable by enabling a broader suite of standardized tests.



Usage Metrics Pipeline

Developed a usage metrics data pipeline and dashboard, providing us more insight into where to conduct outreach to new users and prospective PUDL Sustainers. The usage metrics data additionally enables us to track overall impact and understand which PUDL data sets are most utilized.

Pipeline Performance Improvements

Improved the performance of the PUDL data processing pipeline to be much faster and less resource intense. These improvements eased development of PUDL, lowered the compute and maintenance burden of PUDL, and made it easier to onboard open source contributors to the project.



With additional support from GridLab we integrated or improved upon the following datasets:

EIA-930 Time Series Imputations

Working with researchers at the Climate and Energy Lab at [Carnegie Institute for Science](#), we imputed hourly electricity demand using EIA 930 data from PUDL, producing a complete time series dataset for open energy modelers.

VCE RARE

We integrated another 6 years of hourly, county-level wind and solar generation profiles from the [Vibrant Clean Energy Resource Adequacy Renewable Energy \(RARE\)](#) dataset, which now covers 2014-2024.

FERC EQR Initial Draft

We were able to make the FERC Electric Quarterly Reports (EQR) accessible for programmatic access alongside PUDL data. FERC's EQR records US electricity wholesale market activity in great detail – transactions and contracts on timescales ranging from 5-minutes in real-time spot markets up to multi-year PPAs. In this first pass, we applied a minimal set of transformations to the data, and after soliciting feedback from users, we are eager to take on another round of transformations and cleaning to more deeply integrate EQR into PUDL.

With funding from other orgs and avenues:

Monthly Releases of EIA-860M

We are now integrating monthly updates of EIA 860M data and publishing a monthly release of PUDL data.

PHMSA

With support from the [Future of Heat Initiative](#), we were able to extract and publish normalized (but not cleaned) PHMSA Gas Distribution data going back to 1970 as part of PUDL's annual update. Previously we were only extracting a subset of years.

SEC 10-K Ownership Data

Building off a grant from [the Mozilla Foundation](#), we invested internal funds to finalize a machine learning pipeline to extract corporate ownership data published via PDF attachments to the SEC 10-K filings and publish it in a programmatically accessible form in PUDL. We are seeking funding to do another round of improvements to this data to make it more accurate and accessible.

Influx of New Archives

In response to uncertainty surrounding the continued availability of federal energy data, we collaborated with the [Public Environmental Data Partners](#) (PEDP) with support from RMI, to respond to requests from researchers, NGOs, and government employees to create archives of vital public energy datasets. We have tripled the number of datasets we are archiving this year and helped other data archivists prioritize high-risk data critical to the energy research community. We are seeking funding to continue to support and expand this effort in 2026.

2026 PUDL Sustainers Work Plan

The majority of the PUDL Sustainer budget for 2026 remains focused on ensuring PUDL data products are published in a timely manner. A small portion of the budget will go to cover cloud computing costs. The high level breakdown of the overall PUDL budget for 2026 is as follows:

PUDL Project	2026 Budget
Data Updates	\$63,458
Infrastructure	\$56,093
Community Management	\$25,271
Project Management	\$16,310
Computation	\$8,400
Total	\$169,532

For a detailed breakdown of each task, including expected completion dates and changes from 2025, see the [Appendix](#) below.

Despite progress towards PUDL's financial sustainability, dramatic reversals in U.S. energy and climate priorities since the beginning of 2025 have had both direct and indirect financial impacts. Large reductions in NSF funding, the deprioritization and instability of clean energy & climate research, and restrictions on the diversity and inclusion activities of federal grantees have led to Catalyst's decision not to pursue an NSF POSE Phase II award or other federal grants at this time.

Moving into 2026, we have two gigawatt-level sustainers—GridLab and RMI—and a dozen smaller scale individual sustainers ranging in contribution from \$2000 to \$4. If anticipated foundation funding for PUDL is awarded, our funding gap for 2026 will be approximately **\$42,000**. To bridge the gap, Catalyst is actively recruiting additional Sustainers from the pool of universities and organizations that we know rely on PUDL data.

If you would like to support this work and become a PUDL Sustainer please [support us directly](#) or get in touch at pudl@catalyst.coop.

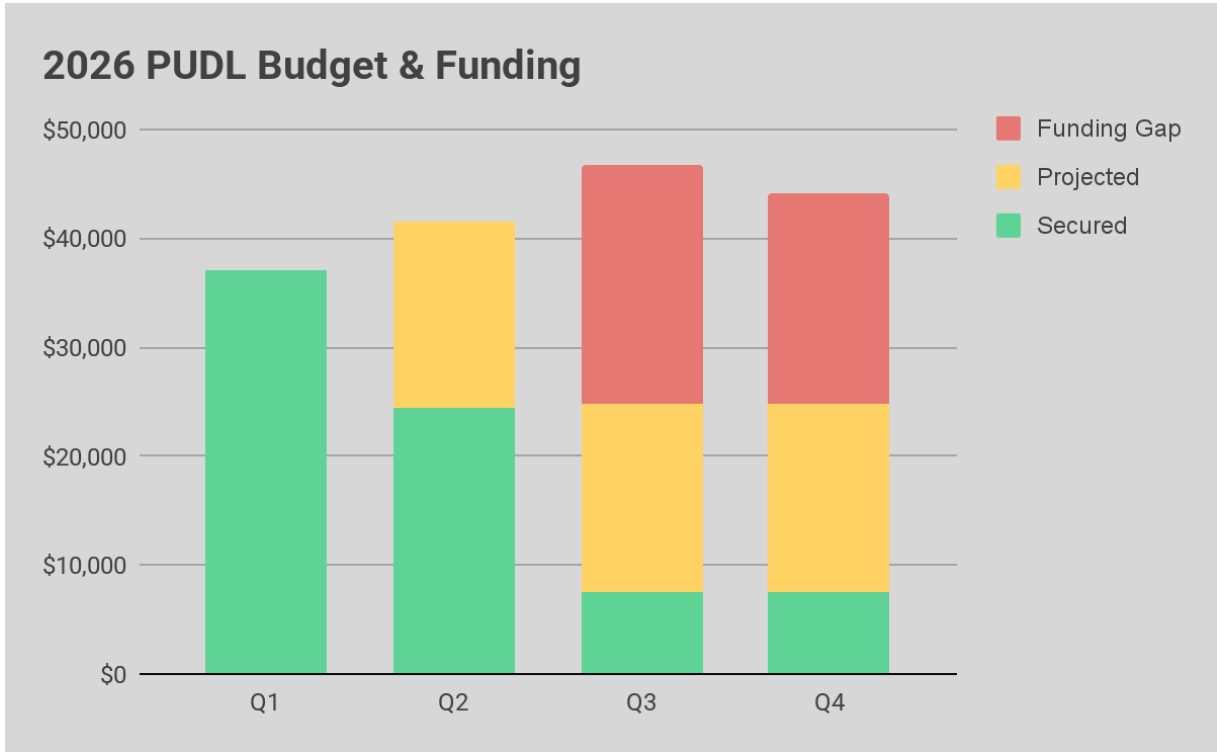


Fig. 2: 2026 PUDL Sustainer Budget broken down by quarter.

Appendix: Detailed Task Breakdown for 2026 Budget

Task Category	Task Name	Task Description / Deliverable	Date: In Nightly Builds	Date: In Versioned Release	Notes
Data Updates - Monthly	Publish Archives	Capture monthly snapshots of all our input data sources, and archive them on Zenodo. Debug when publication locations or formats change.	NA	Published to Zenodo by the 5th of each month.	2026 Update: We now run the archivers weekly due to changing data availability. We do not prioritize debugging until the first archive of the month.
Data Updates - Quarterly	EPA CEMS	Once a quarter update of new data from the EPA CEMS hourly emissions data.	Included by 15th of each quarterly update month: (Feb, May, Aug, Nov)	Included in the release on the 20th of each release month (Feb, May, Aug, Nov)	
Data Updates - Quarterly	EIA Forms 860M & 923	Quarterly update that integrates the year-to-date EIA-923 data and the monthly EIA-860M data.			
Data Updates - Quarterly	EIA Form 930	Quarterly update that integrates any newly released EIA-930 (Grid Monitor) hourly data.			
Data Updates - Quarterly	EIA Bulk API Data	Quarterly update that integrates the bulk EIA API data as a single JSON blob, for use in filling missing fuel prices.			
Data Updates - Quarterly	FERC 2, 6, & 60	Quarterly conversion from new years of the raw XBRL to SQLite, providing access to the data in a more ergonomic format, as well as the older DBF derived data, if any changes have occurred.			2026 Update: We converted this relatively small annual task into a quarterly update to increase access to these datasets.
Data Updates - Annual	Early Release: EIA Forms 860, 923 and 861	Once a year update of the annual early release data from EIA 860 and 923, including extraction and handling any data format changes. Typically early release data is available in the summer.	Two weeks after early release are available.	Included in the August 20th Quarterly Update data release.	2026 Update: We included the early release of EIA 861 after learning through usage metrics and surveys that this is a high priority dataset for many.
Data Updates - Annual	FERC 1	Once a year update of FERC Form 1 data, including converting the DBF and XBRL data to SQLite, and	August 15th		2026 Update: We reduced the expected size of this task based on two relatively

		then the PUDL extraction, transform and validation. FERC Form 1 allows rolling submissions. Most filings will be in by the Q2 release date.			smooth years of integrating FERC 1.
Data Updates - Annual	FERC 714	Integrate an annual update to the FERC-714 tables that have been integrated into PUDL, including the hourly demand by planning area, and the 10-year load forecasts.			
Data Updates - Annual	NREL ATB	Annual update to the NREL Electricity Annual Technology Baseline (ATB). Data is typically released in July.	Three weeks after initial publication.		Note: ATB was not released in 2025 and future releases are uncertain.
Data Updates - Annual	Final Release: EIA Forms 860, 923, 861	Update data for EIA forms 860, 923 and 861. Release of EIA 860 and 923 are expected in September. EIA 861 is expected in October.	Two weeks after data are available.	Included in the November 20th release.	
Data Updates - Annual	EIA-EPA Crosswalk	Once Final Release EIA 860 data has been integrated, add the new year of data to the EIA-EPA crosswalk repo.	November 15th		
Data Updates - Annual	PHMSA & EIA Natural Gas Data	Integrate extracted data for EIA 176, 191, 757a and PHMSA distribution reports. Transform existing tables: 8 PHMSA tables & 2 EIA 176.			2026 Update: Eight PHMSA Gas distribution tables and two EIA 176 tables are now core published tables.
Data Updates - Annual	EIA AEO	Annual updates for the AEO tables currently integrated into PUDL (currently 5 tables).			
Data Updates - Annual	Census PEP	Annual update of Census PEP's FIPS Codes.			New in 2026: This easy to integrate dataset will keep FIPS codes up to date.
Infrastructure	Data Distribution	Publish Nightly Builds to: Kaggle, AWS Open Data Registry in S3, and the PUDL Data Viewer. Publish Quarterly Releases to: Zenodo, AWS Open Data Registry in S3.	Ongoing	Included in the release on the 20th of each release month (Feb, May, Aug, Nov)	
Infrastructure	Bug Fixes	Triage newly reported or identified bugs and data		Ongoing	

		issues, and fix those bugs determined to be urgent.			
Infrastructure	Update Dependencies	Ensure that PUDL relies on recent versions of all our software dependencies, and is not falling behind.			
Infrastructure	PUDL Upkeep	Deal with unexpected nightly build or infrastructure issues that are not related to new features or datasets.			
Infrastructure	Example Notebooks	Ensure that example notebooks continue working with new data as it is published.			
Infrastructure	Usage Metrics	Ensure Catalyst has sufficient visibility into PUDL data usage to direct development priorities. Maintain internal user metrics data pipelines & dashboards. Review at least monthly.			New in 2026: Insight into usage patterns has proven valuable in prioritization decisions and in making the case for PUDL to funders.
Infrastructure	PUDL Data Viewer	Maintain the PUDL Data Viewer. Incremental improvements in response to user feedback.			New in 2026: The PUDL Data Viewer has seen rapid uptake since its launch mid 2025.
Community Management	Contributor Management	Integrate open source contributions - review, test and finalize external contributions.	NA	NA	New in 2026: In 2025, this work was primarily funded through NSF POSE.
Community Management	PUDL Promo	Public outreach and publicity related to the PUDL Project. Blog, newsletter, social media, etc.	NA	NA	New in 2026: In 2025, this work was primarily funded through NSF POSE.
Community Management	Recruitment	Publicize the PUDL sustaining membership program to recruit new sustaining members.	NA	NA	
Computation	Nightly Builds + GitHub CI	Annual cost of nightly build VMs, object storage, live databases, and other cloud infrastructure that's required for our CI/CD pipeline.	NA	NA	
Admin	Project Management	Coordinate efforts to complete each of these line items by their stated deadline.	NA	NA	