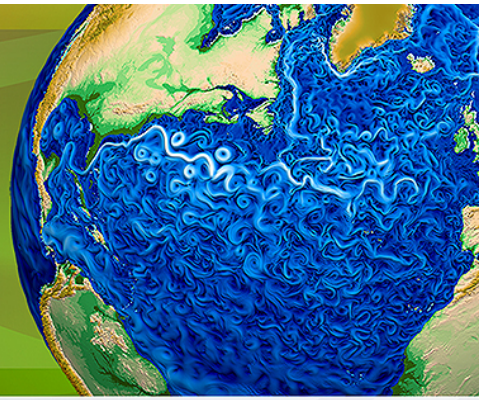
ACME
Accelerated Climate Modeling for Energy



Workflow Group

Overview and Roadmap

Framework and Tools for Supporting Model Integrations and Offline Analysis

U.S. DEPARTMENT OF ENERGY

Overview

Establish an automated infrastructure to enable archiving and comparison of simulation results, diagnostics and validation.

Framework	Description	Impact
Infrastructure	Create a flexible, extensible infrastructure for future ACME efforts and related BER projects	Heightens productivity and user experience
Data Sharing	Support broad data sharing within ACME project teams and with scientific collaborations	Accelerates model development and result dissemination
Model Set Up and Execution	Rule based support for model setup, specialized collaborative portal with a checklist for approvers of new model setups before job launching, Links to User interface and infrastructure for job submission	Enables new users to be effective quickly, allows control over model set ups in distributed, collaborative teams
Provenance Capture	Enable reproducibility and archiving of high-volume simulation data	Increases reproducibility, productivity and credibility of collaboration
Automation	Automate laborious, repetitive simulation data tasks	Increases productivity
Tools	Accelerate model development through data access, visualization, diagnostics and testing	Increases productivity, reproducibility
User Interface	Web-based job submission, running, monitoring, and debugging capabilities on several HPC centers	One stop shop to all needed capabilities, increases productivity, reproducibility

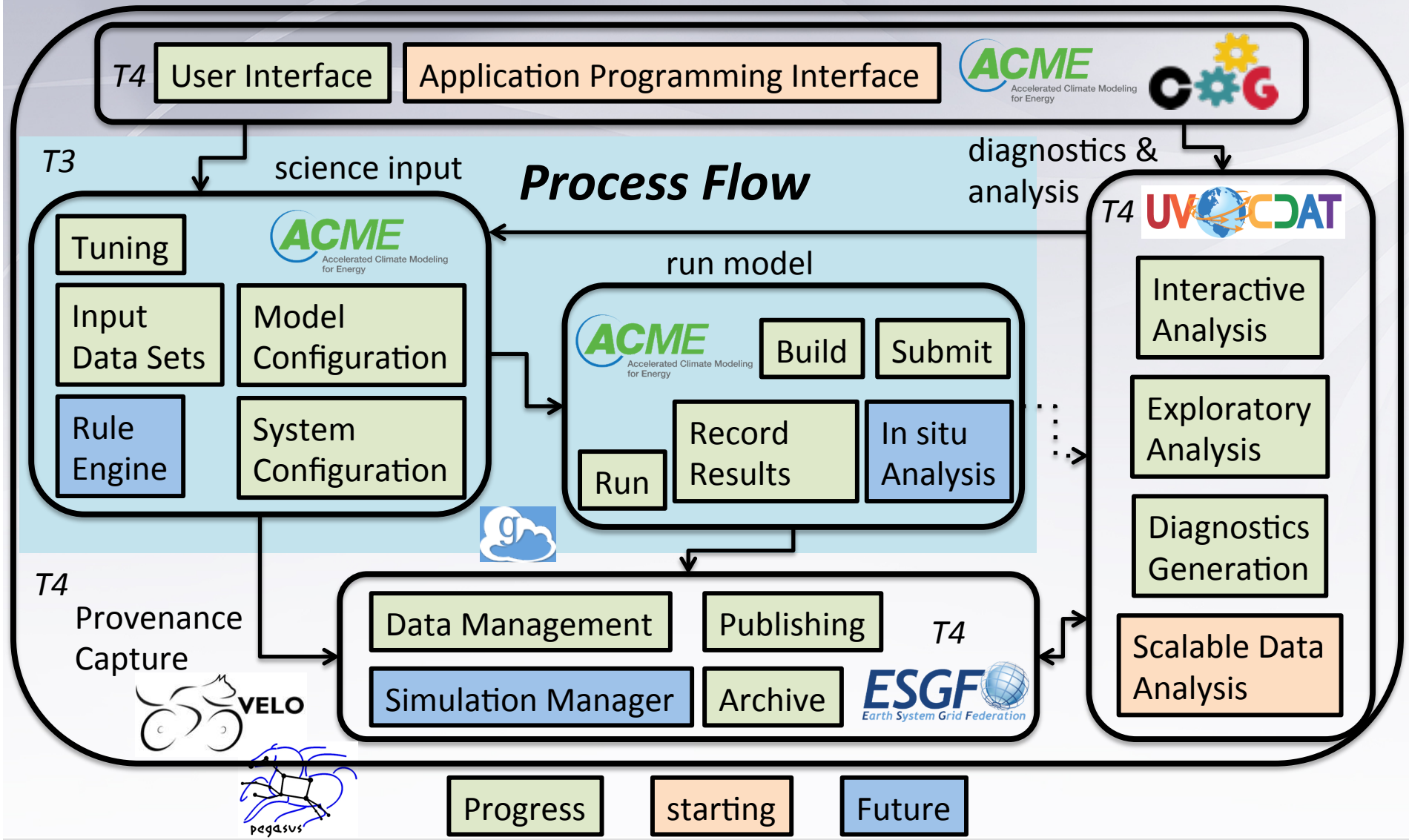
Major Challenges

Challenges	Description
Installation	Software must be adapted to multiple hardware platforms and operating systems located throughout the ACME
Heterogeneous Data Sets	The same infrastructure must allow scientists to access and compare data sets from multiple sources, including from observational satellite and instrument sources
Analysis, Diagnostics, and Visualizations	The generation of new and improved analysis, diagnostics, and visualization techniques for the better model development and evaluation
In Situ Computing	In situ computation is necessary as the increase in data size and complexity of algorithms lead to data-intensive, compute-intensive challenges for ACME diagnostics, UQ, analysis, model metrics, and visualization

Standardizing on Key Components

Component *(Capability)*	**Description**
NetCDF-CF *(Data Mgmt.)*	Conventions and metadata designed to promote the processing and sharing of files created with the NetCDF API
ESGF *(Data Mgmt.)*	Enterprising federated infrastructure to manage, disseminate, and analyze climate and weather simulations, observations, and reanalysis data on a large scale
Globus *(Data Mgmt.)*	Allows users to transfer and share files from lab server, research computing center, national supercomputing facility, or any other storage system in a fast, reliable, and secure way; and provides a service interface for publication
UV-CDAT *(Analysis/Diag.)*	A powerful and complete front-end to a rich set of visual-data exploration and analysis capabilities suited for climate research
ProvEn *(Provenance)*	Provenance capture mechanism that can handle high-velocity provenance information
Pegasus & Velo *(Provenance)*	Workflow and simulation manager that supports data management, modeling & simulation, visualization & analysis, validation, reporting, preservations, publishing & discovery

Conceptual Vision of N-to-N Workflow



Process Flow

Task Lead = Ben Mayer

- Working with Coupled Simulation Group to process ACME model runs
- Automated capture of science configuration is implemented
  - Input data sets, configuration files, system configuration, etc.
  - More fields being added as requested
- All of the below items have manual invocation (person in the loop)
  - Model run management
  - Generation of climatology and diagnostic files (AMWG UV-CDAT and NCL)
  - Use of speed transfer (Globus) to copy files to CADES
  - Publication of ACME model output (history file) to ESGF on CADES
  - Speed and lower cost storage to HPSS

Automated Process Flow

Task Lead = Kerstin Kleese-Van Dam and Ben Mayer

- Pegasus implementation of initial ACME workflow (build, configure, run, periodic analysis)
  - Running at NERSC
  - Pegasus nodes stood up at Oak Ridge, porting workflow
  - TODO: Add transfer, other analysis, ESGF publication
  - TODO: Extensive testing of near production configuration
- Scripting ESGF publication process started
  - Directory based facet transformation
    - Works for one set of output configuration
    - Needs more work to other output styles
  - ESGF publication – functional features to add (other than NetCDF)
- Integration with end-to-end user interface started

Automated Process Flow

- HPSS wrapper script
  - Faster storage
  - More cost optimal than ‘hsi put’
  - Has CRC to check for data correctness
  - More features to add, but functional now
- Process Flow and Data Management go hand in hand. The automated processing of model runs and analysis will be combined with automated data management. This will allow for the reliable and consistent storage of generated data to the ESGF and tape archive systems.

Data Management

Task Leads = Sasha Ames and Rachana Ananthakrishnan

- ESGF v1.8 released (http://esgf.llnl.gov)
  - New content management system user interface (CoG)
  - Publisher upgrades for ACME
- Deployed ESGF at ORNL CADES, ANL, LANL, LLNL, and PNNL
  - LBNL/NERSC coming soon
- Transfer model run data from ANL and ORNL to CADES using Globus
- Published ACME ne30.v0_1 mostly atmosphere data, some ice and ocean data sets to ORNL CADES ESGF site
  - 20 ne30 data sets: 6 climatology, 4 h, 6 h0, 1 h1, 1 h2, 1 h3, 1 h4
  - Model output history files
  - Climatology files
- Publication as a Service prototype
  - Simple interface for scientist to publish data sets

Diagnostics & Metrics

Task Leads = Jeff Painter and Brian Smith

- UV-CDAT diagnostics framework tailored for ACME science domains: Atmosphere, Land, Ocean
  - Atmosphere Diagnostics:
    - Tier 1a: represents the top most wanted 10 diagnostics
    - Tier 1b: a collections of fields; relevant regions or phenomena
    - Tier 2: completed substantially equivalent AMWG diagnostics
  - Land Diagnostics:
    - Tier 2: nearly completed substantially equivalent LMWG diagnostics
    - Plot set 7 in progress
  - Ocean Diagnostics:
    - Resides within ocean model

Diagnostics & Metrics

- Command-line utilities, plus integration in UV-CDAT GUI
  - UV-CDAT-based Python scripts to produce climatology files
    - on unstructured native grid or on a structured grid
    - tested on data from ne30, ne120 grids
  - UV-CDAT-based Python scripts to produce static diagnostics plots
    - diags.py generates single set of static plot files, each with a NetCDF file with data to recreate the plot
    - metadiags.py generates all plots as specified by a Python dictionary with single command line invocation
    - Single plots can also be created “live” in UV-CDAT GUI
  - Diagnostics can be viewed via UV-CDAT scripts or GUI, or via the Classic Viewer web interface

UV-CDAT

https://uvcdat.llnl.gov

Task Leads = Charles Doutriaux and Aashish Chaudhary

- UV-CDAT v2.0, v2.1, and v2.2 released
- Git & Github (repository)
- Ported to CADES & Rhea (OLCF Viz cluster)
- Merge of ACME diagnostics framework (UVCMetrics) into UV-CDAT
  - Climatology script generation
  - Combine NetCDF file with specified grid script
- Update of Kitware's CMake, CTest, and CDash support tools
  - ~400 tests in the test suite
  - Continuous nightly builds and testing

UV-CDAT

https://uvcdat.llnl.gov

- VTK backend graphics
- New visualization capabilities
- 2D and 3D animations
- Documentation and tutorials
  - Plot gallery
  - Askbot support helpdesk
- 2D and 3D widget control on the canvas
- Diagnostics GUI integration
- CDATWeb – user interface to allow remote visualization and computing

User Interface

Task Lead = Matthew Harris

- Completed end-to-end user interface framework
  - User can login and set their user preferences (i.e. panel placement, color, etc.)
  - Displays the ESGF nodes connected to the federation
- Deployed to production server at ORNL and LLNL
  - Include the visualization server
- Integrated with Velo simulation manager API
- Integrated with CDATWeb and ESGF for login and search
- Can generate 2D plots from any ESGF data set located on CADES
- Working on choosing the plot set and diagnostics from CADES (i.e., history and climatology files)

12-Month Roadmap

https://acme-climate.atlassian.net/wiki/display/WORKFLOW/12-Month+Roadmap+-+Workflow+Group

- Process Flow
  - Coordinate with the Coupled Simulation Group and science teams to provide the documentation of the "process flow" for running ACME
  - Document/identify inefficiencies in the "process flow" and make recommendations on how to make improvements; thereby reducing the amount of time spent on menial tasks
- Data Management
  - Upgrade to ESGF v2.0 with new content management system for all ACME sites
  - Provide Publication as a Service to scientists for remove publication of model runs from other sites
  - Provide publisher with Velo simulation manager to aid in the end-to-end automation process
- Analysis & Diagnostics
  - Upgrade to UV-CDAT v3.0 with provenance scripting and remote computing
  - Integrate ocean diagnostics into UVCMetrics
- Provenance Capture
  - Provide format to capture sufficient information for reproducibility and link to specific model runs
  - Integrate into end-to-end system and install at ORNL
- End-to-end User Interface
  - Deliver basic capability end-to-end user interface for running and analyzing ACME model runs

Workflow Hack-a-Thon and Hands-On-Tutorial

- Workflow Hack-a-Thon (Wednesday from 3:30 pm to 5:30 pm):
  - Hack-a-Thon: discuss and work with the Workflow Group on tools and framework needed for ACME
  - Preloading of VMs for (Thursday's hands on tutorials)

- Hands On Tutorials (Thursday from 8:20 am to 9:50 am)
  - Process flow demonstration
  - Hands on Diagnostics tutorial
  - Hands on UV-CDAT tutorial
  - Publisher as a service demonstration

Accelerated Climate Modeling for Energy