Chiltepin: A Federated Workflow System for Earth Science

Christopher W. Harrop
Isidora Jankov

30th Annual CESM Workshop, June 10th, 2025

Acknowledgements

This work is funded by the NOAA Software Engineering for Novel Architectures (SENA) Program

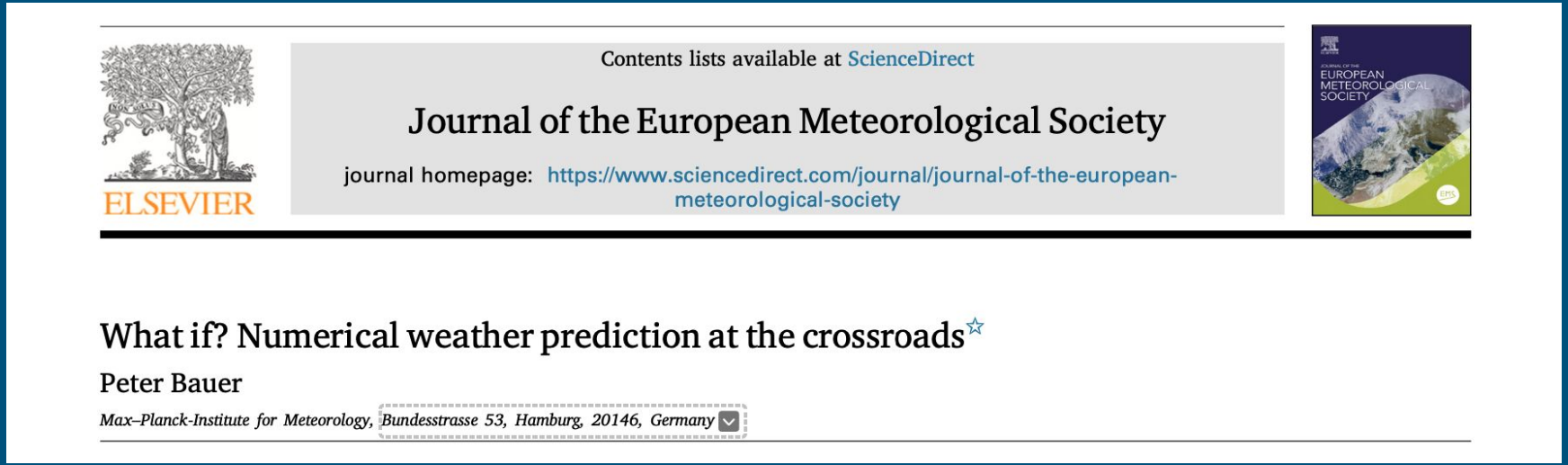
In collaboration with: U.S. Department of Energy, The University of Chicago, and Parallel Works

Special thanks to: Yadu Babuji, Josh Bryan, Kyle Chard, Ben Clifford, Stefan Gary, Kevin Hunter Kesling, James Kube, Reid Mello

Contents

- Introduction
- Chiltepin Preview
- Challenges & Future Work

An Evolving HPC→NWP Landscape



ELSEVIER

Contents lists available at ScienceDirect

Journal of the European Meteorological Society

journal homepage: https://www.sciencedirect.com/journal/journal-of-the-european-meteorological-society

What if? Numerical weather prediction at the crossroads☆

Peter Bauer

Max–Planck-Institute for Meteorology, Bundesstrasse 53, Hamburg, 20146, Germany

Peter Bauer. 2024. What if? Numerical weather prediction at the crossroads. *Journal of the European Meteorological Society* 1 (2024), 100002. https://doi.org/10.1016/j.jemets.2024.100002

Some Existing NWP Workflow Products

- Rocoto (NOAA / GSL)
- ecFlow (ECMWF)
- Cylc (NIWA)
- CIME (NCAR)
- EWOK (JEDI)
- Autosubmit (Barcelona Supercomputing Center)
- uwTools (NOAA / GSL, EPIC)

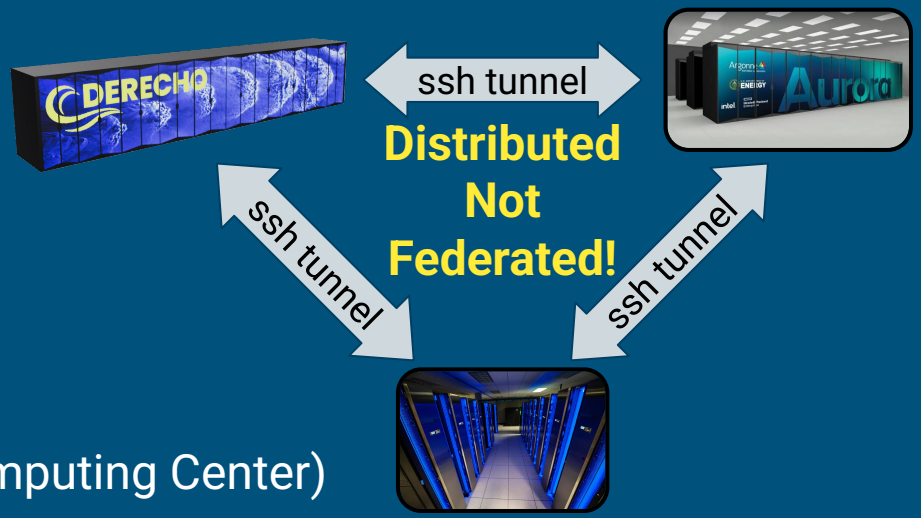
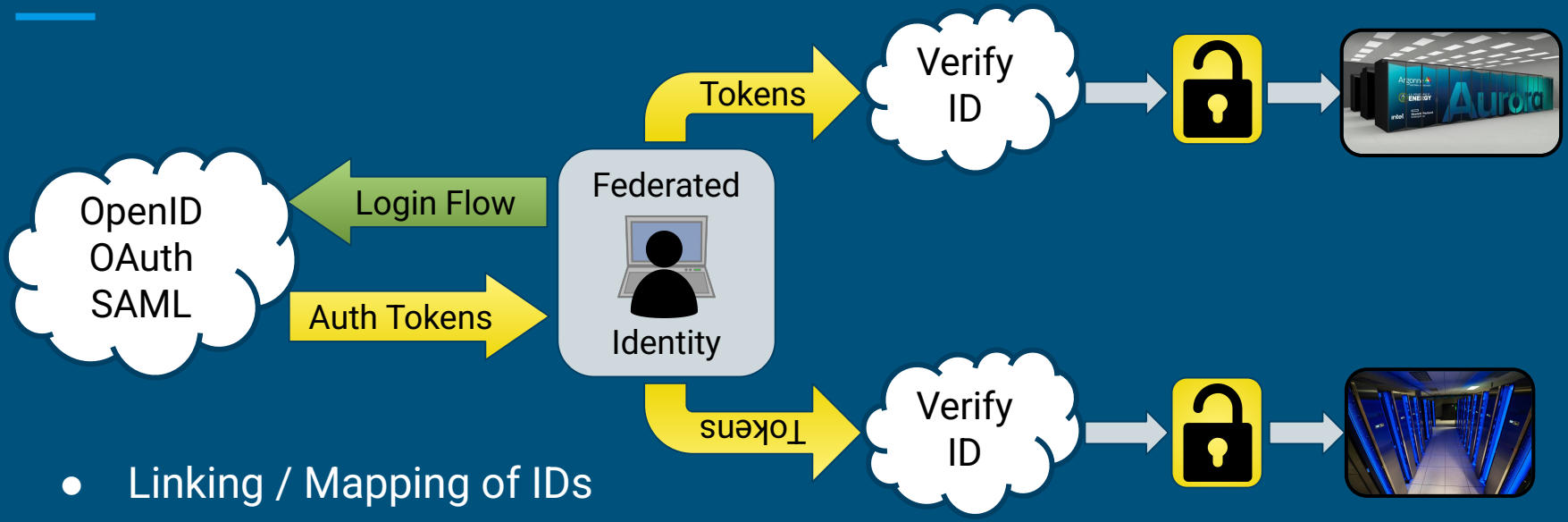


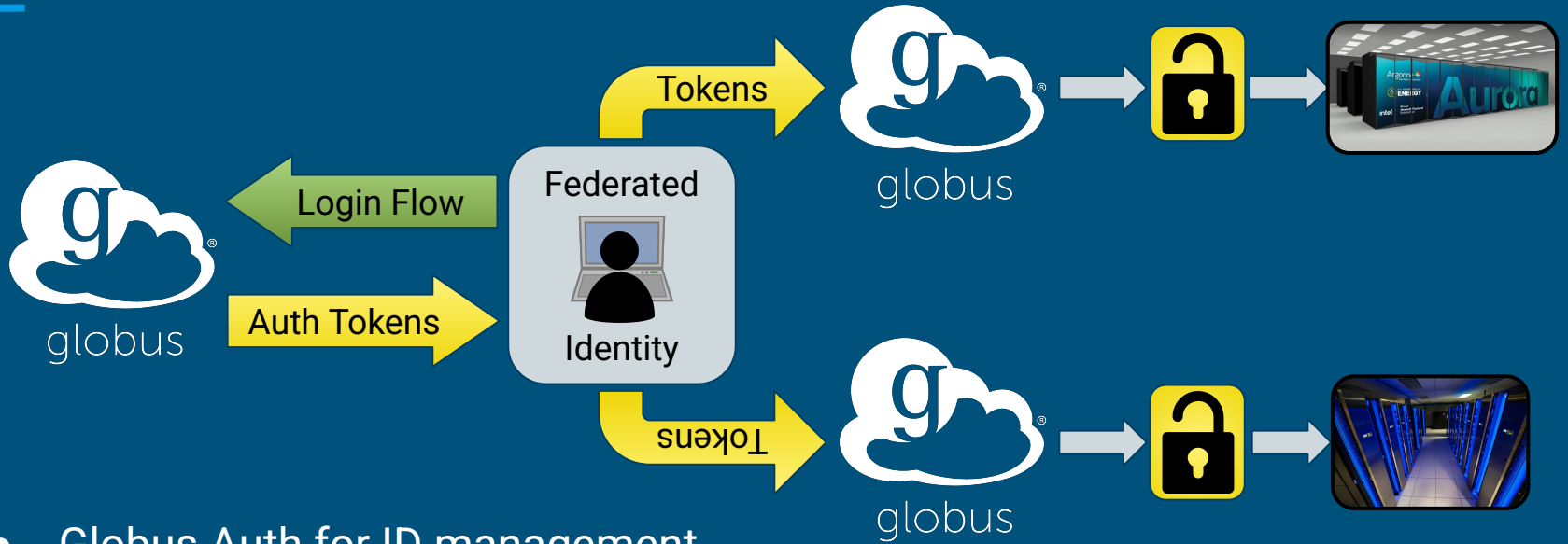
Image credits:
https://ncar.ucar.edu/sites/default/files/inline-images/derecho-ncar-render_0.png
https://www.anl.gov/sites/www/files/2021-03/Aurora_environment_1600x900.jpg
https://www.noaa.gov/media/image_download/309f7ef2-5bf4-4b91-8c00-fcf40dfaea29

What is a Federated Workflow?



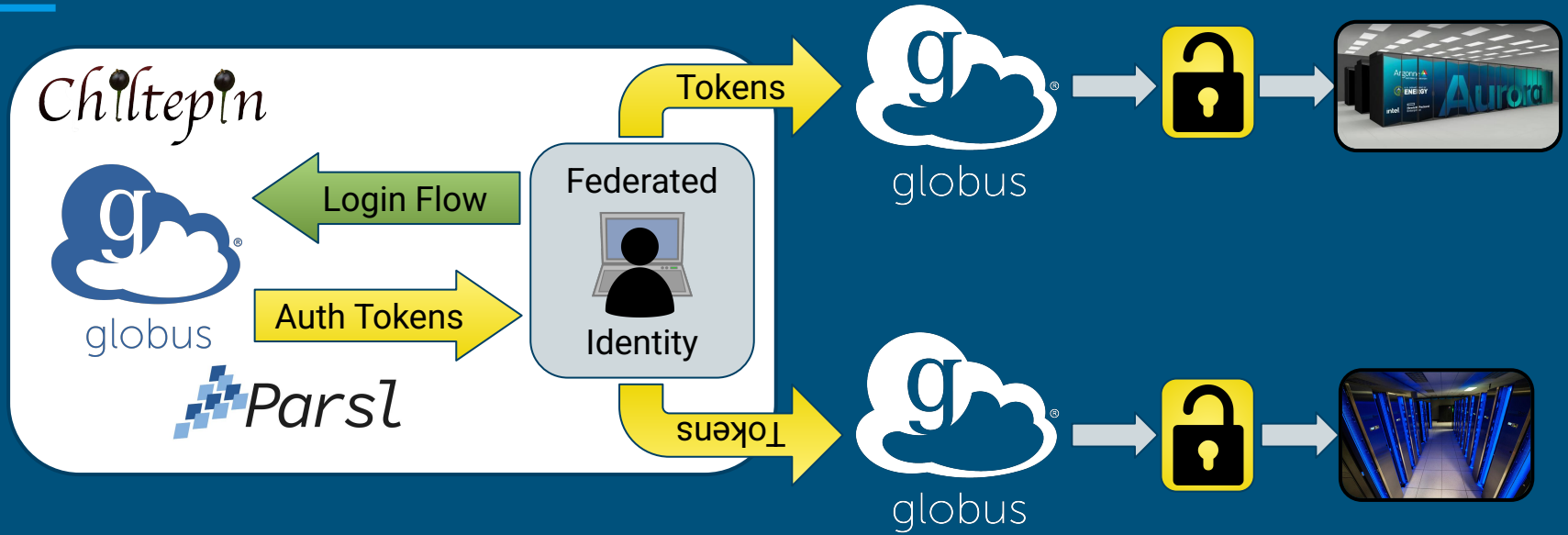
- Linking / Mapping of IDs
- Access control & limited scopes
- Reusable credentials until expiration
- Secure access across autonomous resources

Chiltepin Leverages Existing Solutions



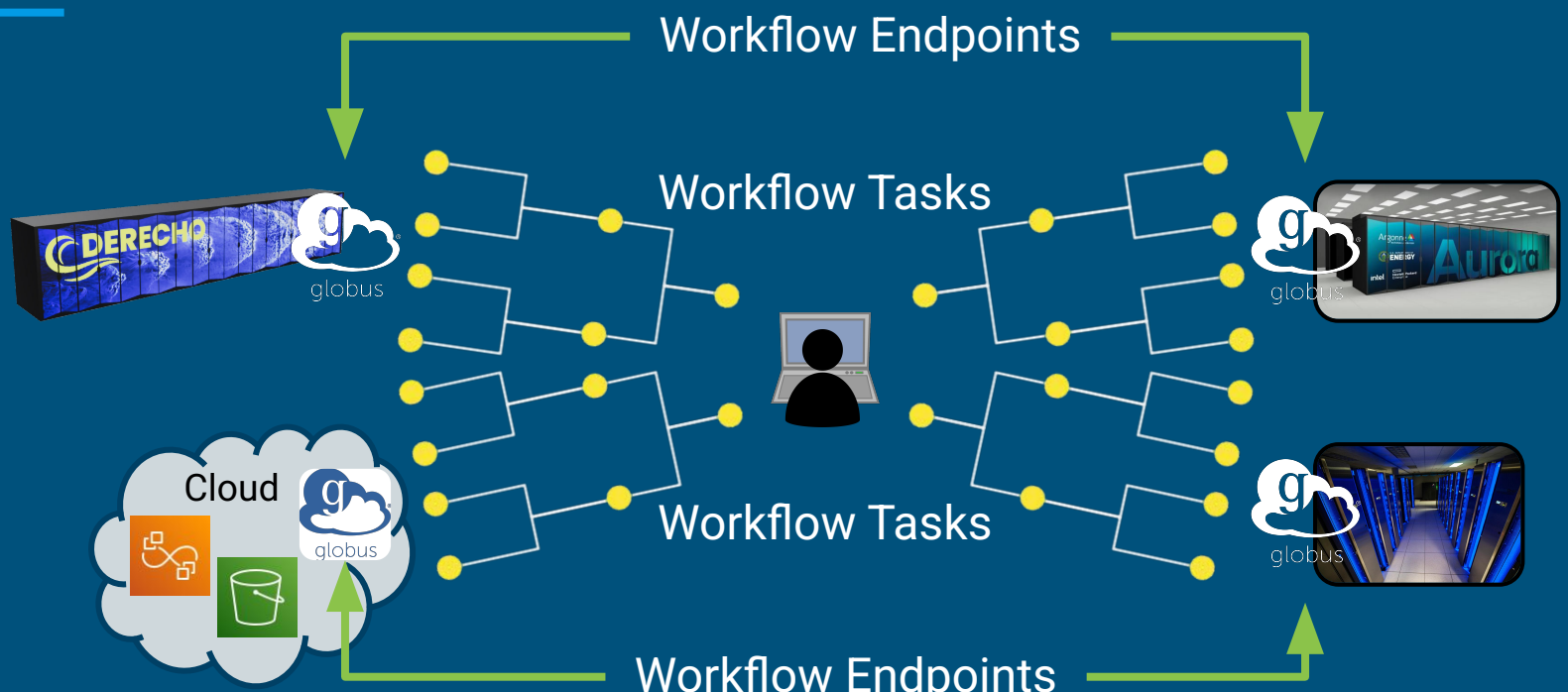
- Globus Auth for ID management
- Globus Compute FaaS for execution at workflow endpoints
- Globus Transfer for data movement

Chiltepin Leverages Existing Solutions



- Parsl for workflow automation
- Parsl's GlobusComputeExecutor for remote execution
- Globus SDK for login & transfers

Chiltepin Vision



Chiltepin Compute Endpoint Lifecycle Mgmt

```
$ chiltepin endpoint configure myendpoint

$ chiltepin endpoint list
myendpoint None Initialized

$ chiltepin endpoint start myendpoint

$ chiltepin endpoint list
myendpoint 46060b64-8383-4577-a74e-a45f928592dc Running
```

Chiltepin Tasks Are Python Functions / Futures

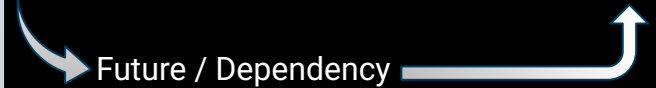
```
class MPAS:

    @bash_task
    def mpas_init(self, …):
        …

    @python_task
    def mpas_forecast(self, ….):
        …
```

```
import chiltepin.configure
import chiltepin.endpoint as endpoint

def main(...):
    endpoint.login()
    resources = chiltepin.configure.load(“config.yaml”)
    with chiltepin.init(resources):
        mpas = MPAS()
        init = mpas.mpas_init(..., executor=”aurora”)
        fcst = mpas.mpas_forecast(..., init, executor=”derecho”)
```



Chiltepin Configuration

```
resources:
  hercules:
    environment:
      - "module load ..."
      - "export MY_VARIABLE=my_value”
    endpoint id: "1234abcd-5678-90123-ab45-abcdef67891234”
    provider: "slurm"
    mpi_launcher: "srun"
    cores per node: 80
    nodes per block: 3
    partition: "hercules"
    account: "foobar"
```

```
endpoint.login()
resources = chiltepin.configure.load(“config.yaml”)
with chiltepin.init(resources):
   mpas = MPAS()
   init = mpas.mpas_init(..., executor=”hercules”)
   fcst = mpas.mpas_forecast(..., init, executor=”derecho”)
```

Challenges & Future Work

- Workflows as Python programs is a new paradigm (for us)
- Need to develop good monitoring tools
- Need to take better advantage of built-in failure recovery
- Develop interoperability with legacy Rocoto workflows
- Continue development to build a sustainable workflow future!
- Publication in preparation

Questions / Discussion