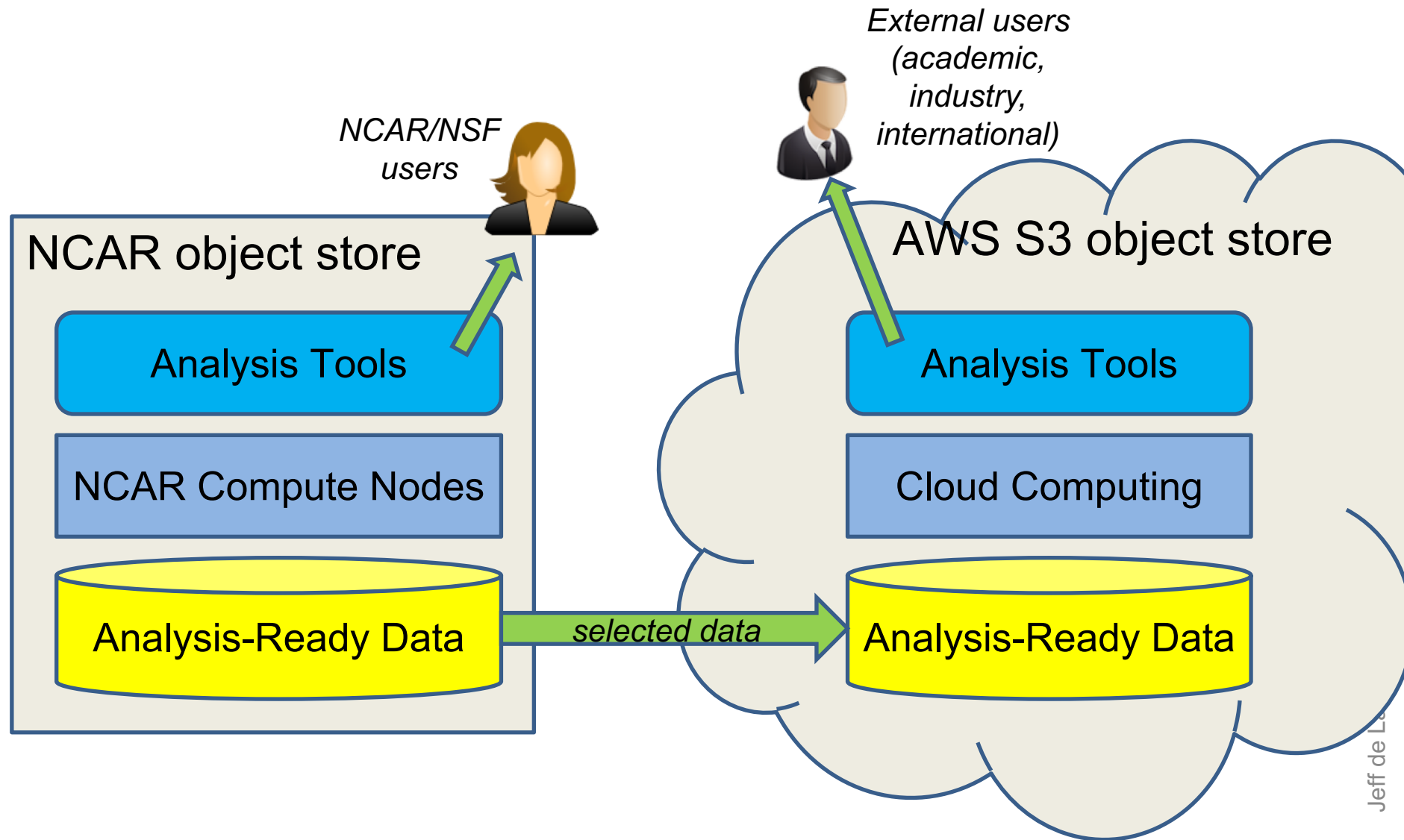


NCAR Datasets in the Cloud

AGU Fall Meeting - IN44A-03
2021-12-16

**J-F de La Beaujardière, Brian Bonnlander, Seth McGinnis,
Max Grover, Anderson Banihirwe, Kevin Raeder, Gary Strand**
National Center for Atmospheric Research

NCAR "Science at Scale" Project: High-Level Concept



Jeff de L

What We Did

- **Converted to Zarr** major subsets of 4 large datasets
 - Binary format optimized for fast parallel reads using Python Xarray
 - Multiple netCDF files combined into single Zarr stores
 - Data are sliced into separate small (100 MB) objects along multiple dimensions: space, time, ensemble member
- **Published data on AWS S3** (us-west-2)
 - Free storage & egress thanks to AWS Open Data Sponsorship Program and Amazon Sustainability Data Initiative
 - Data also on NCAR on-prem Object Store
- **Created Intake-ESM Catalogs** as machine-readable inventory for each dataset
 - Simplifies ingest by Python xarray
 - Mapped unique variable names to CF Standard Names
 - Added info about spatial & temporal coverage
- **Wrote Jupyter Notebooks** showing sample analyses

Published Datasets

- **CESM LENS** - Community Earth System Model Large Ensemble
 - <https://doi.org/10.26024/wt24-5j82>
- **CESM2 LE** - CESM version 2 Large Ensemble
 - <https://doi.org/10.26024/y48t-q717>
- **NA-CORDEX** - N. Am. Coordinated Regional Downscaling Experiment
 - <https://doi.org/10.26024/9xkm-fp8>
- **DART CAM6 Reanalysis** - Data Assimilation Research Testbed
Community Atmosphere Model v6 (CAM6) Reanalysis
 - <https://doi.org/10.26024/sprq-2d04>

Data Characteristics

Dataset	Size	Years	Coverage	Ensemble
CESM LE	83 TB	1920-2100	Global 1°	40 members
CESM2 LE	267 TB	1850-2100	Global 1°	100 members
NA-CORDEX	13 TB	1950-2100	N. Am. ¼ or ½°	50 experiments
DART CAM6	2.5 TB	2011-2019	Global 1°	80 members

AWS Registry of Open Data listing:

<https://registry.opendata.aws/?search=managedBy:national%20center%20for%20atmospheric%20research>

Future Plans

- Add additional variables on request
 - Notably: Plant Functional Type data from DART CAM6
- Notebooks for dataset inter-comparison
- Simple on-demand visualizations using serverless computing

For more info, questions, usage problems,
requests for additional data:
contact cisl-aws-lens@ucar.edu

J-F de La Beaujardière, PhD

Director, NCAR/CISL Information Systems Division

jeffdlb@ucar.edu

<https://orcid.org/0000-0002-1001-9210>

Jeff de La Beaujardiere <jeffdlb@ucar.edu>

