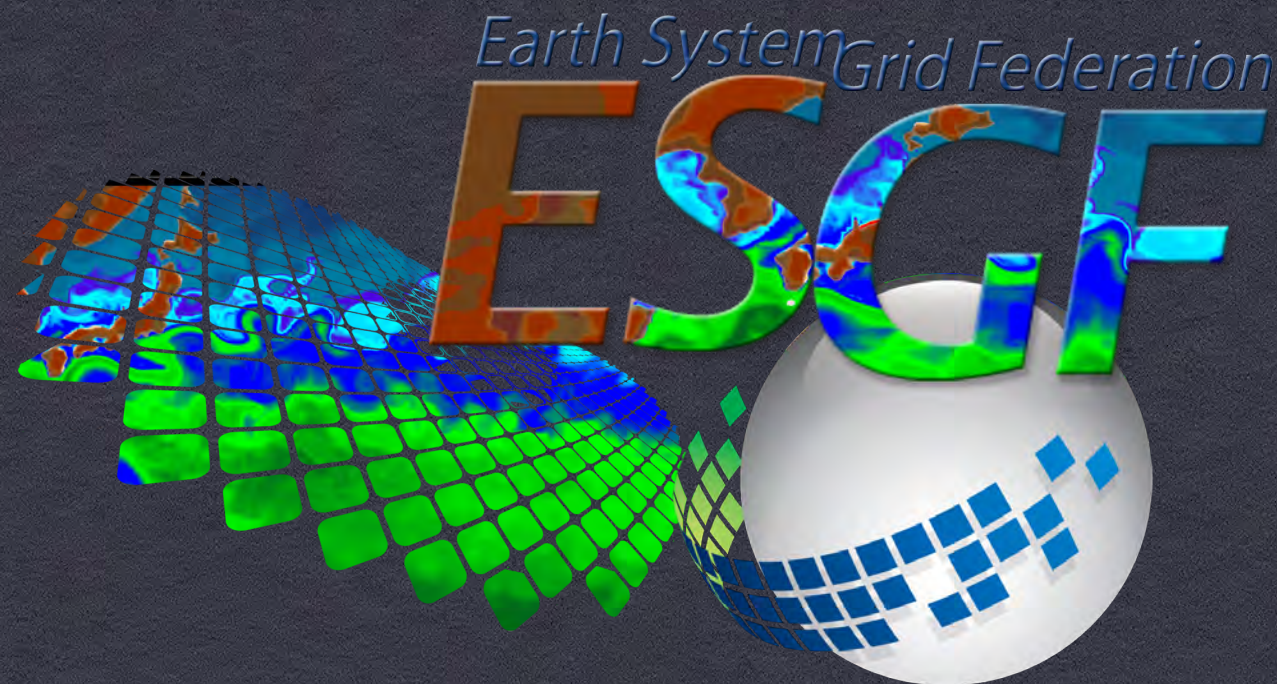




Jet Propulsion Laboratory
California Institute of Technology



THE MAGIC OF DREAMS

(... OR “ON THE USE OF CONTAINERIZATION TECHNOLOGIES IN THE DISTRIBUTED RESOURCES FOR THE ESGF ADVANCED MANAGEMENT (“DREAM”) PROJECT, DISCUSSED DURING THE MIDDLEWARE AND GRID INTERAGENCY COORDINATION (“MAGIC”) TEAM MONTHLY MEETING)

LUCA CINQUINI

NASA JET PROPULSION LABORATORY AND CALIFORNIA INSTITUTE OF TECHNOLOGY

ON BEHALF OF THE ESGF AND DREAM COLLABORATION

JPL UNLIMITED RELEASE SYSTEM CLEARANCE NUMBER: #18-1095

© 2018 CALIFORNIA INSTITUTE OF TECHNOLOGY. GOVERNMENT SPONSORSHIP ACKNOWLEDGED

ESGF Overview

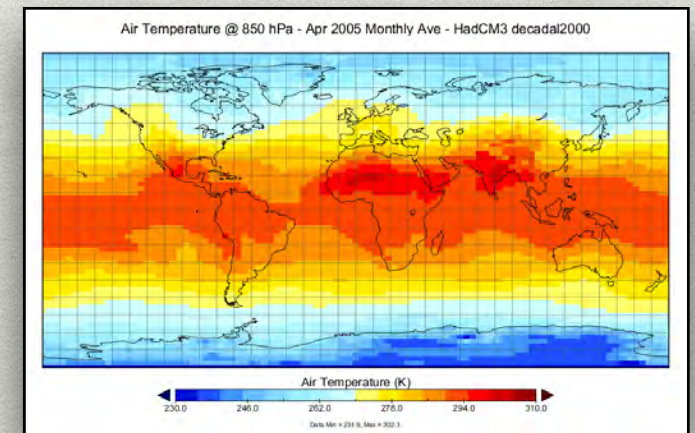
- * ESGF is an international collaboration of climate centers working together to manage and provide access to climate data - models and observations
- * Started more than a decade ago, now the world premier technology infrastructure in support of climate science
- * Spanning several tens of institutions in Europe, North America, Australia and Asia
- * Funding from DOE, NASA (U.S.), Copernicus (EU), NCI (Australia), CRIM (Canada)
- * Winner of the 2017 “R&D 100 Award” - prestigious conference that every year recognizes the top 100 most innovative products in software, science and technology



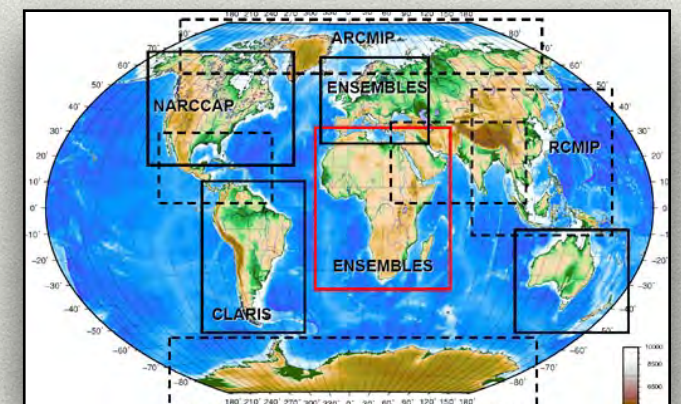
Current and Future Data Holdings

- * ESGF hosts some of the most prominent data collections for climate change research:
 - * **CMIP3, CMIP5** (“Coupled Model Inter-Comparison Project”): output of global climate models used for periodic IPCC assessment reports on climate change
 - * **CORDEX** (“COordinated Regional climate Downscaling EXperiment”): output of regional climate models, grouped by domain (N. America, Europe, Antarctica, etc.)
 - * **Obs4MIPs** (“Observations for Model Inter-Comparison”): observational data from NASA, ESA, etc. formatted to look like climate model output
 - * **Ana4MIPS** (“ReAnalysis for Model Inter-Comparison”): re-analysis data formatted like model output
 - * Many other MIPs: **TAMIP, GeoMIP, DCMIP, ...**
 - * Upcoming **CMIP6**: 25-40 PB of uncompressed data replicated at 4 “Tier-1” Nodes

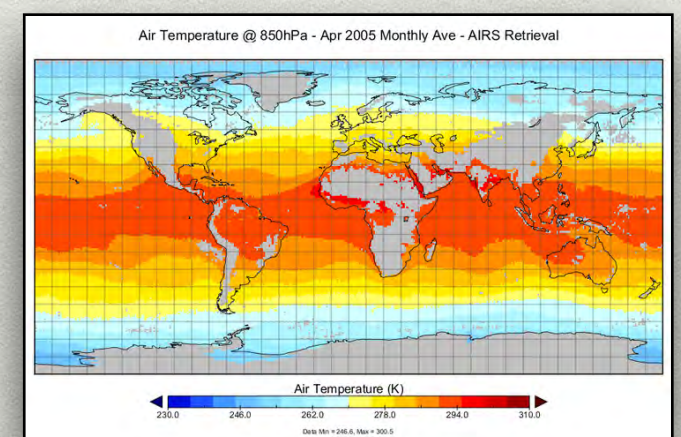
HadCM3



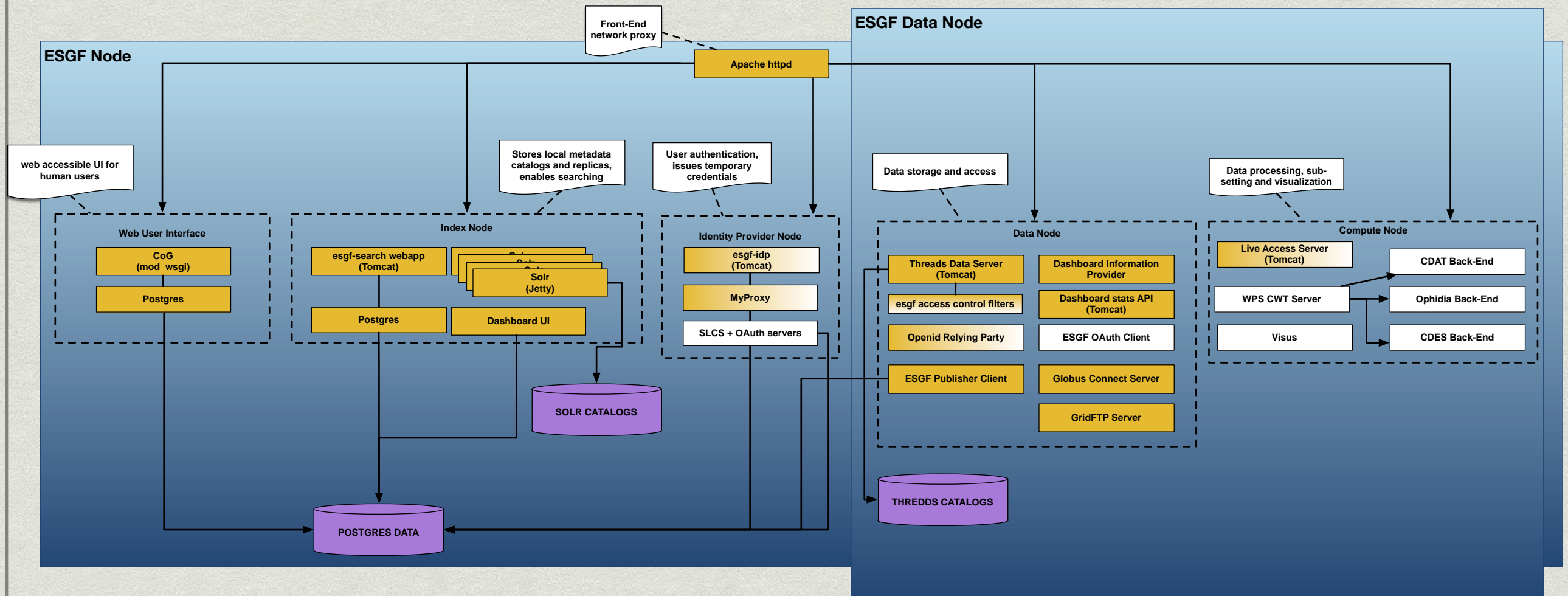
CORDEX



Obs4MIPs



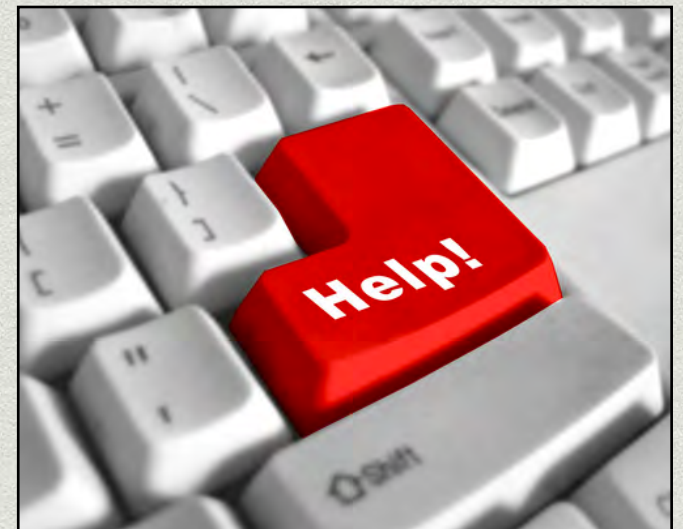
Software Stack



- * Integration of several components from the open source community, climate applications, and developed by ESGF
- * Services are grouped into 5 categories according to functionality: User Interface, Index Node, Identity Provider, Data Node, Compute Node
- * Overall, a very large software stack that is deployed at each Node in the federation

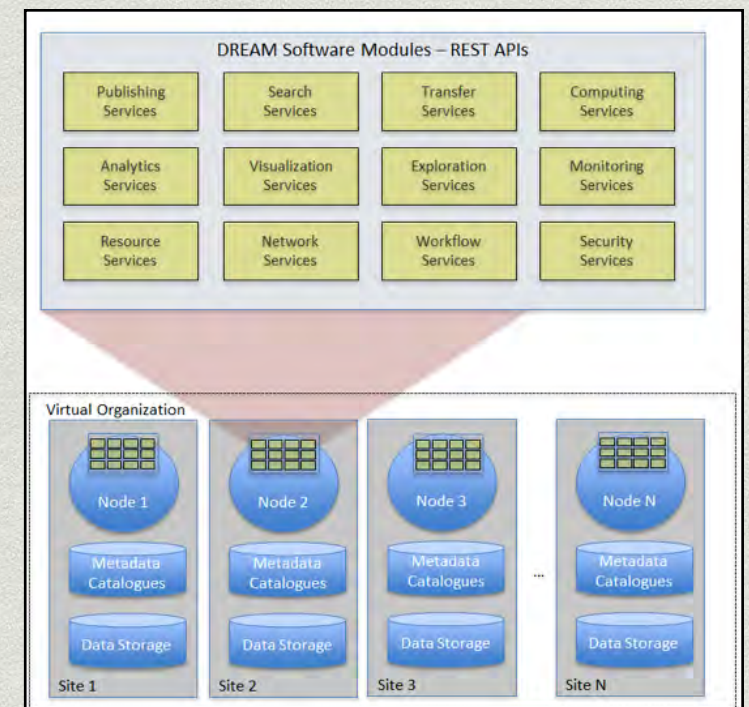
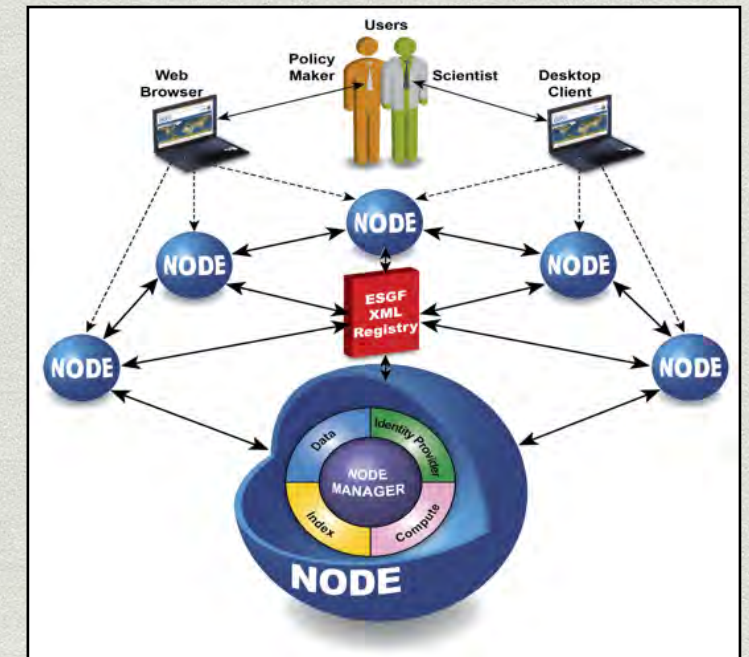
Some Problems with current ESGF System

- * Difficult to install and upgrade, impossible to roll back
- * Installation driven by a gigantic, unmaintainable shell script
- * Architecture is not very flexible or easy to scale
- * Because of the installation barrier, it's difficult to add new software modules and functionality



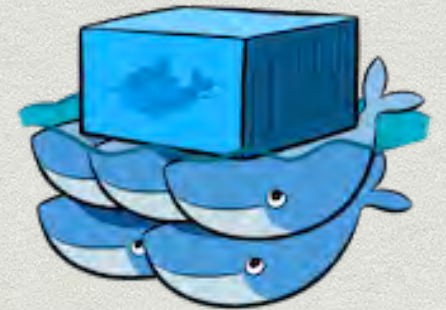
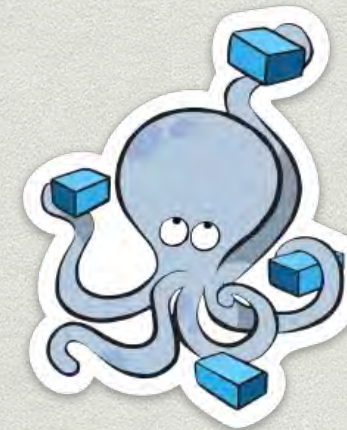
DREAM: Distributed Resources for the ESGF Advanced Management

- ✱ Goal: design and implement the next generation ESGF architecture: more modular, scalable, easier to install and upgrade, applicable to other science domains
- ✱ Strategy: modularization and interaction via RESTful APIs
- ✱ Implementation: use containerization technologies to design a new ESGF architecture based on micro-services



Containerization Technologies

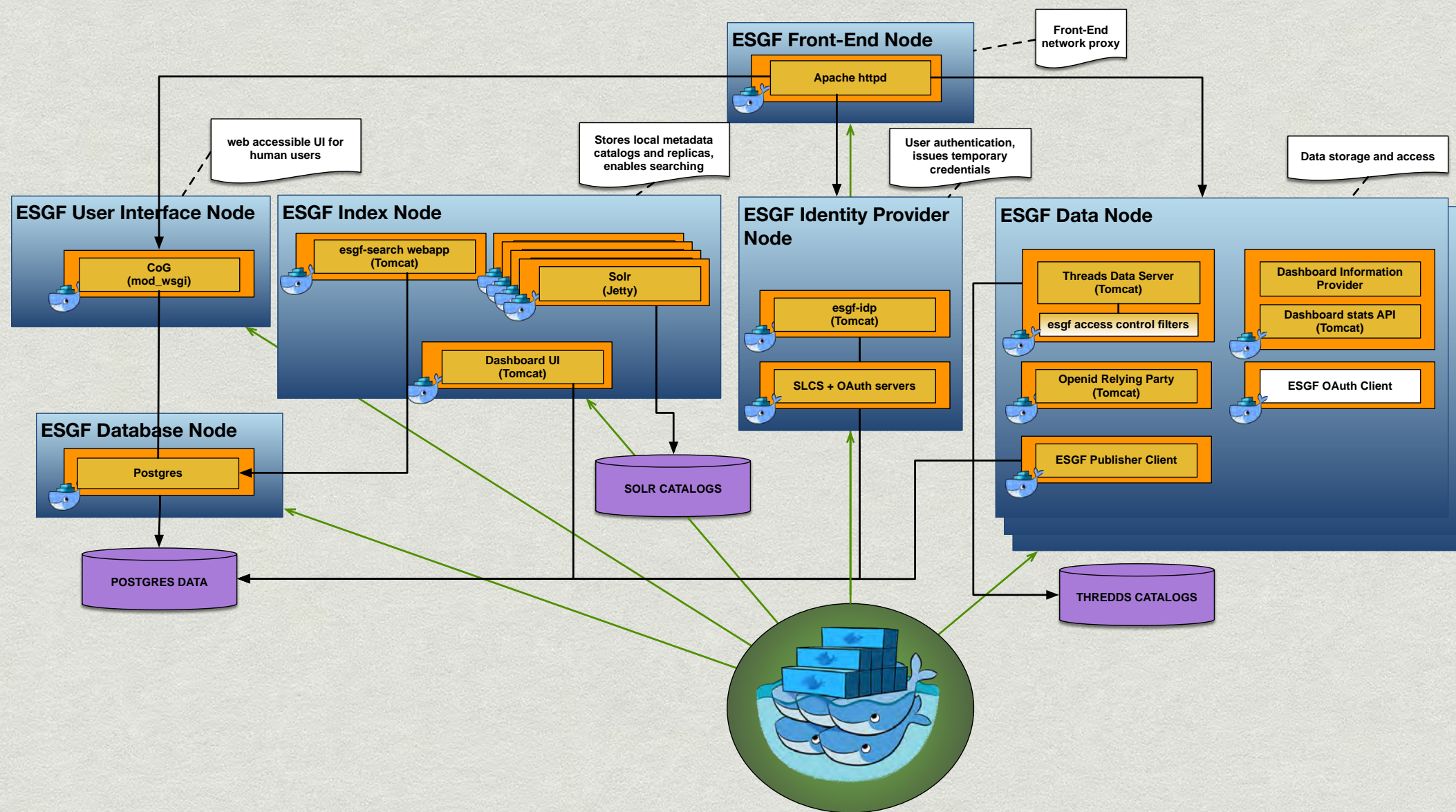
- * Docker to containerize each ESGF service
- * Docker Compose to prototype service interaction on single host
- * Docker Swarm and Docker Stack to deploy all ESGF services on multiple hosts
- * Kubernetes, as alternative orchestration engine on multiple hosts
- * Helm for configuring, packaging and deploying Kubernetes resources
- * OpenShift Origin, as prototype Platform-As-A-Service for managing the K8s cluster



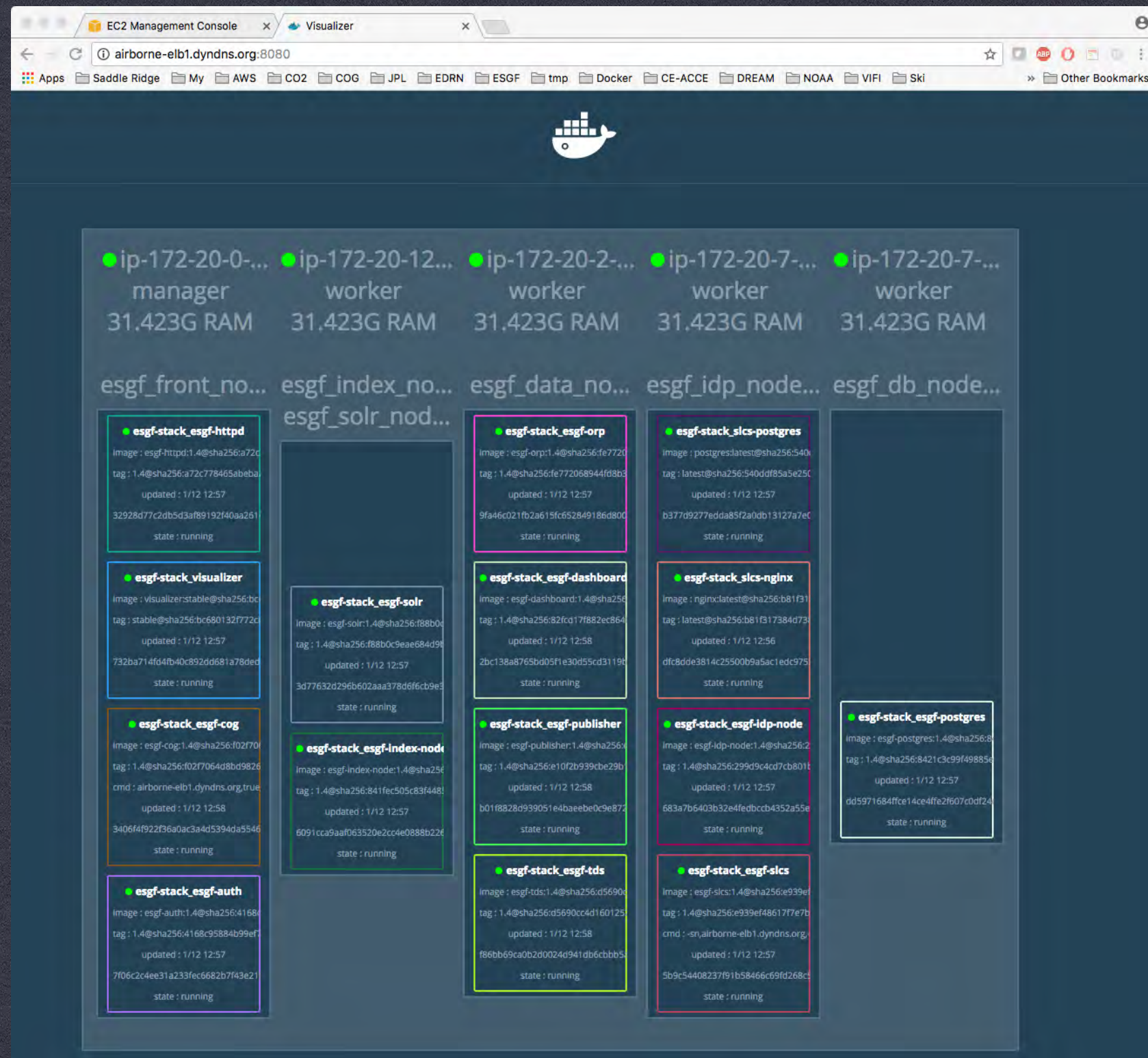
Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

ESGF/Docker Architecture w/ Swarm

- * All ESGF services packaged as Docker containers
- * Services deployed on hosts of specific type using metadata labels



Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.



ESGF/DOCKER/SWARM TESTBED ON AMAZON WEB SERVICES

USING AN ECS (ELASTIC CONTAINER SERVICE) CLUSTER OF 5 NODES

Advantages of Micro-Services Architecture

- * Easier installation: simply download and deploy the Docker images through Docker Stack or Kubernetes (no more complex compilation and installation at each site; also, everybody runs exactly the same binary)
- * Easier upgrade: download a new version of the images, start the containers
- * Rollback an upgrade, if needed
- * Scalable, as multiple containers on the same host, or on multiple hosts
- * More flexible architecture options: Complete Node, Data Node, Index Node, ...
- * Better testing
- * Deployable anywhere

Lessons Learned

- * In most cases, legacy applications are not built to be deployed in a dynamic, distributed environment
 - * Assume application data are stored in a persistent location (which is *not* the default case with containers or pods, which can be moved from host to host) - must configure Docker or Kubernetes “volumes” to persist data
 - * Assume static IP addresses/hostnames (while IP addresses and hostnames are assigned dynamically in a containerized architecture) - must configure Docker or Kubernetes “services”
 - * May use “customized” versions of open source software packages (Postgres, Tomcat, Solr, ...)
- * Consequently, legacy applications need to be partially re-architected or re-configured to be deployed and scaled as containers
- * Steep learning curve for application developers and system administrators
 - * It's relatively easy to write a Docker file and create a Docker image
 - * It's more difficult to run a set of Docker containers with Docker Compose or Docker Swarm, must understand the semantics and capabilities of the framework (especially, networking and persistent storage)
 - * It's even more difficult to run them in a Kubernetes cluster



Best Practices

- * Run a single process per container, whenever possible
- * Use multi-stage builds to create Docker images with reduced size
- * Use official Docker images when available (smaller and more secure)
 - * Tomcat, Solr, http, etc...
- * Persist application data with Docker Volumes or Kubernetes Persistent Volumes
- * Use Docker Services, Kubernetes Services or OpenShift Routes to address the applications with persistent endpoints
- * Run also administration tasks/commands inside containers
 - * No “residual” OS dependencies
- * Express dependencies between containers startup
 - * Use Docker-Compose “dependsOn” semantics
 - * Or Kubernetes “initContainers” to delay the start of dependent containers



Conclusion

- ✱ Despite these difficulties, ESGF and DREAM are fully committed to the micro-services architecture model
- ✱ Goals:
 - ✱ Release a feature-complete ESGF/Docker architecture by the end of 2018
 - ✱ Run at least some operational sites with the ESGF/Docker architecture, either in-premise or on the Cloud



"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development
(NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674,
Fax: 202-459-9673, Email: nco@nitrd.gov, Website: <https://www.nitrd.gov>

