LADCO | LAKE MICHIGAN AIR DIRECTORS CONSORTIUM

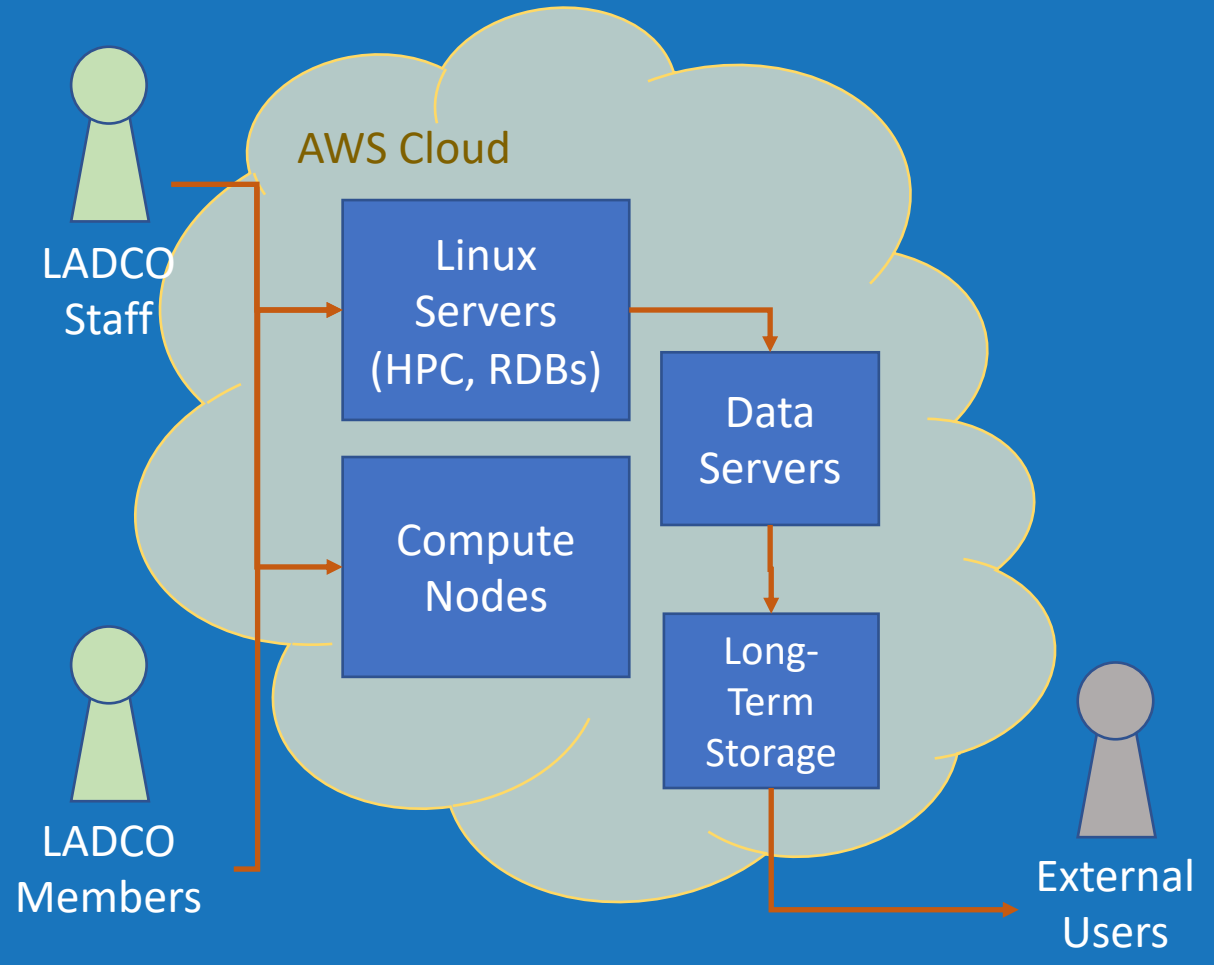
Meteorology and Air Quality Modeling Using the AWS Parallel Cluster Cloud Service

Zac Adelman and Tsengel Nergui

Presented at the 2021 CMAS Conference
November 4, 2021

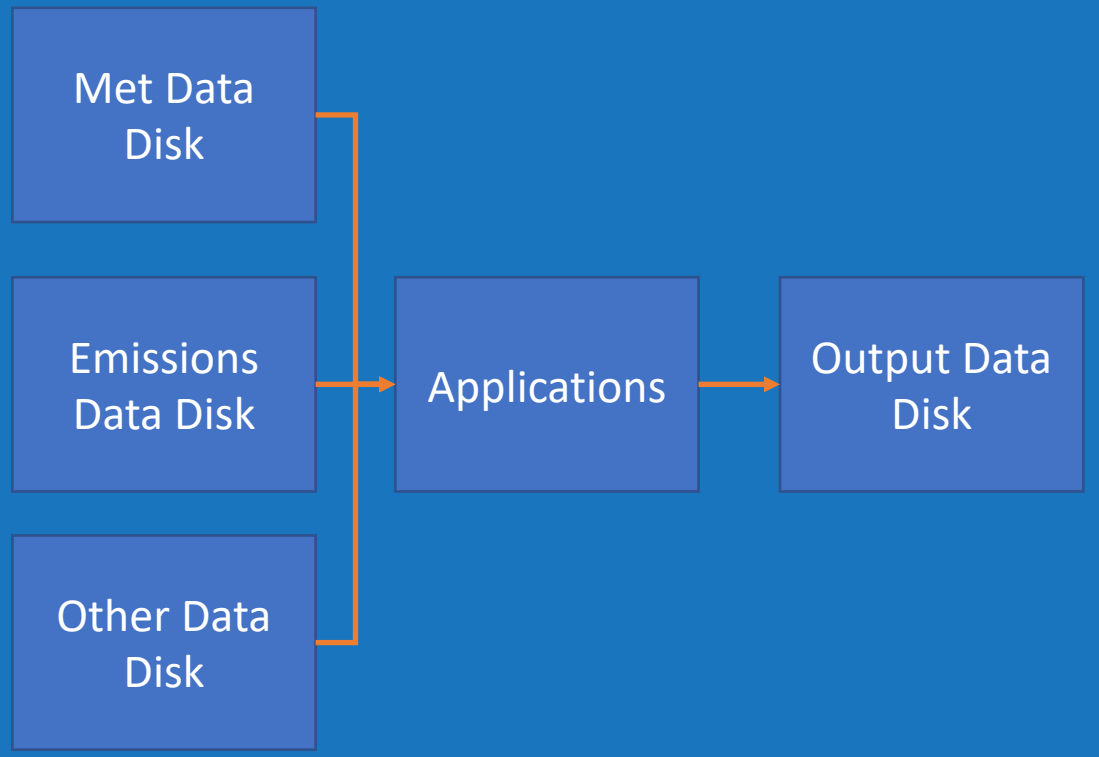
Why does LADCO need cloud computing?

- Supplement/replace the LADCO local compute cluster
- Access scalable high performance computing (HPC) resources
- Circumvent IT barriers at state air programs
- Enable collaboration in modeling and data analysis across the Consortium



Modeling Platform As a Service (PaaS)

- A PaaS is a pre-configured system with all data, software, and computing capacity needed for an application
- AWS Parallel Cluster provisions computing as needed and to meet demand
- LADCO modeling cluster includes all libraries and modeling software: WRF, SMOKE, CMAQ, CAMx, AMET, NCL, VERDI, R



Why Amazon Web Services Parallel Cluster?

- LADCO developer had familiarity with computing @ AWS
- At the time of development, AWS CfnCluster project was migrating to Pcluster
- How it works?

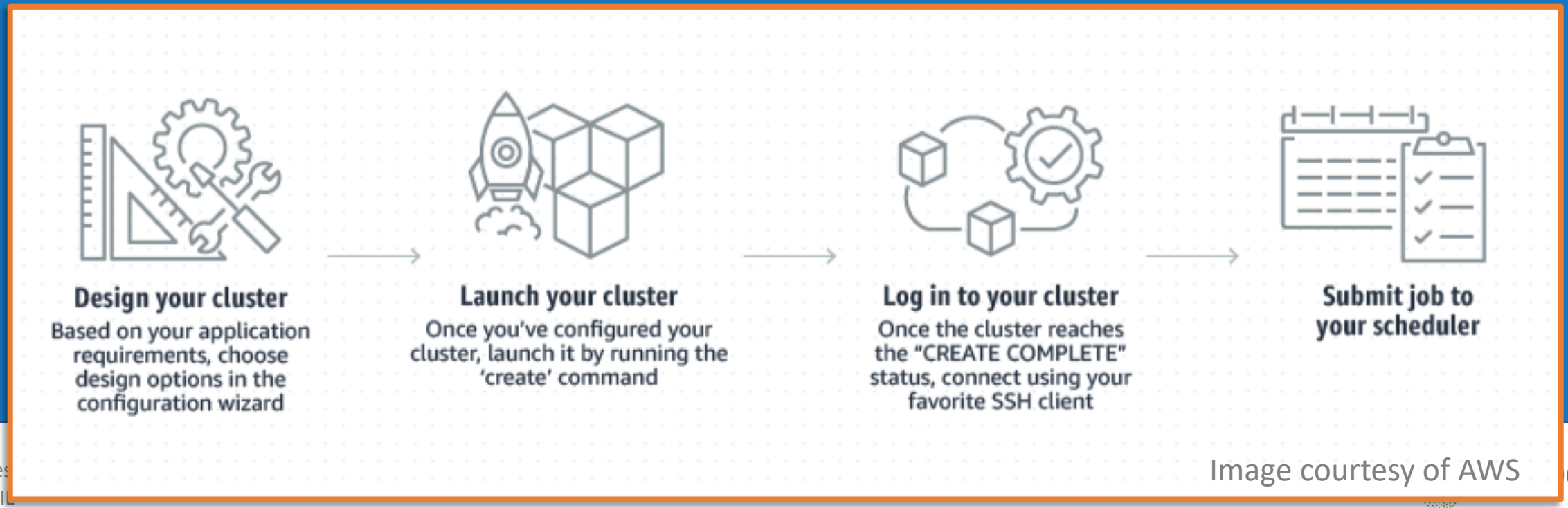
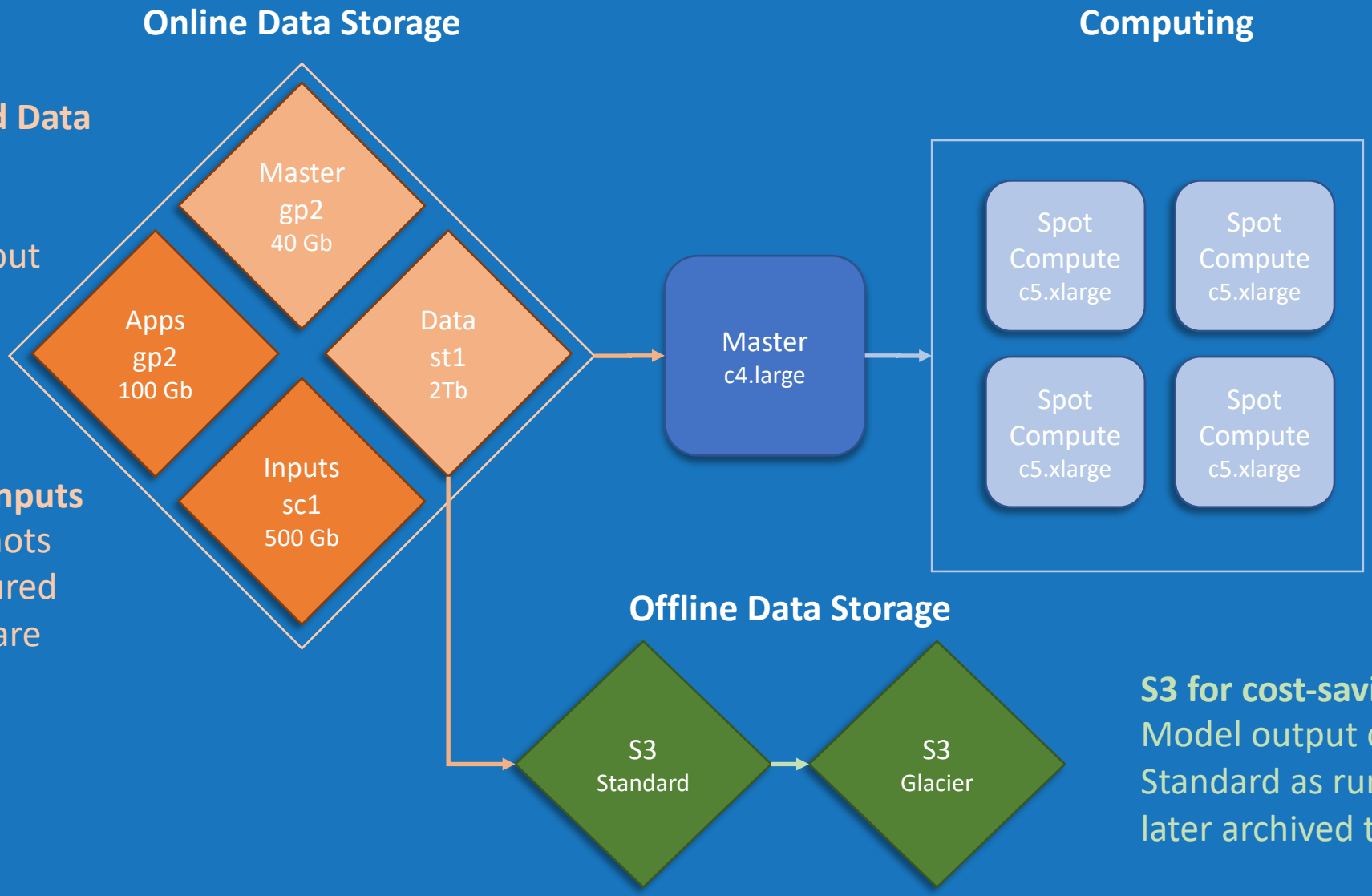


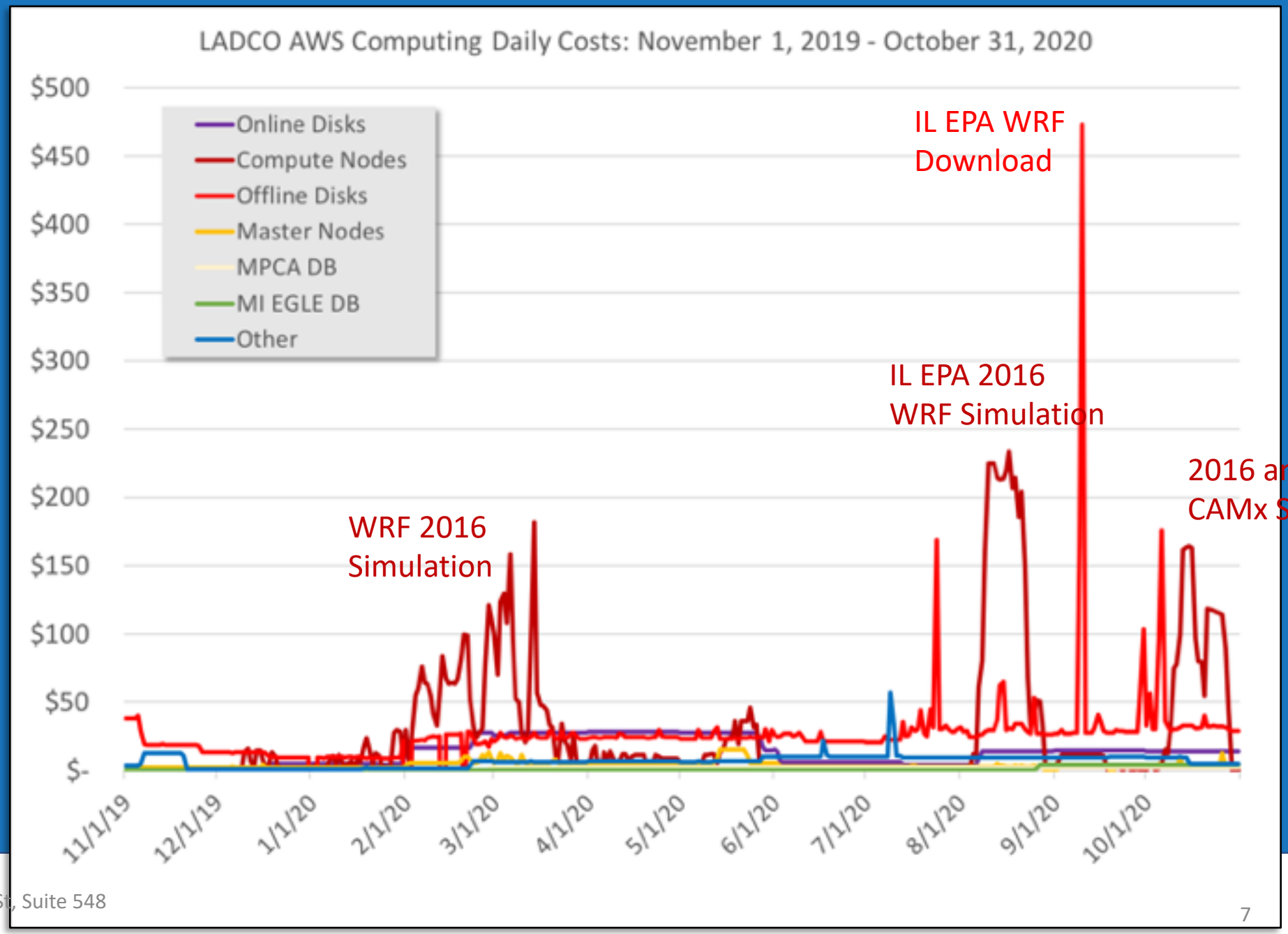
Image courtesy of AWS

LADCO AWS Modeling Schematic



LADCO Modeling Statistics @ AWS

Simulation	Simulation Period	EC2 Compute Instance	CPUs/ Run	Total CPUs	Hrs/ Run Day	CPU Costs ($)	$/Sim Day
WRF 12/4/1.33/1.33 Annual	15 days + 12 hour spin up	c4.2xlarge	48	1248	8	3,325	8.33
CAMx 12km 2016 and 2028	6 months + 10 day spin up	c4.2xlarge	64	256	1	2,220	2.81
CAMx 12km 2028 PSAT (28 srctags x N+S)	3 months + 10 day spin up	c4.2xlarge	96	384	4	3,830	9.22
CAMx 12km 2028 PSAT (31 src tags x N+S)	3 months + 10 day spin up	r5.2xlarge	96	384	4	3,210	7.74
CAMx 12km 2028 PSAT (24 region+src tags x N+S)	3 months + 10 day spin up	r5.2xlarge	96	288	6	3,860	9.53
CAMx 12km 2028 PSAT (25 region+src tags x N+S)	4 months + 10 day spin up	r5.2xlarge	96	384	5.5	2,420	5.83
CAMx 12km 2028 PSAT (25 region+src tags x Primary)	3 months + 10 day spin up	r4.16xlarge	96	384	6.5	4,825	11.63
CAMx 12km 2016 Annual	3 months + 10 day spin up	r4.16xlarge	64	256	1	1,160	2.80

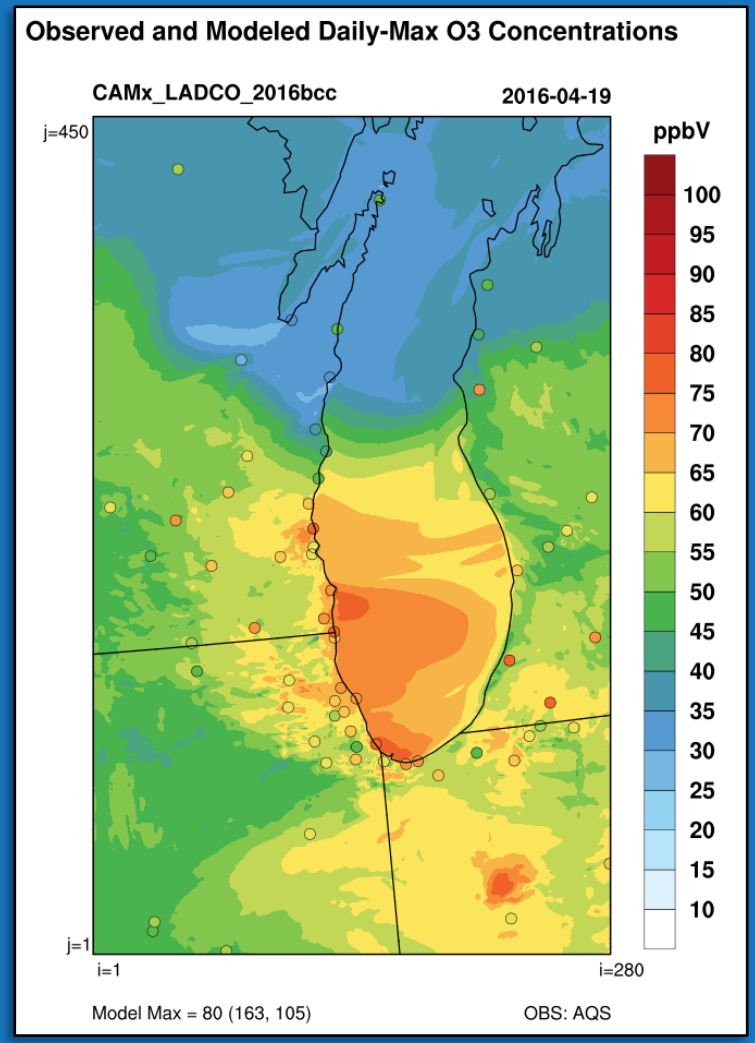

LADCO AWS Computing Daily Costs: November 1, 2019 - October 31, 2020
$500
$450
$400
$350
$300
$250
$200
$150
$100
$50
$-
Online Disks
Compute Nodes
Offline Disks
Master Nodes
MPCA DB
MI EGLE DB
Other
11/1/19
12/1/19
1/1/20
2/1/20
3/1/20
4/1/20
5/1/20
6/1/20
7/1/20
8/1/20
9/1/20
10/1/20
IL EPA WRF Download
IL EPA 2016 WRF Simulation
WRF 2016 Simulation
2016 and 2028 CAMx Simulations

Notes on Data Management

- Automate workflows
  - Online EBS volumes are the most expensive storage unit, need to automate moving data between less expensive offline and more expensive online disks
- The AWS Simple Storage Solution (S3) is a relatively cheap place to store data, and it's great for transferring large volumes of data
  - Egress fees ~$100/Tb to download
  - AWS Command Line Interface (CLI) uses multiple threads to transfer data, up and down from S3; terabyte scale electronic transfers are practical with S3
  - Example: > 4 Tb of WRF data downloaded by IL EPA in ~24 hours @ $500

Key Outcomes

- All LADCO modeling on AWS is done using Spot Instances
  - Learned to read the AWS Spot Market to find available, last-generation instances
- AWS is a practical solution for meteorology and air quality modeling
  - Annual 12km U.S. CAMx simulation costs ~$2.80/model day
  - Adding nested grids or instrumented models (i.e., PSAT) makes runs more expensive because they require more CPU-hours
- Don’t be afraid to scale up
  - Larger instances are more expensive per hour, but about the same cost per CPU-hour (the jobs run more quickly)
- LADCO invested ~$20k over a year to develop modeling capabilities on AWS



Next Steps

- Optimize AWS costs
  - Better manage offline and online storage costs
  - Shut down our compute servers more often
- Collect more data points on best practices
  - Transfer LADCO capabilities to other interested agencies
  - Learn from others using AWS (or other cloud services)
- Knowledge transfer to others in the community
  - MPI expertise to help determine the best configuration for AWS hardware
- National air quality modeling cloud computing workgroup

LADCO Cloud Computing Development Timeline