



# ARM Data Center

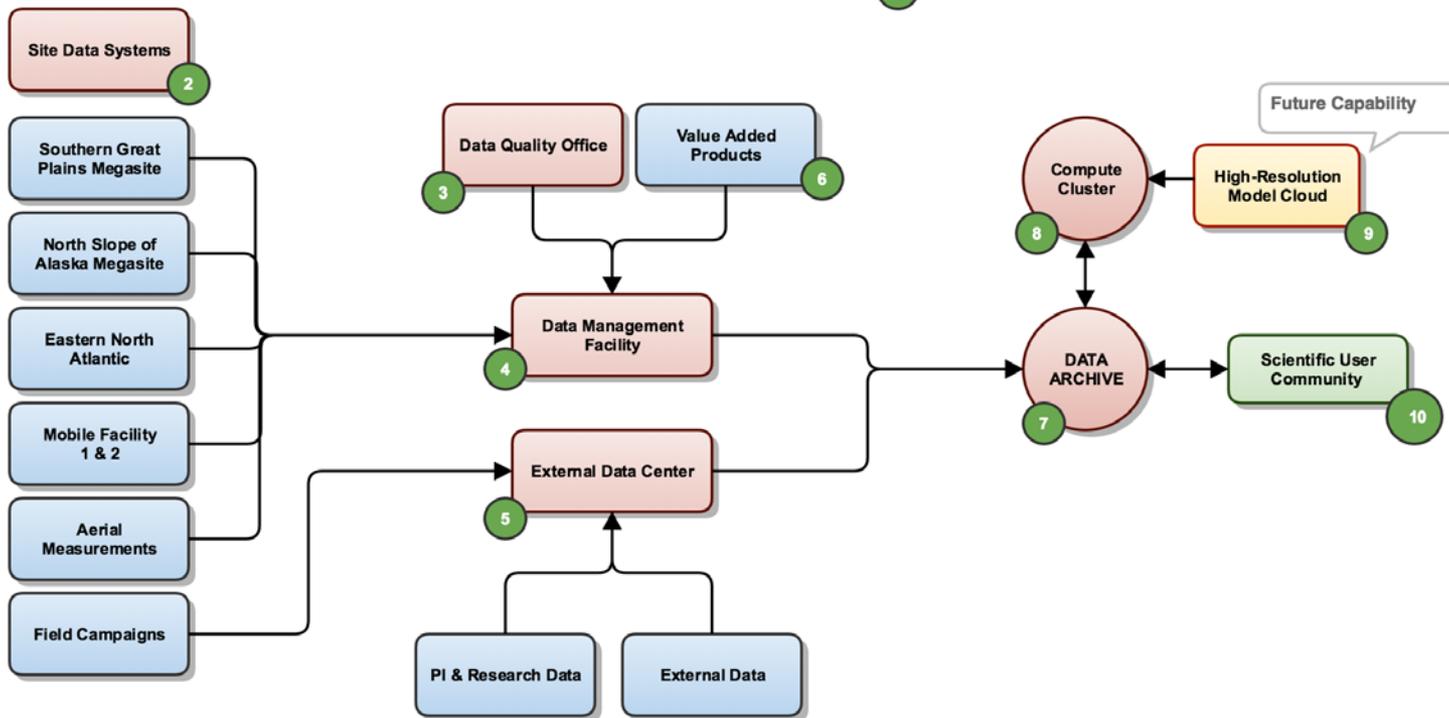
ARM Data and Computing Capabilities for PIs and Users

**Giri Prakash**

**ARM Data Center**

Future Capability

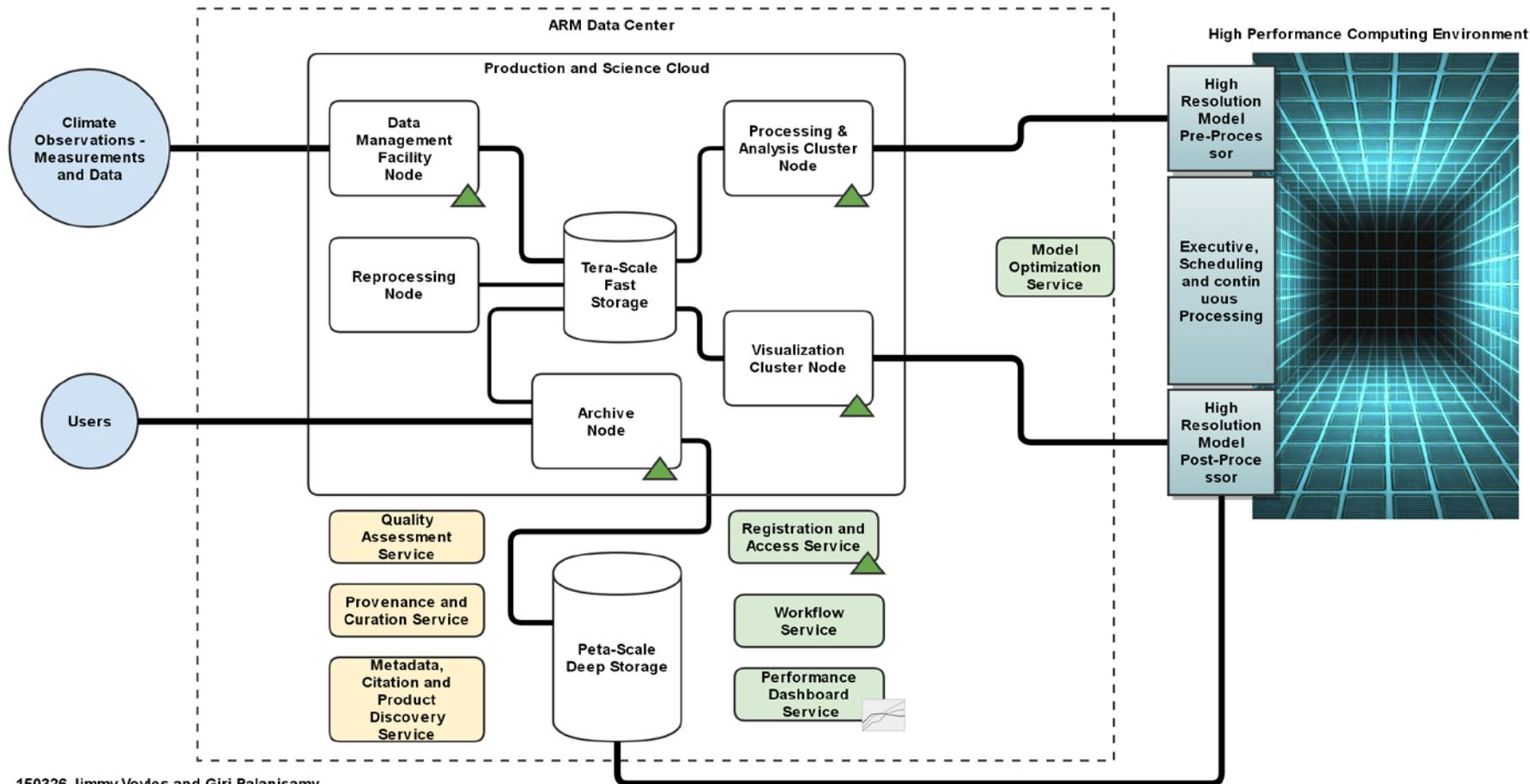
# ARM Adaptive Computing Architecture



1	<b>ARM Adaptive Computing Architecture</b>	This concept is being specified and evolved to provide the data tools, connections, and software for scalable microservices.
2	<b>Site Data Systems</b>	Collects instrument and measurement data from all sites and transmits it to DMF for processing. Also, provides on-site real-time processing to provide data locally to PIs.
3	<b>Data Quality Office</b>	Provides review of over 5000 data products per day for data quality. Automated QA documentation and diagnostic tools are provided.
4	<b>Data Management Facility</b>	Processes ~25 TB of measurement data per month. 7.7k files and 174Gb is the daily average for processing raw data to higher order products. A unified software development environment is used for processing, Value Added Processing, and reprocessing.
5	<b>External Data Center</b>	Processes ~25 external data products per month. Collection and ingest of external climate data, field campaign, and PI data products. Provides program wide assistance for Metadata management.

6	<b>Value Added Processing</b>	To produce high impact synthesis data products from observations to facilitate scientific discovery.
7	<b>Data Archive</b>	Manages the storage and discovery of all ARM data. Data rate to the Archive is ~15 TB/Month and increasing geometrically. Total storage is currently ~710TB.
8	<b>Compute Cluster</b>	Provides a managed and unified software development and computing environment for the science community and infrastructure for high-volume processing, analysis, and visualization jobs.
9	<b>High-Resolution Modeling Cloud</b>	A future processing framework and Elastic cloud based architecture to initialize, execute, evaluate, and improve High Resolution Models.
10	<b>Scientific User Community</b>	Register for and subscribe to use ARM data products for research and scientific discovery. Data quality and product metadata are provided. Scientific data users provide feedback to the program on the utility of products and suggest improvements.

# Cyberinfrastructure for ARM LES Model Development and Operations



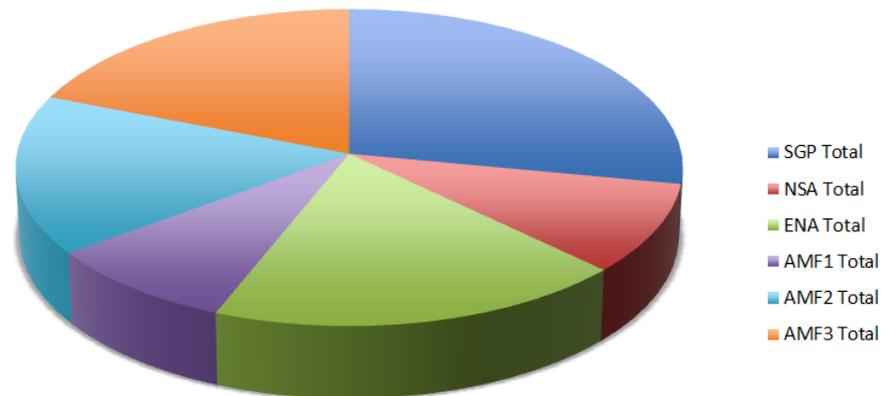
150326 Jimmy Voyles and Giri Palanisamy

# Meeting Data Growth Challenges

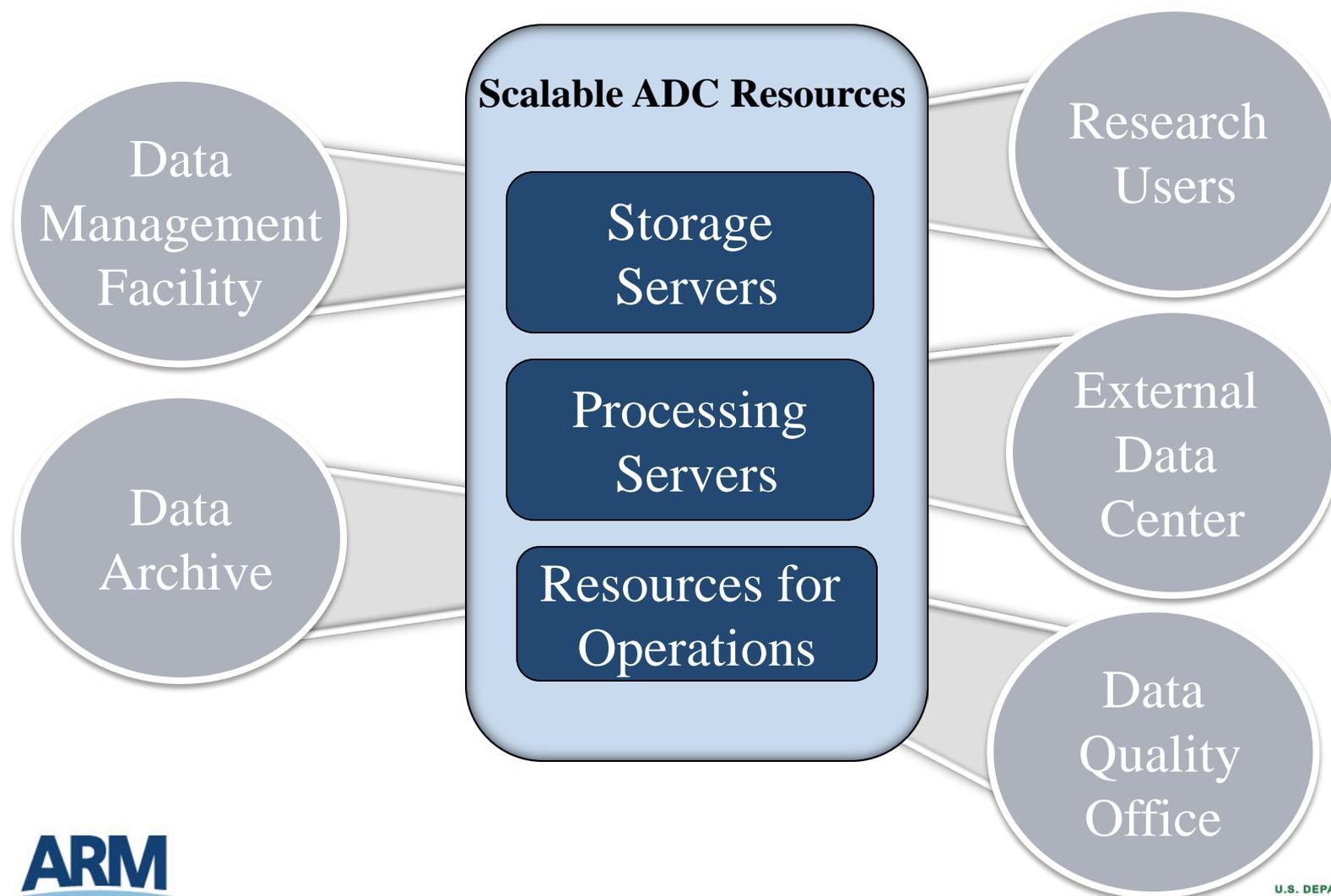
## Anticipated Data Volumes

Site	# of sites	MB/month	Total MB/month	2 Month	4 Month
SGP C1	1	9,983,963	9,983,963	19,967,926	39,935,852
SGP XSAPR	3	2,678,400	8,035,200	16,070,400	32,140,800
SGP CSAPR	2	2,678,400	5,356,800	10,713,600	21,427,200
SGP RWP	3	245,300	735,900	1,471,800	2,943,600
SGP EF	13	1,184	15,392	30,784	61,568
SGP Profiling Site	4	54,844	219,376	438,752	877,504
SGP KA/X SACR Site	2	6,032,538	12,065,076	24,130,152	48,260,304
SGPC1 AOS	1	152,761	152,761	305,522	611,044
		<b>SGP Total</b>	<b>36,564,468</b>	<b>73,128,936</b>	<b>146,257,872</b>
NSA C1	1	11,980,171	11,980,171	23,960,342	47,920,684
NSA EF	2	800	1,600	3,200	6,400
		<b>NSA Total</b>	<b>11,981,771</b>	<b>23,963,542</b>	<b>47,927,084</b>
ENA	1	24,376,062	24,376,062	48,752,124	97,504,248
ENA AOS	1	152,761	152,761	305,522	611,044
		<b>ENA Total</b>	<b>24,528,823</b>	<b>49,057,646</b>	<b>98,115,292</b>
AMF1	1	10,797,434	10,797,434	21,594,868	43,189,736
MAOS	1	622,970	622,970	1,245,940	2,491,880
		<b>AMF1 Total</b>	<b>11,420,404</b>	<b>22,840,808</b>	<b>45,681,616</b>
AMFC2	1	21,087,006	21,087,006	42,174,012	84,348,024
AMF2 AOS	1	152,761	152,761	305,522	611,044
		<b>AMF2 Total</b>	<b>21,239,767</b>	<b>42,479,534</b>	<b>84,959,068</b>
AMF C3	1	24,563,415	24,563,415	49,126,830	98,253,660
AMF3 AOS	1	152,761	152,761	305,522	611,044
		<b>AMF3 Total</b>	<b>24,716,176</b>	<b>49,432,352</b>	<b>98,864,704</b>
		<b>All Sites Total</b>	<b>130,451,409</b>	<b>260,902,818</b>	<b>521,805,636</b>

## Data Volume Breakdown

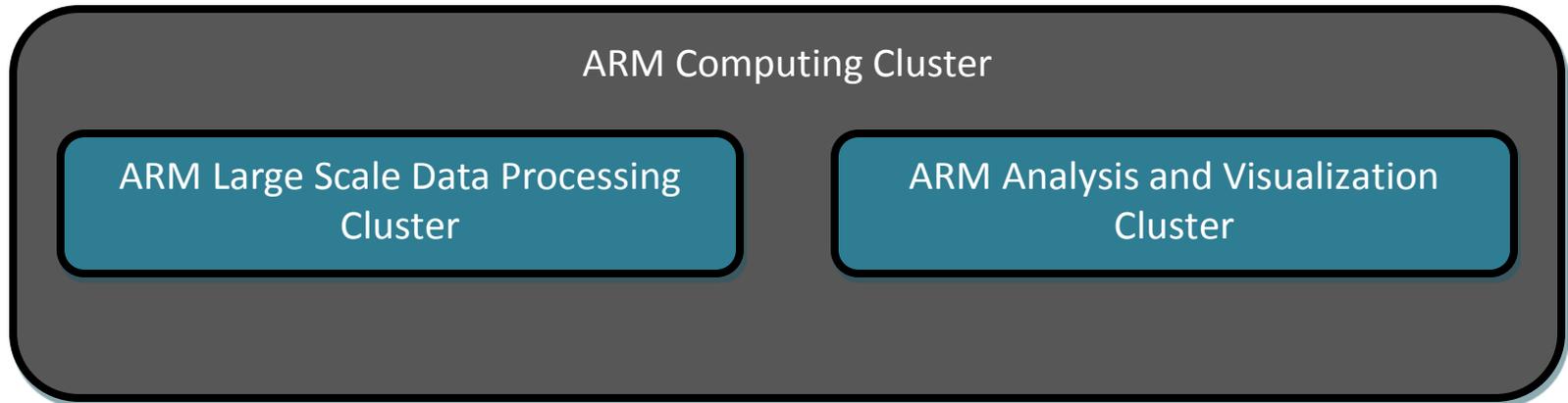


# Meeting Data Growth Challenges

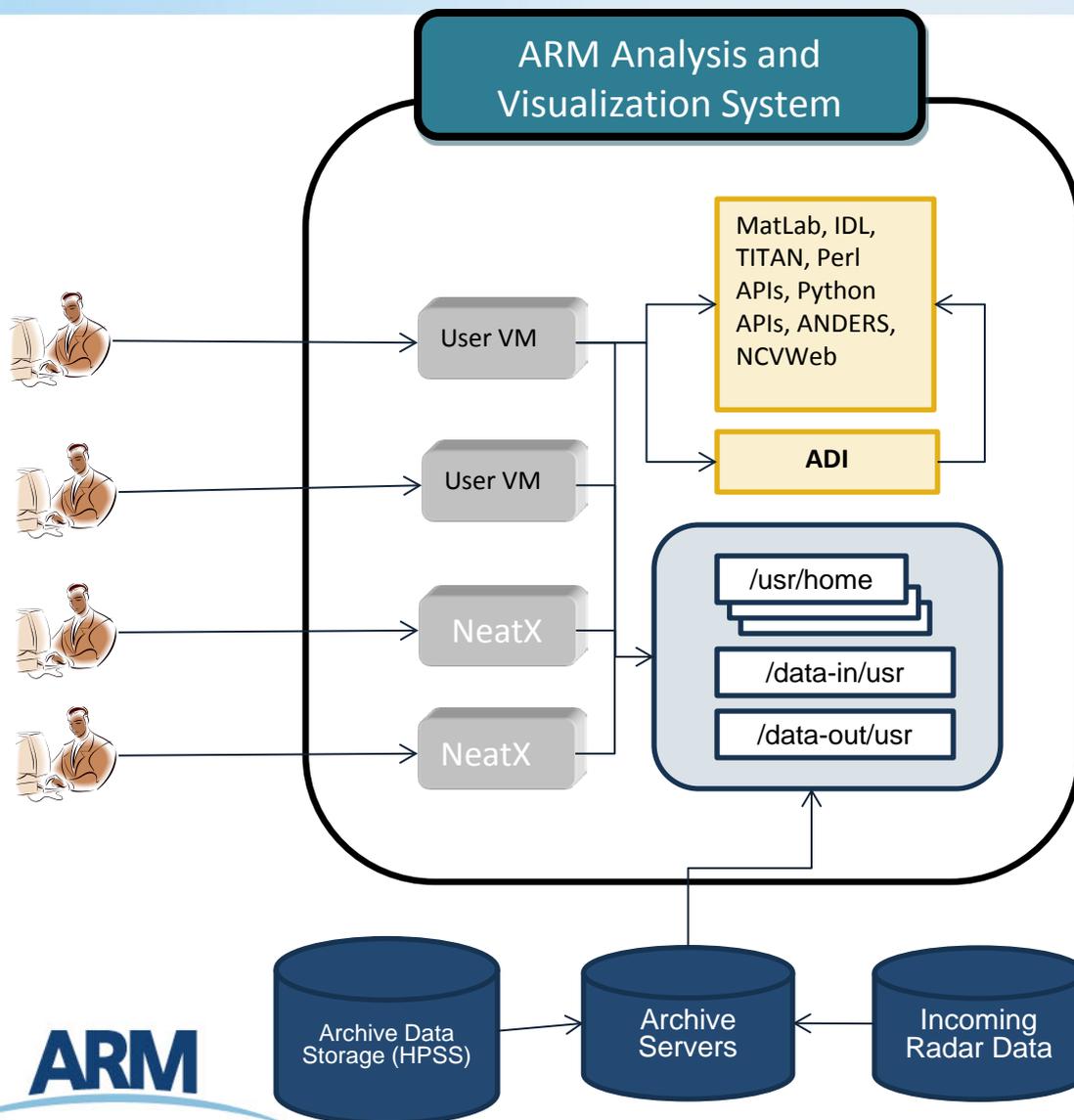


# ARM Computing Cluster

A Community Resource for ARM Data Processing,  
Visualization and Analysis



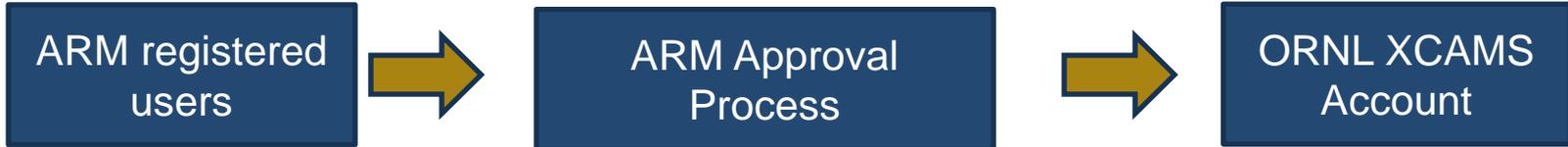
# ARM Computing Cluster for Large - Scale Radar Data Analysis



- **Interactive visualization**
  - Users request access to the cluster
    - Almost instance access to the Archive users
    - need to create one more account (can use the same archive user id)
  - User-saved “views” and data extractions can be made available via web services
- **Software Development**
  - Access to local copy of ARM Development Interface

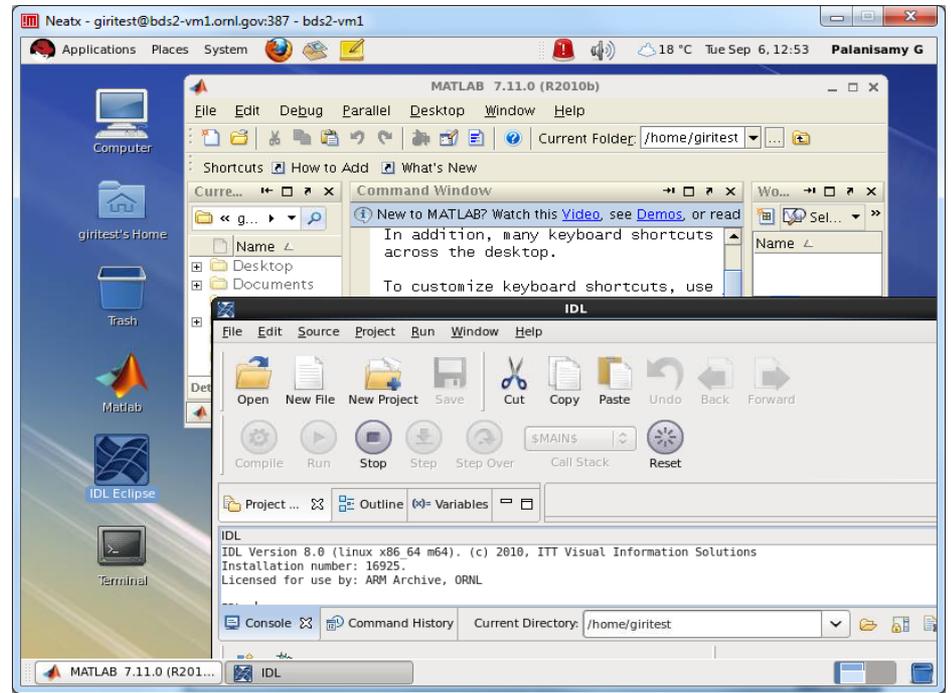
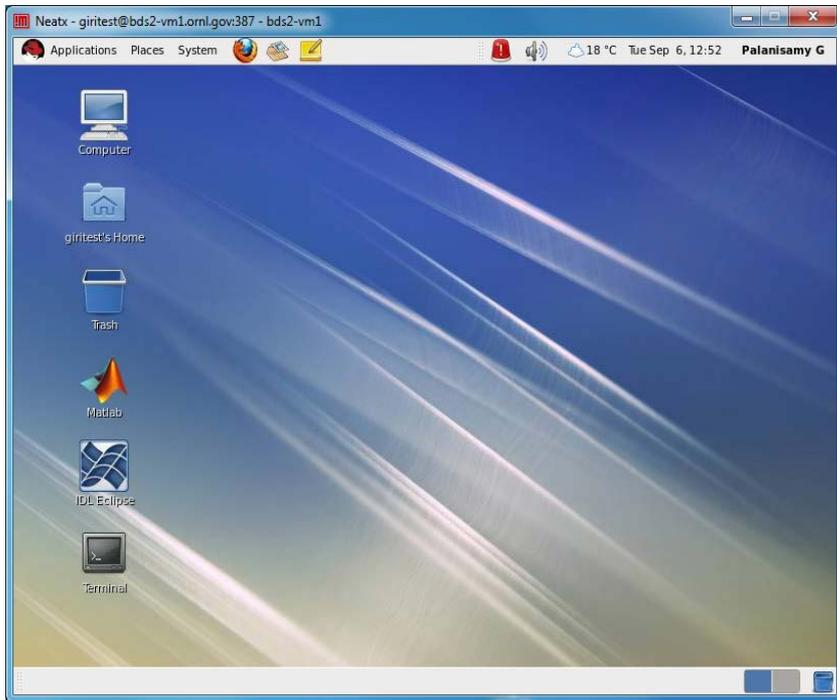
32 cores, 256 – 512 GB memory, 40-100 TB of local disk

# ARM User Workflow for ARM Cluster (BDS)



Remote login ARM cluster

Use tools to analyze/visualize data





# ENG-1061 - ARM Next Generation - Computing Architecture for Big Data Processing and High Resolution Modeling

- Collecting requirements for computing clusters to help
  - Routine data processing of big data (example: scanning Radars)
  - Data analytics and visualization of big data
  - Research user data processing (example: users of BDS clusters)
  - Routine operation of LASSO
- Information Needed:
  - Summary of your use case
  - # Nodes/cores
  - Memory
  - Storage within the cluster
  - Software (processing, queuing etc..)
  - Compilers
  - GPU

# Preliminary findings from ENG 1061

High performance computing needs for ARM data processing, analysis and modeling can be served by a combination of:

- Mid-scale compute clusters
  - 512-1024 cores (Intel Xeon)
  - 8-16 GB RAM per core
  - 0.5 - 1 PB storage
  - A single rack mount node can look like: 16 (2x8) or 32 (or 2x16) cores, 256 - 512 GB RAM, 100 TB HDD
- Leadership class computing facility
  - OLCF / ALCF allocations would provide access to a larger systems for large jobs (like LES simulations)
  - Shared Lustre system accessible from the mid-scale computing system would provide seamless access to all data
- Long term data archival
  - HPSS accessible from both mid-scale and LCF systems

# ARM PI data Product Registration Tool (OME)



\*This record has been approved

The tool only supports ASCII characters and UTF-8 formatting. If you are copying text from other documents, please, avoid any special characters and other formats. But, HTML tags like hyperlinks are supported

[Back](#)

[Preview Metadata](#)

\* = required

[\[logout\]](#)

- [Data Type](#) >
- [Author](#) >
- [Describe Data](#) >
- [Spatial Info](#) >
- [Measurements](#) >
- [Data Quality](#) >
- [Constraints](#) >
- [Tools](#) >
- [Additional](#) >

### Data Product

First, within the ARM program, is this data product/set only associated with ARM field campaign?\*

Answer 'yes' to this question if the dataset was collected for an ARM field campaign. Answer 'no' if this is an independent dataset that did not use ARM infrastructure, or a product developed for regular ARM data, or for use with more than one field campaign.

Yes  No

Please select the ARM campaign:

New Particle Formation Study 2013 (NPF2013)

### Contact information about the author of this metadata record.

Full Name (First Name followed by Last Name):\*

*Begin typing a letter (slowly) to see the list of existing fields. This list is supplied from arm database.*

Smith, James

ID

Organization Name: \*

NCAR

Position Name:

Telephone Number: \*

3034971468

Email: \*

jmsmith@ucar.edu

Address:

Street: PO Box 3000

# ARM DOI Tool (For PIs)

**ARM** CLIMATE RESEARCH FACILITY **DOI CREATION TOOL** U.S. DEPARTMENT OF **ENERGY** Office of Science

[Home](#) [logout](#)

**Title\***

**Author/PI\***

**Author's/PI's email addresses\***

**Keywords\***

**Description\***

Just reserve the DOI

**Site uri\***

**Product size\***

**Publication date\***

**Is your data?**

**Do you have an instrument code?**  Yes  No (instrument code is auto-generated for you)

**Instrument code\***

# ADC Contacts

- Data Ingest and Reprocessing (DMF)– Nicole Keck
- ADC data/project resource: Rob Records
- ADC Clusters: Anthony Clodfelter
- PI and Field Campaign Data – Alice Cialella
- Metadata – suggestions, issues: David Troyan