

The Virtual Climate Data Server (vCDS): An iRODS-Based Data Management Software Appliance Supporting Climate Data Services and Virtualization-as-a-Service in the NASA Center for Climate Simulation

John L. Schnase¹, Glenn S. Tamkin^{2,3}, W. David Ripley, III^{2,3}, Savannah Strong^{2,3}, Roger Gill^{2,4}, and Daniel Q. Duffy²¹

Office of Computational and Information Science and Technology² NASA Center for Climate Simulation (NCCS)³ Computer Science Corporation (CSC), ⁴ Innovim, LLC NASA Goddard Space Flight Center Greenbelt, MD 2077 I



Part 1 – Background (5 Minutes) •vCDS Concept and Rationale •vCDS 0.9 Anatomy / vCDS 0.9 Products Part 2 – Where We Are Now (10 Minutes) •NetCDF/IPCC Toolkit •Adminstrative Extensions •Repetitive Provisioning •Operational Deployment •Amazon Cloud vCDS-IPCC-ESG-v0.9

 $Part \ 3 - Wrap \ Up \ (5 \ {\rm Minutes} \ +)$

•Next Steps

Discussion





Concept and Rationale

Scenario

A customer approaches the NCCS with a new dataset they want us to manage ...

Q. What technology is needed to quickly meet that customer's requirement under the follow constraints:

•The solution should be: simple, fast, and cheap;

•provide core capabilities to get started, but extendable to accommodate future needs;

•be flexible, with the ability to use, optimize, and change deployment configurations in response to resource availability;

•allow the new dataset to be integrated into an existing data collection; and

•come with a help desk and user support?

The DMS Project has been looking at **iRODS data** grid software as a potential solution ...

Definitions

Customer – an individual scientist, a lab, project, or mission.

Dataset – may be products generated by a GCM, may be observational data or a subset thereof, reanalysis data, or specialized products of value to an individual scientist or lab.

Manage – may refer to short-term file storage, long-term archival preservation; data may be used online by a person or application.

Examples

Abound:
IPCC AR5 data for ESG.
MODIS Atmospheres data for CMIP5.
MERRA downscaled meteorological and environmental data.
AgMIP, CERES, SMOS, Laboratory for Atmospheres, the Snowfake project ...



We begin the next phase by working with IPCC AR5 products ...

FY12 Q1/Q2

Managed Collections: (1) Publication Datasets

•GISS Collection - Served to ESG in Amazon

- Ingest IPCC AR5 data into CDS 0.9
- Operationally harden CDS 0.9 to CDS 1.0 (TRL 9)
- Expose collection to ESG publication system
- Develop Collection Administrator's interface





A production system

in Nebula that

mirrors the NCCS's

current

capabilities ...

Develop requirements, implement Collection Administration policies and mechanisms, and specify OAIS Policy Metadata – all relatively easy with Publication Datasets.

Published Collections	Estimated Current Size (TB)	Estimated Final Size (TB)
GISS IPCC	3.5	60
MERRA	0.5	0.5
CERES	0.1	0.1
AgMIP	0	0
GMAO IPCC	0	60
Total	4.1	120.6

Current estimates as of 9/1/2011. Total estimated size of IPCC AR5 200 TB.





vCDS V0.9 Anatomy



The DMS Project Team - NASA Goddard Space Flight Center

The Data Management System Project



vCDS V0.9 Products

IPCC / NetCDF Module	iRODS microservices, rules, configuration settings, and software utilities required to implement canonical CRUD operations for IPCC/NetCDF system kernel	 Microservice Code Microservice Utilities IPCC / NetCDF Rules Configuration File 	Build Drv Tree 090 000 000 000 0000 000 000 000 0000 000 000 000 0000 000 000 000 0000 000 000 000 0000 000 000 000 0000 000 000 000 0000 000 000 000 0000 000 000 000	IPCC Microservices IPCC iCAT IPCC Rules
Administrative Extensions	iRODS Postgres extensions and utilities to log system-level object provenance and provide QA for OAIS metadata compliance (plus associated Rich Web Browser GUI extensions)	 OAIS Object Views Object Action Logging PHP Browser Extensions 	RI PDI DM	iCAT Rule Engine iRODS 2.5
Repetitive Provisioning	RPM script to build software stacks for the SLES 11 SP1 (IaaS), iRODS AE (PaaS), and CDS/IPCC (SaaS) virtual images	 Automatic Installation VaaS Architecture 	AND	IPCC Microservices IPCC iCAT IPCC Rules
Deployment and Distribution	Product library, documentation, and SLA infrastructure for distribution, deployment, and help desk support	• Tech Transfer Plan • Tech Transfer Team • UNC RENCE Partnership		iRODS 2.5 AE SLES 11 SP 1 Climate Data Server V0.9

The DMS Project Team - NASA Goddard Space Flight Center

6



IPCC / NetCDF Module



CDS 0.9 Products

1.IPCC / NetCDF Module *iRODS microservices, rules, configuration settings, and software utilities required to implement canonical CRUD operations for IPCC/NetCDF system kernel ...*

•Administrative Extensions

iRODS Postgres extensions and utilities to log systemlevel object provenance and provide QA for OAIS metadata compliance (plus associated Rich Web Browser GUI extensions) ...

•Repetitive Provisioning

RPM scripts to build software stacks for the SLES 11 SP1 (IaaS), iRODS AE (PaaS), and CDS/IPCC (SaaS) virtual images ...

•Deployment and Distribution

Product library, documentation, and SLA infrastructure for distribution, deployment, and help desk support ...



The DMS Project Team - NASA Goddard Space Flight Center

The Data Management System Project



Open Archival Information System (OAIS)

An OAIS is an archive, consisting of an organization of people and systems that has the responsibility to preserve information and make it available for a designated community ...

The reference model addresses a full range of <u>archival</u> <u>information preservation functions</u> including:

ingest, data management, access, and dissemination;
the migration of digital information to new media and forms;
the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives.

And it identifies both internal and external interfaces to the archive functions;

•it identifies a number of high-level services at these interfaces;

•it provides various illustrative examples and some '<u>best</u> <u>practice</u>' recommendations;

•and it defines a minimal set of <u>responsibilities</u> for an archive to be called an OAIS.







Administrative Extensions

Major Functions

Basic system-level capabilities to log object provenance and provide OAIS package views of object metadata ...

- •iRODS Postgres extensions
- •iRODS microservice extensions
- •Rich Web Browser extensions

N. Alverting Received.	1347/merra_Zo.							
						rode@koahost134715aa.0M		
	- V :	Lelect All 🧳	Browse Up	New * 🥥 Delete +	Q Uplead * P	tore		
ma_Zone	Marrie			Resource	Dise.	Date Medified -		
home	100.0		Lassim Institu 24	dencResc	125.86 MB	July 14, 2010, 11:02 am		
goutik:			Lassim Instit. 3d		126.24 MB	July 14, 2010, 11:07 am		
C-62323					127.29 MB	July 14, 2010, 11.11 am		
a 🛄 1980			Lassim.instH_3d_					
iii 🗀 1561			Lassim.instH_3d_		128.23 MB	July 14, 2010, 11.15 am		
a Co 1962 a Co 1963	180 ×	ACRIANCE prov	hassiminath_34_	dencResc	126.73 MB	July 14, 2010, 11.19 am		
a C 1964	litt +	ERRA100 proc	.assim.instH_3d_	denoResi	126.46 MB	July 14, 2010, 11 24 am		
a - 1965			A MARCH INSTRUMENT		124.78 MB	July 14, 2010, 11 29 am		
a 🛄 1966	100	10 P		of a constant of the				_
iiii 🦢 1967	and the state of		gooknarks Both I					
a 🗀 1968	Do For							
a 📥 1960	÷ +	· 6 0 :	😤 🗑 https://14	9.354.348.315/reds/brows	e php#ruci=rods@ioca	ites/h3A2347/mena_2sne/home/public	men 🗆 🕈 🚺 🖬 🖓	
1990	12		etimena Ze O					
a Co 1990	W HAR LINE							-
a Co 1990		MERRALDO.	prod.assim.inst94,	34_asm_Cp.197991	Liner 🔉		- 3	
ar 📥 1994	Collectors	C free						
a 📥 1996	a camera J							
a 🔁 1996	ia cahone	0 AM 9	Remove 5 Relo	ad 📷 Save				
	901	Name -		Malue			(deal)	
Inter-				Sea-level pressure, Surfa-				1
iiii 🧰 1998		a variables				otential, Geopolential height, Coone Marry/R	ate	
a Co 1998		tariables 15e		MERROA reanalysis. GEDD		potential, Geogradential height, Crome Maing Ru	ala .	
a Ca 1998 a Ca 1998 a Ca 1999				MERROR reanatysis. GEOD Citobal Modeling and Assim	620		*	
2000 2000 2000 2000 2000		104			1523 Mater Office, GEOSign.,		*	
		100 Anarite		Global Modeling and Assir	1523 Mater Office, GEOSign.,		*	
 1998 1999 2001 2001 2002 2003 2003 2004 		Elle scorrie references		Clobal Mixleling and Assir Mgs (gmac.galt, ness.gov 8.00000004+147	1520 Idelon DRice: GEOSope.) Instanch/Imercal			
 1998 1998 2001 <li< td=""><td></td><td>Ste source references restriction</td><td></td><td>Clobal Mixleling and Assir Mgs (gmac.galt, ness.gov 8.00000004+147</td><td>1520 Idelon DRice: GEOSope.) Instanch/Imercal</td><td>w</td><td></td><td>2 2 8</td></li<>		Ste source references restriction		Clobal Mixleling and Assir Mgs (gmac.galt, ness.gov 8.00000004+147	1520 Idelon DRice: GEOSope.) Instanch/Imercal	w		2 2 8
 1996 1996 2001 2001 2000 2000 2000 2000 2000 2005 2006 		59 source offerences missing, value trailinder		Cinhal Multiley and Asse Mp. (grave get), near get 3.3000006+147 Global Mobiling and Assi File witten by DFIO HDFEOE, VD. 14	6523 Haton DRice: GEO/Gopt_ Insearch/Internal Haton DRice: NASA Godt	1,3,3 and Space Fight Garlier, Graaniteit, MD 2017		
 in 1996 in 2000 2001 2001 2001 2001 2002 2002 2003 2004 2005 2005 2005 2005 2005 2005 		Ma source ordenances manag_value haitbake haitbake		Cinhal Multiley and Asse Mp. (grave get), near get 3.3000006+147 Global Mobiling and Assi File witten by DFIO HDFEOE, VD. 14	6523 Haton DRice: GEO/Gopt_ Insearch/Internal Haton DRice: NASA Godt	w		
		Ne marie relevences relation beliador latin) falloscorector		Cinhal Multiley and Asse Mp. (grave get), near get 3.3000006+147 Global Mobiling and Assi File witten by DFIO HDFEOE, VD. 14	6523 Haton DRice: GEO/Gopt_ Insearch/Internal Haton DRice: NASA Godt	1,3,3 and Space Fight Garlier, Graaniteit, MD 2017		
 a 1986 a 2000 a 2000 a 2001 a 2001 a 2002 a 2002 a 2003 a 2005 a 2006 a 2006 a 2006 a 2006 a 2006 		De suurie rebances restig valle zelluter tatory falseverster dramatre		Color Modeling and Asser Mp (gene get), raise gev 3.30000000+147 Colora Noteling and Asser File within by CFIO HOFEOS, v2.14 TIME 2016/40 + 1, VDM	1523 Nation DRice: SECOlogie, J Insulant DRice: NASA Goot SECOSOFIE = 144, IDand	1,3,3 and Space Fight Garlier, Graaniteit, MD 2017		
a _ 1988 a _ 1988 a _ 1988 a _ 2000 a _ 2000 a _ 2000 a _ 2000 a _ 2006 a _ 2006 a _ 2006 a _ 2006 a _ 2007 a _ 2006 a _ 2		De surre réleances restrig value ballador faitury fallesconstan dimensione consertions		Contactificating and Asset Mg (Ignae gelt near gelt 3.0000006-147 Contactificating and Asset File within 1y CPIO HOPEOL_V2.141 TIME_CONSIND = 1, YDW CPI.10	1523 Nation DRice: SECOlogie, J Insulant DRice: NASA Goot SECOSOFIE = 144, IDand	1,3,3 and Space Fight Garlier, Graaniteit, MD 2017		
a _ 1988 a _ 1988 a _ 1988 a _ 2000 a _ 2000 a _ 2000 a _ 2000 a _ 2006 a _ 2006 a _ 2006 a _ 2006 a _ 2007 a _ 2006 a _ 2		Be searce relevances relevances relevances relevances devences convertiens contect		Contactificating and Assor Mg ((group get); name get) 8 mm2000+140 Contactificating and Assor File written by CFIO HOFFIOR, VO.14 Table; DOGG/VO.14, 1/Der CFI.10 Mg ((group get); name get)	1523 Halan Office OEOSays, Insularchinama' Halan Office, NASA Gold (2006/HD = 144, HOrnS	1,3,3 and Space Fight Garlier, Graaniteit, MD 2017		
a _ 1988 a _ 1988 a _ 1988 a _ 2000 a _ 2000 a _ 2000 a _ 2000 a _ 2006 a _ 2006 a _ 2006 a _ 2006 a _ 2007 a _ 2006 a _ 2		19e source oderances realization realization realization realization constant constant constant constant		Contactificating and Asso Mg (Ignae, gel), ness geo 8.0000000-140 Gotosi Noteing and Asso Pile without by CP10 HOFE005, V2.14 Table:0006/400 = 1, Y(be CP110 Mg (Ignae, gel), ness geo 0000-52.0	1523 Halan Office OEOSays, Insularchinama' Halan Office, NASA Gold (2006/HD = 144, HOrnS	1,3,3 and Space Fight Garlier, Graaniteit, MD 2017		
 a 1996 a 1996 a 2000 a 2000		58 Nutria Internetation Patholocum Patholocum Natura Siternetation Contect Contect Contect Contect Contect Contect Contect Contect Contect	minida Coper.	Golar Hickey and Asset High Ignac gale, rese, gale Streamber H-H Golar Hickey and Asset Pile with the O'HO HICKEY, O'HO HI	1523 Halan Office OEOSays, Insularchinama' Halan Office, NASA Gold (2006/HD = 144, HOrnS	1,3,3 and Space Fight Garlier, Graaniteit, MD 2017		

CDS 0.9 Products

1.IPCC / NetCDF Module iRODS microservices, rules, configuration settings, and software utilities required to implement canonical CRUD operations for IPCC/NetCDF system kernel ...

To appear in iRODS 3.0 release ...

•Administrative Extensions / iRODS Postgres extensions and utilities to log systemlevel object provenance and provide QA for OAIS metadata compliance (plus associated Rich Web Browser GUI extensions) ...

•Repetitive Provisioning

RPM scripts to build software stacks for the SLES 11

//rods(localhost:1247/merr			#ruri=rods@localhost%3A1247/i	erra_Zone/home/public/mei 🏠 🔹 🚼			pisauri - Pisating Status 🔝. est 1347/acc2e +	afest Headlines	~ D	uholla Prefer Y						
MERRA100.prod.assim.instM_3d_asm_Cp.197901.hdf					Sign Out		3 4, Marcal (17) p1, 19111-1920	11.ee									
ns rra_Zc home		131996336 Bytes) localhost:1247/merra_i	Zone/home/public	:/merra/1979/MERRA100.prod	issim.instM_3d_asm_Cp.197901.hdf	,⊊ Ned → 102 am	Rope: 14.05.60 (1973) Robis URA: 1053 (210 Research at feedback Type: parters: Data Maddhad: 2115	Lathout 124/Npcc2onathomatrodulture		542.4,%	Ancal, drivel, 1	R0107-192512-AK					
	Resource: demoR	esc				:11 am	Rame .	Value	Unit	туре	Category	CAS Subcategory	Bound	Ormanatan	Function	HEAT	
	Type: generic Date Motified: We	d Jul 14 2010 11:02:48	8 GMT-0400 (EST)			model_14	05542A		grout	PDI	CTX				10086	
		i Jul 14 2010 11:02:48				:15 am parent_experime modeling_realm	parent_experiment	pre-industrial control sealce		good	PDI	CTX				10086	
				host:1247/merra_Zone/home/p		:19 am	realization	3		grout		CTX				10086	
	Audit History					:24 am	experiment of	Network		anna	PDI	CTX				10086	
	Action Name	Action Description	Actor Name	Audit Date	3	:29 am											
	Create Data		rods	2010-07-14T11:02:48-04:00		:33 am											
	Access Granted	modify metadata	rods	2010-07-14T11:04:47-04:00	T	:37 am											
	Access Granted	read object	rods	2010-07-14T11:04:47-04:00		:41 am											
*	Create Metadata	-d	rods	2010-07-14T11:04:47-04:00		:46 am											
	Create Metadata	-d	rods	2010-07-14T11:04:47-04:00	Ļ	:50 am											
	A	-	d+	0040 07 41744-04-17 04-00	×												
							increase, instants	a mere serve meree	Teacher 14	der spool							



SVN

LIB

BAK

Repetitive Provisioning

First, about our development environment ...

The DMS Project has worked in a virtualized environment – including MacTops with VMware Fusion and a VMware vSphere dev/test server farm.

Dev

DR

VIRTUALIZED BY

mware[.]

Test

RG

VMware vSphere™

Build

F

Ĩ

SS

CDS 0.9

iRODS 2.5

SLES 11

RPM

GT

This environment has influenced the way we are thinking about building, distributing, and deploying CDS components ...

IPCC Microservices

IPCC iCAT

IPCC Rules

iRODS 2.5 AE

SLES 11 SP 1 Climate Data Server V0.9

/ME

Rationale: Why create an RPM?

•Automate installation

Goal is to be able to conveniently install iRODS and our vCDS software stack in different environments and on different platforms ...

•Reduce installation errors and eliminate the user inteface

Installing iRODS "out of the box" is cumbersome, and manual installation leads to errors and unstable systems ...







FUSION

10

The Data Management System Project



Operational Deployment - Amazon Cloud vCDS-IPCC-ESG-v0.9



The DMS Project Team - NASA Goddard Space Flight Center

The Data Management System Project



... then move to other primary and derived products ...

FY12 Q3/Q4

Managed Collections: (1) Publication Datasets

•MERRA Collection

- Develop iRODS MERRA kit
- Develop iRODS HDFS, Swift, S3 drivers
- Create testbed HDFS, Swift, S3 repositories
- Expose collection to ESG publication system

•CERES, AgMIP, GMAO Collections

Replicate/refine the Managed Collections processes as needed to accommodate customer interest, response, and needs – and budget, time, and political constraints ...

•NASA / RENCI jointly developed products:

CDS V1.0 Enterprise Edition
 NetCDF, HDF science kits
 IPCC, MERRA climate kits
 HDFS, Swift, S3 drivers
 iDROP Collection Administrators GUI

	Published Collections	Estimated Current Size (TB)	Estimated Final Size (TB)
	GISS IPCC	3.5	60
	MERRA	0.5	0.5
	CERES	0.1	0.1
	AgMIP	0	0
-	GMAO IPCC	0	60
	Total	4.1	120.6

Current estimates as of 9/1/2011. Total estimated size of IPCC AR5 200 TB.







... then to the challenging task of active research collections.

FY13 and beyond ...

Managed Collections: (2) Research Datasets

•Operational Research

- Transition from "Archive" to "Managed Collections"
- Approach will be stepwise, incremental, logical
- Need established technology, a process, and an expert team
- And a conceptual model for how this is done ...





Develop requirements, implement user policies and mechanisms, and specify OAIS Policy Metadata and Discovered Metadata – this is where we add layers to the kernel.

Research Collections	Estimated Current Size (TB)	Estimated Final Size (TB)			
Person	???.?	???.?			
Person	???.?	???.?			
Project	???.?	777.7 777.7 777.7			
Project	???.?				
Lab	???.?				
Total	???.?	???.?			

Circumscribed Datasets

amazon webservices Public Cloud - -**Provisioning Paths** Hybrid Cloud NEBULA Community Cloud Migration Paths **Federation Paths** VME Private Cloud **Provisioning Paths** Hybrid Cloud Data Center

NCCS



Discussion



14