

Clouds and the Earth's Radiant Energy System (CERES)

Data Management System

CATALYST Test Plan Version 6

Primary Authors

Nelson Hillyer, Joshua Wilkins

Science Systems and Applications, Inc. (SSAI)
One Enterprise Parkway
Hampton, Virginia 23666

NASA Langley Research Center
Climate Science Branch
Science Directorate
21 Langley Boulevard
Hampton, VA 23681-2199

SW Delivered to CM: October 2016
Document Date: February 2017

Document Revision Record

The Document Revision Record contains information pertaining to approved document changes. The table lists the date the Software Configuration Change Request (SCCR) was approved, the Version Number, the SCCR number, a short description of the revision, and the revised sections. The document authors are listed on the cover. The Head of the CERES Data Management Team approves or disapproves the requested changes based on recommendations of the Configuration Control Board.

Document Revision Record

SCCR Approval Date	Version Number	SCCR Number	Description of Revision	Section(s) Affected
09/25/12	V1	931	<ul style="list-style-type: none"> New document. Formatting issues were fixed and long hyphens were replaced with short hyphens for easier testing. (12/11/2012) 	All All
03/07/13	V2	937	<ul style="list-style-type: none"> Updated for CATALYST Build 1. Updated for CATALYST Build 1.0.3. (08/27/2013) Updated for CATALYST Build 1.0.4. Added new information for activation code in the installer. (11/19/2013) Removed manual steps that are now automatically handled by the CATALYST server. (11/19/2013) 	All All Sec. 2.1 Sec. 3.1
04/28/14	V3	1008	<ul style="list-style-type: none"> Added "NOTE" for installing server updates. Added '-daemonize' flag to server start instructions. Added instructions for testing JIRA ticket CER-146. Added instructions for testing JIRA ticket CER-122. Added instructions for testing JIRA ticket CER-83. Added additional promotion information. 	Sec. 2.1 Sec. 3.1 Sec. 3.1.4 Sec. 3.1.5 Sec. 3.1.6 Sec. 4.0
06/14/14	V4	1021	<ul style="list-style-type: none"> Added new appendix to hold JIRA ticket specific test cases. Moved existing JIRA ticket test cases to the new appendix. Added test cases for CER-95 and CER-159. A few minor corrections were made concerning formatting and heading titles. (06/24/2014) 	Sec. 3.1 & App. D App. D

Document Revision Record

SCCR Approval Date	Version Number	SCCR Number	Description of Revision	Section(s) Affected
04/02/15	V5	1065	<ul style="list-style-type: none"> Removed Appendix D. Added explicit mentioning of actual user names of 'CATALYST administrator' to test cases and installation instructions. (09/16/2015) Added instructions to installation steps to install patch. (09/29/2015) Added instructions to installation steps to install patch #02. (10/22/2015) Modified a few formatting issues. (10/28/2015) 	<p>App. D Secs. 2.0 & 3.0</p> <p>Sec. 2.1</p> <p>Sec. 2.1</p> <p>Sec. 2.1</p>
09/28/16	V6	1176	<ul style="list-style-type: none"> Updated installation instructions. Updated test instructions. 	<p>Sec. 2.1</p> <p>Secs. 3.1, 3.2, & 3.3</p>

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
Document Revision Record	ii
1.0 Introduction.....	1
1.1 Document Overview	1
1.2 CATALYST Overview	2
1.2.1 Perl Library Modules	2
1.2.2 Architecture/Location Dependent Code	2
2.0 Software Installation Procedures	3
2.1 Installation.....	3
3.0 Test and Evaluation Procedures	5
3.1 CATALYST Server v2.1 and Operator's Console v2.2 Update Description	5
3.2 Test strategy at a glance	6
3.3 Executing the CATALYST Test Suite.....	7
3.3.1 Preliminary Checks	7
3.3.2 Setup	8
3.3.3 Testing [CER-253]:.....	8
3.3.4 Testing [CER-211, CER-301]:.....	9
3.3.5 Testing [CER-53, CER-300, CER-318].....	9
3.3.6 Testing [CER-299] Part 1:	10
3.3.7 Testing [CER-278, CER-316, CER-317]:	10
3.3.8 Complete testing of [CER-299]:	10
3.3.9 Wrapping up:	11
Appendix A - Acronyms and Abbreviations	A-1
Appendix B - CATALYST Directory Structure Diagram.....	B-1
Appendix C - File Description Table.....	C-1
C.1 Executable Scripts	C-1

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 3-1. Three PRs have been created to test these seven tickets, and those PRs are available in the PPE PR Tool. CER-299 is a two part test case, which is why it's listed twice above.	7
Figure B-1. CATALYST Library Directory Structure	B-1

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table C-1. \$PERL5LIB directory	C-1

1.0 Introduction

CERES is a key component of EOS and NPP. The first CERES instrument (PFM) flew on TRMM, four instruments are currently operating on the EOS Terra (FM1 and FM2) and Aqua (FM3 and FM4) platforms, and NPP (FM5) platform. CERES measures radiances in three broadband channels: a shortwave channel (0.3 - 5 μm), a total channel (0.3 - 200 μm), and an infrared window channel (8 - 12 μm). The last data processed from the PFM instrument aboard TRMM was March 2000; no additional data are expected. Until June 2005, one instrument on each EOS platform operated in a fixed azimuth scanning mode and the other operated in a rotating azimuth scanning mode; now all are typically operating in the fixed azimuth scanning mode. The NPP platform carries the FM5 instrument, which operates in the fixed azimuth scanning mode though it has the capability to operate in a rotating azimuth scanning mode.

CERES climate data records involve an unprecedented level of data fusion: CERES measurements are combined with imager data (e.g., MODIS on Terra and Aqua, VIIRS on NPP), 4-D weather assimilation data, microwave sea-ice observations, and measurements from five geostationary satellites to produce climate-quality radiative fluxes at the top-of-atmosphere, within the atmosphere and at the surface, together with the associated cloud and aerosol properties.

The CERES project management and implementation responsibility is at NASA Langley. The CERES Science Team is responsible for the instrument design and the derivation and validation of the scientific algorithms used to produce the data products distributed to the atmospheric sciences community. The CERES DMT is responsible for the development and maintenance of the software that implements the science team's algorithms in the production environment to produce CERES data products. The Langley ASDC is responsible for the production environment, data ingest, and the processing, archival, and distribution of the CERES data products.

1.1 Document Overview

This document, CATALYST Test Plan, is part of the CATALYST delivery package provided to the Langley Distributed Active Archive Center (DAAC). It provides procedures for installing and testing the CATALYST software. A list of acronyms and abbreviations is provided in [Appendix A](#), a directory structure diagram is contained in [Appendix B](#) and a description of the software and data files is contained in [Appendix C](#).

This document is organized as follows:

[Section 1.0](#) - Introduction

[Section 2.0](#) - Software Installation Procedures

[Section 3.0](#) - Test and Evaluation Procedures

[Appendix A](#) - Acronyms and Abbreviations

[Appendix B](#) - Directory Structure Diagram

[Appendix C](#) - File Description Tables

1.2 CATALYST Overview

The CATALYST service contains no PGEs. Rather it is a framework for coordinating the execution of CERES PGEs in the ASDC production environment. This build of CATALYST contains the functionality necessary for running the Clouds and Inversion Edition 4 processing chain as well as the login interface required for the PR Web Tool.

1.2.1 Perl Library Modules

CATALYST is written primarily in Perl. Many of the routines required by CATALYST are located in CERES's Perl_Lib. The **\$PERL5LIB** environment variable must be defined before attempting to run the CATALYST server. **\$PERL5LIB** can be set using **\$CERESENV**.

1.2.2 Architecture/Location Dependent Code

There is only one location that CATALYST is to be installed: the *AMI-P* X86_64 cluster head node. This requirement exists because CATALYST must be able to communicate both internally, within the *AMI-P* cluster for job coordination, and externally to the PR Web Application, CATALYST Operator's Console, and sub-programs built using the CATALYST XML-RPC API for operator interaction and control.

2.0 Software Installation Procedures

This section describes how to install the CATALYST software in preparation for making the necessary test runs at the Langley DAAC. The installation procedures include an executable installation script which unpacks CATALYST software, configures run time environments, and adjusts the logging database as necessary.

2.1 Installation

1. The scripts and makefiles in the CATALYST delivery package expect the CERES environment variable, **\$CERESENV**, to point to a file which sets the following environment variables:

CERESHOME - Top directory for CERES software
CERESLIB - Top directory for CERESlib software (this location will be different for the different CERESlib versions)
PERL5LIB - Directory containing CERES Perl module

2. CATALYST Server 2.1 Installation Steps:

Login to *catalyst.larc.nasa.gov* (zb27.cluster.net) as the appropriate CATALYST administrative user for the environment for which you are running (vobadm in PPE, catalyst in Production). All steps below are to be performed on catalyst.larc.nasa.gov (zb27.cluster.net):

- a. Before installing CATALYST v2.1 shutdown the existing CATALYST v2.0:

\$CERESHOME/catalyst/bin/stop_catalyst.sh

- b. Unpackage the CATALYST v2.1 update:

tar xf catalyst_server-1176.tar.gz -C \$CERESHOME

- c. Apply the configuration file update for CATALYST Server v2.1:

Unpackage the CATALYST v2.1 ancillary package:

tar xf catalyst_server_anc-1176.tar.gz -C \$CERESHOME

Depending on your environment, perform one of the following:

1. If CERES CM, run the following:

\$CERESHOME/catalyst/catalyst_conf_update-1176.sh CM

2. If ASDC PPE, run the following:

\$CERESHOME/catalyst/catalyst_conf_update-1176.sh PPE

3. If ASDC Production Environment, run the following:

\$CERESHOME/catalyst/catalyst_conf_update-1176.sh Production

Report any error messages, should they appear. You should see “update complete” if the script has run successfully.

- d. Start CATALYST v2.1:

\$CERESHOME/catalyst/bin/start_catalyst.sh

- e. Start all (“start all”) of the CATALYST subcomponents via the CATALYST manager utility:

\$CERESHOME/catalyst/bin/catalyst_manager.sh

- f. Wait at least 10 seconds, and then check to make sure that all the CATALYST subcomponents are still running using the “status” command in the CATALYST manager utility. If any one of the components is not running, please stop here and report this information to the development team.
- g. If all CATALYST subcomponents are running then CATALYST Server v2.1 was successfully.

3. CATALYST Operator’s Console v2.2 Installation Steps:

- a. Download the latest version of the CATALYST Operator’s Console from CATALYST Home Page (http://ceres.larc.nasa.gov/Internal/intern_docs_catalyst.php) to your desktop.
Note: Older CATALYST Operator’s Console(s) can still technically function, but they will not be utilizing the new features available with CATALYST Server v2.1.

3.0 Test and Evaluation Procedures

This section provides instructions for executing the CATALYST test suite. (See Section 2.1 for an explanation of the **CERESENV** environment variable.)

Please note: The test suite must be executed on *catalyst.larc.nasa.gov*, on which CATALYST has been installed. If problems are encountered in any of the tests, immediately contact one of the CATALYST developers.

3.1 CATALYST Server v2.1 and Operator's Console v2.2 Update Description

CATALYST Server v2.1 introduces several new features and stability improvements, many of which go hand-in-hand with user interface updates in CATALYST Operator's Console v2.2. New features include:

1. [CER-53] Better PR Load balancing within CATALYST. This feature allows CATALYST to spread its attention more broadly across active PRs, such that one PR with runnable items does not totally saturate production prohibiting other PRs from executing runnable items. This also improves upon the predictability when a job from a PR will run, since PRs will be dealt with in a round-robin fashion.
2. [CER-211, CER-301] This gives the ability for the CATALYST logdb process, which ordinarily interfaces with the CATALYST Log Database, to directly inspect ASDC_archive for existing input products. This should reduce the need for log database updates.
3. [CER-253] Allow for multiple PRs for a PGE that share Sampling Strategy, Production Strategy, and Configuration Code, but do not allow date ranges to overlap. This will allow the submission of a new PR that extends the processing range of another (previously submitted) active PR.
4. [CER-278, CER-316, CER-317] Visual epilogue queuing order. This server feature allows the Operator's Console to do two things with regards to epilogues.
 - a. In the PR's job display, any jobs that are in state "epilog_queued" have a numeric suffix, such as "[12]", in their state. This number tells you that this job is 12th in line to have its epilogue process executed. As the previous 11 epilogues are processed, you will see this number decrease until "epilog_running" is reached. This should give a better indication as to why a job's queued epilogue has not run, and should give better estimations as to when it will run. Note that this feature only shows you items for the currently selected PR; for information across all PRs, the next item will be of benefit. The testing of this covers Operator's Console v2.2 ticket [CER-316].
 - b. This aspect of [CER-278] allows the CATALYST Operator's Console to display a listing of all epilogue tasks in order of execution, across all active PRs, in a unified display. For instance, you will be able to see that there are 12 epilogue jobs from PR #1 which come before the 5 that are in PR #2 which come before the 24 from PR #3. This will give you additional ability to estimate when items will be archived. The testing of this covers Operator's Console v2.2 ticket [CER-317].

5. [CER-287] This is a fix for a rare issue where a job's state displays that it is running, but in reality, it has not started. This is caused by a timing bug between the CATALYST kernel and cluster processes, for which internal measures such as request timeouts have been added to rectify the problem. Because of the rarity with this issue, successfully reproducing this bug is difficult and required significant "rigging" (forcibly inducing lag between the cluster and kernel processes) in the development test system in order to reproduce the symptoms. There will be no test cases in this document specifically for this item, other than observing that this issue does not reappear over time.
6. [CER-289] This adds a configurable property to CATALYST that should make occurrences of SGE_QDELS in jobs to be less frequent. There is a periodic lag between a job's completion (i.e. dropping off the qstat list), and its recording in the SGE accounting record (examined with qacct). CATALYST v2.0.x had retries for checking this, but the number of retries and the delays between the retries has been found to be insufficient in operation. This update adds these retry counts and delays as configurable parameters in the CATALYST server configuration file, so that if the updated values are still insufficient in practice, a simple two line change in the CATALYST configuration file would be needed, instead of a code update. Other than adding this new parameter to the CATALYST server configuration file, there is no associated test case because of the difficulty to reproduce this issue (which seems to be related to high I/O activity on the disk where the SGE accounting record is stored).
7. [CER-299] Ability to flush SGE jobs in CATALYST. This provides the functionality to prevent new SGE jobs from starting, but allows the currently executing jobs to run to completion. This feature can be toggled on and off at runtime depending on its necessity, such as a planned cluster/SGE downtime where grid engine queues need to be deactivated.
8. [CER-300, CER-318] Ability to view a qstat-like listing of CATALYST jobs in the Operator's Console. This server-side feature provides methods for the user on the Operator's Console to view currently running CATALYST SGE jobs. This display is a graphical form of the familiar "qstat" command. The testing of this covers Operator's Console v2.2 ticket [CER-318].

3.2 Test strategy at a glance

At a glance, here is the order with which these tickets will be tested (CER-287 and CER-289 are not listed for reasons described in the previous section):

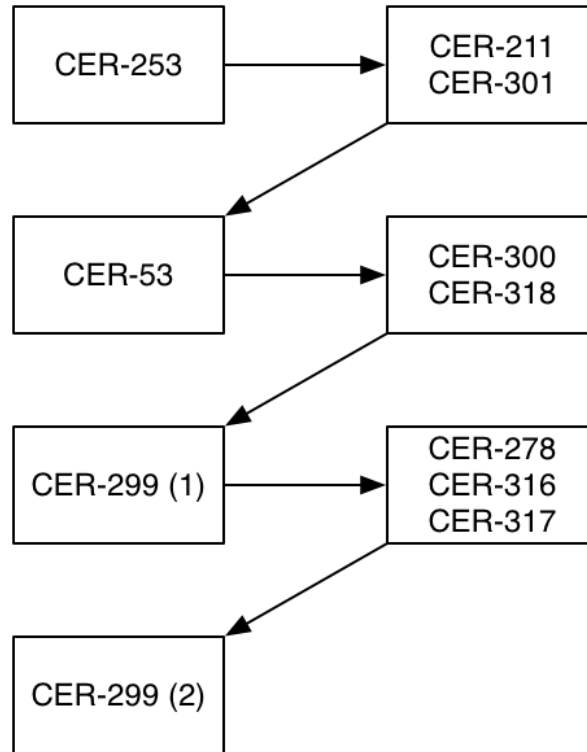


Figure 3-1. Three PRs have been created to test these seven tickets, and those PRs are available in the PPE PR Tool. CER-299 is a two part test case, which is why it's listed twice above.

3.3 Executing the CATALYST Test Suite

3.3.1 Preliminary Checks

1. Operator's Console v2.2:
 - a. Start the Operator's Console and connect to the appropriate CATALYST Server on catalyst.larc.nasa.gov (port 4021 for PPE, and port 4020 for Production).
 - b. If you have any trouble connecting to the CATALYST server, please stop here and report the problem to the CATALYST development team.
 - c. If there were any PRs that were open in CATALYST v2.0, you will be able to see them in CATALYST v2.1.
2. PR Tool Login Verification:
 - a. Verify that the PR Tool can still reach the CATALYST server for which you are testing by confirming that you can log in with your AMI username and password (PPE PR Tool for the PPE, Production PR Tool for the production environment).
 - b. If your credentials are rejected (and you're positive that you've typed them correctly), or the PR Tool reports that it cannot reach the CATALYST server, please stop and report this to the CATALYST development team.

- c. Successfully logging in to the PR Tool verifies that the PR Tool can successfully connect to and query the CATALYST v2.1 server installation.

3.3.2 Setup

1. Select “Start Processing” from the “CATALYST Server” menu in the CATALYST Operator’s Console.
2. Stop the CATALYST cluster process using either the Operator’s Console or the command line manager utility. This provides the opportunity to see the PR balancing [CER-53] more quickly regardless of how fast or slow the test steps are performed.
3. Ensure that the SARB subsystem is enabled and epilog-enabled.
4. Ensure that PGE CER7.2.1P2 is enabled and the epilog is *disabled*. It is important that this PGE’s epilog is *disabled* because that allows tickets [CER-278], [CER-316], and [CER-317] to be verified with many epilogues queued up.
5. Stop the CATALYST logdb process using either the Operator’s Console or the command line manager utility.
6. In order to perform some of the testing to follow, we need to remove a set of records from the CATALYST log database in order to force CATALYST to try to find certain items on the DPO. To do this, a script included with the ancillary tar file can perform this action. Only run this in the CM or PPE environment. Do not run this in production.

\$CERESHOME/catalyst/catalyst_db_test_prep-1176.pl

7. Report any errors should they appear. This script should complete with the message “update successful”.
8. Restart the CATALYST logdb process.

3.3.3 Testing [CER-253]:

1. Using the PPE PR Tool submit PR #634-17 to CATALYST. This PR should appear within the CATALYST Operator’s Console momentarily covering the month of September 2004.
2. Attempt to submit PR #635-17. This PR will be rejected. If it is not, please report this to the CATALYST development team. It is important that this PR be rejected because it covers the same date range as #634-17 with the same sampling strategy, production strategy, and configuration code. This type of rejection prevents the situation simultaneously running the same job twice under two PRs.
3. Attempt to submit PR #636-17. This PR should be accepted and will appear in the CATALYST Operator’s Console momentarily. In CATALYST v2.0, this PR would’ve been rejected because it shares a sampling strategy, production strategy, and configuration code with #634-17. In v2.1 this PR is accepted because CATALYST’s duplication checks now factor in the date range with the new PR as well; you will see that this PR covers the month of October 2004. If this PR is rejected, please stop and report this information to the CATALYST development team.

3.3.4 Testing [CER-211, CER-301]:

1. Note that PRs #634-17 and #636-17, which were submitted while testing (CER-253], are available in the Operator's Console.
2. Examine all the jobs in these two PRs using the Operator's Console. All of them should eventually enter the state "ready_to_launch". This means that the jobs have finished their input existence checks and are ready to hand-off to run on the cluster. If after several minutes, you do not see this occur, please stop and notify the development team.
3. In the setup step the log database items that have existed for a long time that are used by this PGE were removed. This forced the CATALYST logdb process to look directly on the DPO to determine if the required input files are available. The jobs being in the "ready_to_launch" indicates that the data was found in the DPO. Since this PGE uses many different types of data from the DPO, you have successfully tested these two tickets.
4. Do not proceed beyond this point until all the conditions mentioned above have been verified.

3.3.5 Testing [CER-53, CER-300, CER-318]

Testing the next several tickets below should occur within an hour of each other to best see the features in action:

1. Testing [CER-53]:
 - a. Since there are many items from two PRs that are "ready_to_launch" PR load balancing can be easily observed. Start the CATALYST cluster process using either the Operator's Console or the command line management utility.
 - b. In the Operator's Console, observe jobs from PRs #634-17 and #636-17 switch from "ready_to_launch" to "running". CATALYST v2.0 would've launched jobs from PR #634-17 in its entirety before launching jobs from PR #636-17.
 - c. This serves as a preliminary confirmation for this ticket. Proceed to the next step to get a better visual indication of the load balancing in action.
2. Testing [CER-300, CER-318]:
 - a. In the CATALYST Operator's Console, go to the "Tools" menu and select the "SGE Jobs" items. A new window will appear that brings up a qstat-like list of running SGE jobs. Since the load balancing from [CER-53] is in effect, observe that SYNI jobs from both September 2004 and October 2004 are running together. This list will closely look like it alternates from September, October, September, October...
 - b. This list shows all jobs running under CATALYST at any given time and was designed to be a graphical version of the output from the 'qstat' command. The number of SGE jobs running is displayed in the title of this window, and the list can be sorted by any column by clicking on the column to switch from ascending/descending sorting. This list updates itself automatically, and the window can be closed and re-opened at any time.
 - c. If you have reached this point, you have successfully tested these two tickets and you have also performed a secondary confirmation for the previous ticket [CER-53].

3.3.6 Testing [CER-299] Part 1:

1. To begin testing the job flushing capability, select “Start SGE Job Flush” from the “CATALYST Server” menu. Select “Yes” in the subsequent dialog box to confirm the intent to perform the job flushing action.
2. After a few moments the Operator’s Console’s title bar will now contain “[Flushing Jobs]”. This indicates that the SGE Job Flush mode is active. All Operator’s Console clients will see this in their window title.
3. Open the “SGE Jobs” window again; “Tools->SGE Jobs” in the Operator’s Console. Notice that there are several jobs running, but that CATALYST is no longer submitting any new jobs.
4. Continue to monitor this window over time until all SGE jobs have been completed (the list should eventually become empty). Given this PGE, CER7.2.1P2, this might take several hours (this is normal). The empty list proves that the SGE Job Flush action [CER-299] was successful.

3.3.7 Testing [CER-278, CER-316, CER-317]:

1. There are now several jobs with their science processing complete and epilogues queued because this PGE’s epilogue is disabled.
2. Navigate through the two PRs and view the two months. Notice that there are jobs “epilog_queued [##]”, where “##” is their processing order. Seeing the numbers in brackets confirms [CER-316] and the first half of [CER-278].
3. Select “Epilog Queue List” from the “Tools” menu in the Operator’s Console. A new window will appear titled “Epilog Queue”.
4. In this window there will be a list of the CATALYST jobs that are queued for epilogues to run. This list will also show which PR the jobs are from, which stream they are a member of, and the data date of the job itself. This list will automatically refresh itself while it is open. Seeing jobs from the two PRs in this list successfully confirms [CER-278] and [CER-317].
5. Enable the epilogue for PGE CER7.2.1P2. After doing so, items from this list disappear, meaning that they have finished processing.

3.3.8 Complete testing of [CER-299]:

1. Disable the SGE Job Flush mode by selecting “Stop SGE Job Flush” from the “CATALYST Server” menu in the Operator’s Console, selecting “Yes” to confirm the action. This will allow the CATALYST cluster process to start submitting more jobs to the cluster.
2. The “[Flushing Jobs]” text will disappear from the title bar. This means CATALYST has returned to its normal processing mode. All other connected clients will also see the “[Flushing Jobs]” text eventually disappear.
3. Open the “SGE Jobs” window and notice that CATALYST is now submitting new jobs to the cluster.
4. Seeing that CATALYST is submitting more jobs to the cluster confirms the feature introduced by [CER-299] is working as expected.

3.3.9 Wrapping up:

1. Allow CATALYST to complete processing the jobs in these two PRs. Expect these jobs to complete within a few days. Use the new SGE and epilogue task monitoring features as CATALYST runs these jobs to completion. Report any issues to the development team.
2. At this point, all testable JIRA tickets associated with CATALYST Server v2.1 and CATALYST Operator's Console v2.2 have been verified.

Appendix A

Acronyms and Abbreviations

API	Application Programming Interface
ASDC	Atmospheric Science Data Center
CATALYST	CERES AuTomAted job Loading sYSTem
CERES	Clouds and the Earth's Radiant Energy System
CERESlib	CERES library
DAAC	Distributed Active Archive Center
LDAP	Lightweight Directory Access Protocol
NASA	National Aeronautics and Space Administration
Perl_Lib	CERES's Perl module library
PR	Processing Request
TRMM	Tropical Rainfall Measuring Mission
XML-RPC	Extensible Markup Language – Remote Procedure Call

Appendix B

CATALYST Directory Structure Diagram

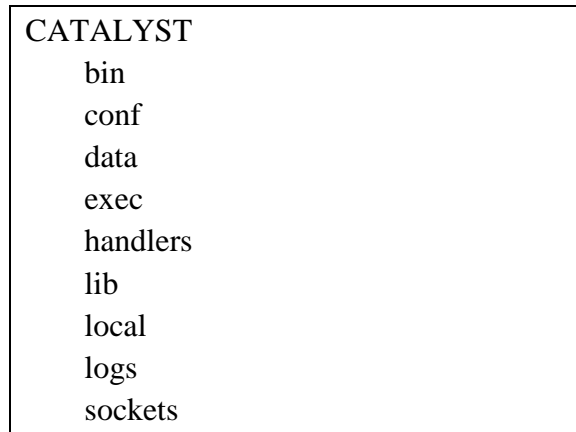


Figure B-1. CATALYST Directory Structure

Appendix C

File Description Table

C.1 Executable Scripts

Table C-1. \$CATALYST/bin directory

File Name	Format	Description
catalyst_manager.sh	bash	Text user interface utility for managing CATALYST processes.
send_email.sh	bash	Sends a CATALYST status email to the group.
send_email_if_issue.sh	bash	Sends a CATALYST status email to the group only if there is a problem.
start_catalyst.sh	bash	Start the CATALYST master process.
stop_catalyst.sh	bash	Stops the CATALYST master process and other CATALYST subprocesses.