

423-SPEC-003, Original, Revision -  
Earth Science Data and Information Systems (ESDIS)  
Project, Code 423

# **Common Metadata Repository Workload Specification**

## CMR Workload Specification Signature/Approval Page

**Prepared by:**

***Signature obtained on file***

Chris Durbin  
CMR Technical Lead  
NASA GSFC, Code 423

***03/29/2019***

Date

**Reviewed by:**

***Signature obtained on file***

Dana Shum  
Task lead - EED-2 Task 35  
NASA GSFC, Code 423

***04/10/2019***

Date

**Approved by:**

***Signature obtained on file***

Valerie Dixon  
EOSDIS CMR Manager  
NASA GSFC, Code 586

***03/15/2019***

Date

***Signature obtained on file***

Karen Michael  
EOSDIS Systems Engineer  
NASA GSFC, Code 400

***03/18/2019***

Date

***Signature obtained on file***

Andrew Mitchell  
ESDIS Project Manager  
NASA GSFC, Code 423

***04/18/2019***

Date

**[Electronic] Signatures available in B32 Room E148  
online at: / <https://ops1-cm.ems.eosdis.nasa.gov/cm2/>**

## **Preface**

This document is under ESDIS Project configuration control. Once this document is approved, ESDIS approved changes are handled in accordance with Class I and Class II change control requirements described in the ESDIS Configuration Management Procedures. Changes to this document shall be verified by a document change notice (DCN) and implemented by change bars or by complete revision.

Any questions should be addressed to: [esdis-esmo-cmo@lists.nasa.gov](mailto:esdis-esmo-cmo@lists.nasa.gov)

ESDIS Configuration Management Office (CMO)

NASA/GSFC

Code 423

Greenbelt, Md. 20771

## **Abstract**

A guide for load testing the Common Metadata Repository (CMR), intended to describe starting conditions, testing parameters, and reporting requirements for these tests, but not to prescribe implementation or technology strategies.

*Keywords: CMR, Workload*



## **Table of Contents**

### **List of Figures**

---

---

---

# 1 INTRODUCTION

## 1.1 Purpose/Scope

This document is intended to guide load testing of the Common Metadata Repository (CMR).

## 1.2 Related Documentation

The latest versions of all documents below should be used. The latest ESDIS Project documents can be obtained from URL: <https://ops1-cm.ems.eosdis.nasa.gov>. ESDIS documents have a document number starting with either 423 or 505. Other documents are available for reference in the ESDIS project library website at: [http://esdisfmp01.gsfc.nasa.gov/esdis\\_lib/default.php](http://esdisfmp01.gsfc.nasa.gov/esdis_lib/default.php) unless indicated otherwise.

### 1.2.1 Reference Documents

The following documents are not binding on the content but referenced herein and, amplify or clarify the information presented in this document.

423-RQMT-003	Metadata Requirements Base Reference Document
--------------	---

---

---

## 2 OVERVIEW

This document is intended to describe starting conditions, testing parameters, and reporting requirements for these tests, but is not intended to prescribe implementation or technology strategies.

### 2.1 Defining Sub-Second Search

The CMR is being designed to provide sub-second searching, and as such the definition of sub-second search should be clarified. Based on initial prototype investigations outline the qualities of sub-second search:

*Sub-second search implies that average search time is < 1s for both collection and granule level searches as measured from the time the CMR receives a query until it has completed streaming the response.*

*Sub-second performance will be provided for queries that have the following:*

- *Granule queries that include a collection identifier such that it identifies a single collection. That could be ECHO Collection Id or provider id with dataset id or short name and version id.*
- *Number of results requested is 100 or fewer.*
- *Requested format is either a strict reference or the format given by the provider or the format has been pre-computed and cached.*
- *Any of the granule conditions including temporal and spatial will be supported with sub-second search.*
- *Any number of conditions will be supported.*

### 2.2 Environmental Considerations

From an environmental viewpoint, both the hardware and software employed in the workload need to reflect the operational buildout. This means that the following items need to be kept in sync between systems:

- Hardware (or VM) counts and specifications: database nodes, elastic nodes, etc.
- Network configuration/capacity/security OS
- Versions and Patching Levels
- Software Platform and Versions: applications containers and runtime environments such as the JVM and component application servers
- 3rd Party Library Versions and Patching: any libraries and software packages (e.g. Oracle, ElasticSearch, Quartz, etc) used by the CMR

If there are any discrepancies among these items, any reports resulting from workload tests need to document these discrepancies.

---

---

## 2.3 Search and Retrieval Testing Criteria

This section outlines the expectations of the CMR under various search loads. As stated earlier, the CMR is expected to sustain sub-second average search times under normal system load and continue to provide acceptable performance under a stress test. The tests outlined here will simulate these load environments and establish criteria for determining success. There are two tests included in this document:

- Sustained Load Testing (AKA 24 Hour Load Test)
- Spike Load Testing

### 2.3.1 Initial System State

The system should be preloaded with a mirror of the operational search index and database prior to the execution of the test. This mirror will be periodically updated as described in the "Frequency of Refresh".

### 2.3.2 Search Load Characteristics

#### 2.3.2.1 Types of Searches

Load on the system should represent actual system load, meaning that query parameters should mimic searches performed on the production system. The workload system should use actual operational queries to generate search requests.

This load includes a mix of searches to include various combinations of spatial, temporal and keyword searches at both the collection and granule level as well as direct collection and granule retrieval using concept-ids. In addition, queries may be returned in various supported formats (e.g. json, ECHO10, ISO 19115, DIF)

#### 2.3.2.2 Concurrent Requests

The number of concurrent requests in the system should be reflective of concurrent requests from the operational system rather than evenly distributed throughout the run.

#### 2.3.2.3 Access Control Lists and Load

For the purposes of these tests it is to be assumed that the token(s) used to generate load have the same Access Control Lists (ACLs) as a normal registered User Registration System (URS) Earthdata user accessing the CMR. Administrators and non-authenticated guest users should not be used for generating search load. Future versions of the workload system should work to accommodate a more representative mix of user and access levels.

### 2.3.3 Sustained Load Testing

This test involves executing a predetermined number of queries on the system and reporting statistics based on those queries. Sustained load

---

---

should reflect 125% of the most active search day over the past three months or the highest target established since the outset of load testing. This load target can be determined via log queries and should be updated quarterly as needed.

#### 2.3.3.1 Establishing a Baseline for Target Load

Use Splunk logs to find the busiest single day within the past three months. If the current workload target is higher than any day within the past three months then keep the same target. Otherwise target 25% more granule and collection queries from the busiest single day.

#### 2.3.4 Spike Load Testing

(Note: This has not been implemented in CMR at this point, only sustained load tests are currently supported)

If the intention of the Sustained Load Test is to ensure sub-second performance, the intention of Spike Load Testing is to push that performance to its limits and observe system performance under times of extreme stress.

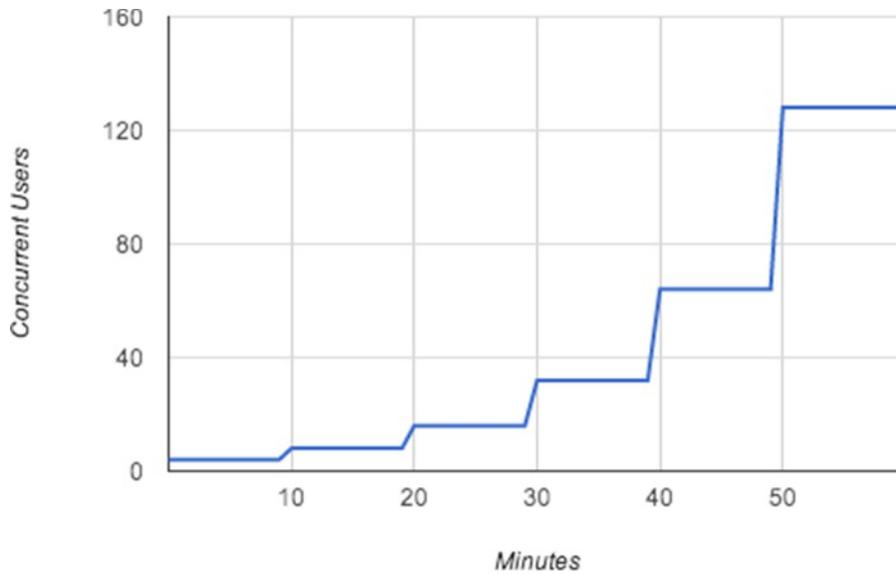
##### 2.3.4.1 Load Ramp Up

This test will begin with a small number of concurrent workers running at sustained load as described above. The number of concurrent users will then be periodically increased over set intervals. These the number and spacing of timing intervals should be configurable, as should the initial number of concurrent users.

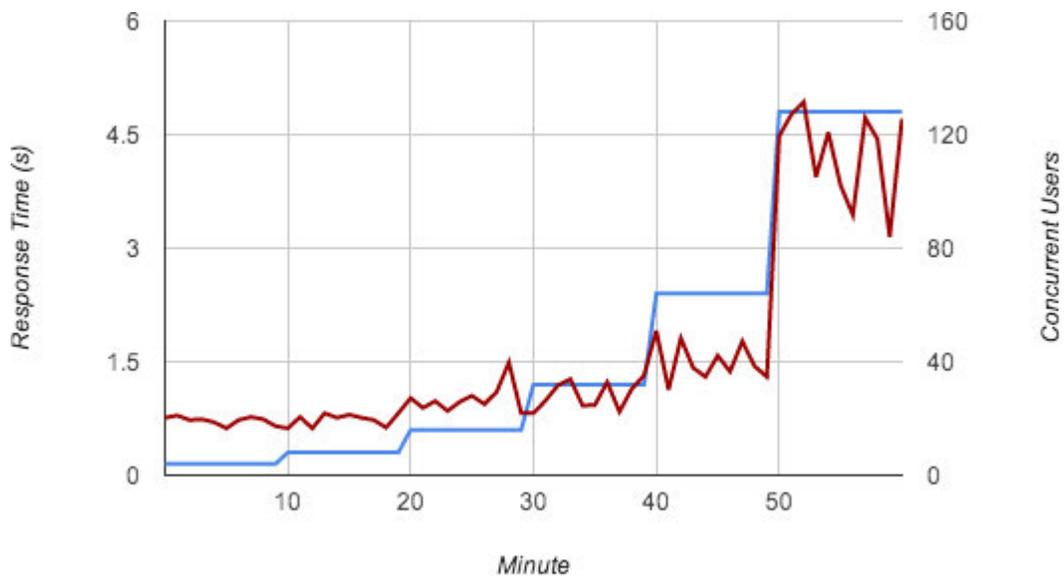
#### Example

(Note: These data are not based on any actual metrics gathered via CMR load testing and are randomized numbers being used to illustrate principles of the spike load test.)

Figure 2 11 shows an example ramp up, starting with 4 users for the first 10 minutes of the test and then doubling in number for each successive 10 minute interval. This data is collected over 6 intervals. The next figure, Figure 2 2 depicts system response time overlaid on the concurrent users line. This example data shows that system performance degrades significantly as the number of users increases beyond 32 concurrent users.



**Figure 21. Example Spike Load Test Ramp Up**



**Figure 22. Example Spike Load Test Ramp Up**

### 2.3.5 Post-Test Reporting and Pass/Fail Criteria

The intent of the workload performance tests is to ensure no degradation as code changes are deployed each release. As such the performance of each workload run will be compared to the original query performance of those same queries in production. Analysis of the query performance will be based on the following metrics:

- Average query time
- Query time standard deviation
- 95th percentile

- 
- 99th percentile

If any of the metrics are more than 20% degraded when compared to operational performance the test will be considered failed. Further investigation as to the cause of the performance degradation will be determined by detailed investigation of the logs.

The following dashboard will be used to report workload run metrics:

[https://logs.earthdata.nasa.gov/en-US/app/search/cm\\_r\\_ngap\\_wl\\_search\\_dashboard](https://logs.earthdata.nasa.gov/en-US/app/search/cm_r_ngap_wl_search_dashboard)

### 2.3.6 Frequency of Baseline Refresh

Once the baseline performance for each criterion outlined in this document has been established, it is expected that the environment used to execute search load testing is refreshed and mirrored from the operational system at least 4 times annually. This ensures that the testing is kept in step with production with respect to inventory size, system load, as well as hardware and software versions.

## 2.4 Ingest Load and Index Criteria

The current ingest workload target is 900K granules ingested over 24 hours. The 900K target is well above the sustained ingest rate from all missions from all providers forward processing. We test with additional ingest load to account for reprocessing campaigns that occur for each provider. We've seen cases where a provider needs to resend all of their metadata (more than 100 million granules for some providers) and during these times we've observed ingest rates at more than 7 million a day. In the future we should determine what our ingest target load should be.

## 2.5 Ordering/Services Criteria

This document will be updated at a later time to reflect workload specification for ordering and service invocation performance.

## 2.6 References

Performance Testing Guidance for Web Applications

{+} <http://msdn.microsoft.com/en-us/library/bb924375.aspx> Load Testing

---

## Appendix 6.1.1.1.1.A Abbreviations and Acronyms

ACL	Access Control Lists
AKA	Also Known As
CCB	Configuration Change Board
CCR	Configuration Change Request
CMR	Common Metadata Repository
DCN	Document Change Notice
DIF	Directory Interchange Format
ECHO	Earth Observing System (EOS) Clearinghouse
ESDIS	Earth Science Data Information System
GSFC	Goddard Space Flight Center
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
JVM	Java Virtual Machine
K	thousand
OS	Operating System
TBD	To be determined
URL	Uniform Resource Locator
URS	User Registration Service
VM	Virtual Machine