Granule Queries Now Require Collection Identifiers

On April 4, 2018, we will push a change to CMR operations to require that all granule queries include some kind of collection identifier in the search. Queries that do not specify any field to limit to a subset of collections will be rejected with the message, "The CMR does not currently allow querying across granules in all collections. To help optimize your search, you should limit your query using conditions that identify one or more collections, such as provider, concept_id, short_name, or entry_title." This change was made in order to prevent any single query from using too many resources in the CMR. Any one of the following parameters will satisfy this condition:

- provider
- concept_id (either granule or collection or the alias echo_collection_id)
- short_name
- version
- entry_id
- entry_title or its alias dataset_id

More details on why the change is being made

The problem with unrestricted granule queries is that the CMR needs to search against all granule indexes instead of just the granule index for that collection.

In operations today we have about 1350 unique shards across 265 granule indexes. Consider the following three queries:

1) /search/granules?producer_granule_id=MOD09GQ.A2015220.h00v08.006.2015303173444.hdf
2) /search/granules?producer_granule_id=MOD09GQ.A2015220.h00v08.006.2015303173444.hdf&provider_id=LPDAAC_ECS
3) /search/granules?producer_granule_id=MOD09GQ.A2015220.h00v08.006.2015303173444.hdf&short_name=MOD09GQ

All of the above queries achieve the same result; they return a single granule with the provided producer_granule_id. However, in Elasticsearch the resource profile is drastically different for each of those queries.

#1 requires 1350 searches because it needs to search against every shard.
#2 requires 325 searches because there are 61 LPDAAC_ECS collection indexes which each have 5 unique shards plus the small collections index which has 20 unique shards.
#3 requires 5 searches because that collection has its own index with 5 unique shards.

Providing a collection identifier drops the query load on Elasticsearch by a factor of 270 in this case.