selective SOQL & non-selective SOQL Queries

MARCH 31, 2018 / PAVAN

Your SOQL query sometimes returns so many sObjects that the limit on heap size is exceeded and an error occurs. To resolve, use a SOQL query for loop instead, since it can process multiple batches of records by using internal calls to query and queryMore.

```
// Use this format if you are not executing DML statements
// within the for loop
for (Account a : [SELECT Id, Name FROM Account
                  WHERE Name LIKE 'Acme%']) {
    // Your code without DML statements here
}

// Use this format for efficiency if you are executing DML statements
// within the for loop
for (List<Account> accts : [SELECT Id, Name FROM Account
                            WHERE Name LIKE 'Acme%']) {
    // Your code here
    update accts;
}
```

Instead of using a SOQL query in a for loop, the preferred method of mass updating records is to use batch Apex, which minimizes the risk of hitting governor limits.

For more information, see SOQL For Loops.

More Efficient SOQL Queries

For best performance, SOQL queries must be selective, particularly for queries inside triggers. To avoid long execution times, the system can terminate nonselective SOQL queries. Developers receive an error message when a non-selective query in a trigger executes against an object that contains more than 200,000 records. To avoid this error, ensure that the query is selective.

Selective SOQL Query Criteria

- A query is selective when one of the query filters is on an indexed field and the query filter reduces the resulting number of rows below a system-defined threshold. The performance of the SOQL query improves when two or more filters used in the WHERE clause meet the mentioned conditions.
- The selectivity threshold is 10% of the first million records and less than 5% of the records after the first million records, up to a maximum of 333,333 records. In some circumstances, for example with a query filter that is an indexed standard field, the threshold can be higher. Also, the selectivity threshold is subject to change.

Custom Index Considerations for Selective SOQL Queries

- The following fields are indexed by default.
  - Primary keys (Id, Name, and OwnerId fields)
  - Foreign keys (lookup or master-detail relationship fields)
  - Audit dates (CreatedDate and SystemModstamp fields)
  - RecordType fields (indexed for all standard objects that feature them)
  - Custom fields that are marked as External ID or Unique
- When the Salesforce optimizer recognizes that an index can improve performance for frequently run queries, fields that aren’t indexed by default are automatically indexed.
- Salesforce Support can add custom indexes on request for customers.
- A custom index can’t be created on these types of fields: multi-select picklists, currency fields in a multicurrency organization, long text fields, some formula fields, and binary fields (fields of type blob, file, or encrypted text.) New data types, typically complex ones, are periodically added to Salesforce, and fields of these types don’t always allow custom indexing.
- You can’t create custom indexes on formula fields that include invocations of the TEXT function on picklist fields.
- Typically, a custom index isn’t used in these cases.
  - The queried values exceed the system-defined threshold.
  - The filter operator is a negative operator such as NOT EQUAL TO (or !=), NOT CONTAINS, and NOT STARTS WITH.
  - The CONTAINS operator is used in the filter, and the number of rows to be scanned exceeds 333,333. The CONTAINS operator requires a full scan of the index. This threshold is subject to change.
  - You’re comparing with an empty value (Name != ”).

    However, there are other complex scenarios in which custom indexes can’t be used. Contact your Salesforce representative if your scenario isn’t covered by these cases or if you need further assistance with non-selective queries.

Examples of Selective SOQL Queries

To better understand whether a query on a large object is selective or not, let’s analyze some queries. For these queries, assume that there are more than 200,000 records for the Account sObject. These records include soft-deleted records, that is, deleted records that are still in the Recycle Bin.

Query 1:

```
1    SELECT Id FROM Account WHERE Id IN (<list of account IDs>)
```

The WHERE clause is on an indexed field (Id). If SELECT COUNT() FROM Account WHERE Id IN (<list of account IDs>)returns fewer records than the selectivity threshold, the index on Id is used. This index is typically used when the list of IDs contains only a few records.

Query 2:

```
1    SELECT Id FROM Account WHERE Name != ''
```

Since Account is a large object even though Name is indexed (primary key), this filter returns most of the records, making the query non-selective.

Query 3:

```
SELECT Id FROM Account WHERE Name != '' AND CustomField__c = 'ValueA'
```

Here we have to see if each filter, when considered individually, is selective. As we saw in the previous example, the first filter isn’t selective. So let’s focus on the second one. If the count of records returned by SELECT COUNT() FROM Account WHERE CustomField__c = ‘ValueA’ is lower than the selectivity threshold, and CustomField__c is indexed, the query is selective.