Designing, Testing & Tuning Objectstore Queries

This document describes best practice for designing queries within FileNet Content Manager

Version: 1.1
Dominik Baer dominik.baer@ch.ibm.com

Contents

1 Document information

1.1 Open points

No table of contents entries found.

1.2 Decisions taken

No table of contents entries found.

1.3 Versions

Version	Changes / Status	Author(s)	Date
1.0	Draft	D.Baer	2.7.2018
1.1	Updated and reformatted	D.Baer	10.3.2021

2 Scope of this document

2.1 Querying meta data

IBM FileNet Content Manager / Case Manager supports searching on metadata either by using:

- content based queries (Fulltext)
- relational queries (DBMS)

2.1.1 Fulltext searches– out of scope

This document doesn't discuss the approach using full text searches, even it might be a good alternative addressing the burden to build numerous database indices to address various search strategies.
For information about building full text based searches see:
CBR queries
Reasons to use content-based queries instead of relational queries are:

- case insensitivity
- highly customized searches with constantly changing search arguments

Reasons not going this approach might include the additional need of maintaining CSS Servers and the need to maintain additional storage for the Full text index.

2.1.2 DBMS searches – in scope

2.1.2.1 Database vendors

An Objectstore is always created within a supported database system like

- MS SQL Server,
- Oracle
- DB2
- PostresSQL

Searches are either been formulated against the content subsystem or the workflow system.

You find further information regarding the preparation steps on the database side
https://www.ibm.com/support/knowledgecenter/en/SSNW2F_5.5.0/com.ibm.p8.planprepare.doc/p8ppi084.htm

Please check the compatibility matrix at
https://www.ibm.com/software/reports/compatibility/clarity/softwareReqsForProduct.html

Most of the discussion in this document is aimed at customers using Oracle. Most techniques can be applied as well on the other database vendors.

2.2 DBMS

2.2.1 Case sensitivity

The case sensitivity depends on the database vendor and version. You find most relevant information googling for “Collation” and your database system vendor.

MS SQL: https://docs.microsoft.com/en-us/sql/relational-databases/collations/collation-and-unicode-support?view=sql-server-ver15

Oracle: https://docs.oracle.com/database/121/NLSPG/ch5lingsort.htm#NLSPG005

DB2: https://www.ibm.com/support/knowledgecenter/SSEPGG_11.5.0/com.ibm.db2.luw.admin.nls.doc/doc/r0050489.html

PostgresSQL: https://www.postgresql.org/docs/13/collation.html

In general MS SQL come out of the box with a case insensitive collation and is not as problematic as DB2 and Oracle. DB2 has introduced case insensitive collations a few years ago and finally Oracle comes with the same mechanism and ever more detailed variants:

2.2.2 Oracle

Oracle provides today very differentiated options to configure the sort and filter behavior with different mechanisms:

1. Functional index
2. Changed behavior per session
3. Case insensitive index
4. Case insensitive database collation

2.2.3 FileNet Content Manager

1. FileNet Content manager can either work with the setings on the database or
2. has an option which is called “enforce case insensitive searches” which translates queries in a way, that the searchcondition is slightly changed and a search clause is added of the type lower(search_argument) = lower(string_value)

If you are selecting option 2 of above, you are responsible to ensure the relevant functional index is available on the table. Typically, these indices should be created accordingly to following pattern including the search argument and the object_id:

```
lower(uxy_propertytemplate, object_id)
```

This allows the optimizer to choose this index instead of traversing the whole table by a full table scan.

1) Since 10gR2, Oracle allows to fine-tune the behavior of string comparisons by setting the NLS_COMP and NLS_SORT session parameters: (changed session behavior)

```
SQL> SET HEADING OFF
SQL> SELECT *
  2  FROM NLS_SESSION_PARAMETERS
  3  WHERE PARAMETER IN ('NLS_COMP', 'NLS_SORT');

NLS_SORT
BINARY

NLS_COMP
BINARY

SQL>
SQL> SELECT CASE WHEN 'abc'='ABC' THEN 1 ELSE 0 END AS GOT_MATCH
  2  FROM DUAL;

         0
SQL>
SQL> ALTER SESSION SET NLS_COMP=LINGUISTIC;

Session altered.

SQL> ALTER SESSION SET NLS_SORT=BINARY_CI;

Session altered.

SQL>
SQL> SELECT *
  2  FROM NLS_SESSION_PARAMETERS
  3  WHERE PARAMETER IN ('NLS_COMP', 'NLS_SORT');

NLS_SORT
BINARY_CI

NLS_COMP
LINGUISTIC


SQL>
SQL> SELECT CASE WHEN 'abc'='ABC' THEN 1 ELSE 0 END AS GOT_MATCH
  2  FROM DUAL;
         1
```

So this could be an approach which could simplify the documented and supported approach in the future:

```
create or replace

trigger set_nls_onlogon

AFTER LOGON ON SCHEMA

DECLARE

BEGIN

  EXECUTE IMMEDIATE 'ALTER SESSION SET NLS_SORT=''BINARY_CI''';

  EXECUTE IMMEDIATE 'ALTER SESSION SET NLS_COMP=''LINGUISTIC''';
END set_nls_logon;
```

There might be as well possibilities to overwrite the session default behavior on the database pool definition of the application server.

2) You can also create case insensitive indexes:

```
create index
   nlsci1_gen_person
on
   MY_PERSON
   (NLSSORT
      (PERSON_LAST_NAME, 'NLS_SORT=BINARY_CI')
   )
;
```

3) From later Oracle versions a case insensitive collation is available:
https://oracle-base.com/articles/12c/column-level-collation-and-case-insensitive-database-12cr2

Summary/Recommendations:
If you chose a strategy which is not documented on the IBM knowledge center, please ensure that you test your approach and that you contact your local IBM contact and ask for a blessing of your approach by the support team. This has been done in the past several times.

The case insensitive collation approach has been documented and supported on DB2
The functional indices have been documented and supported for DB2 and Oracle.

2.3 Objectstore (Content part - non workflow part)

At the creation of an Objectstore database table, indices, sequences and views are generated.
It would be a good idea to keep the DDL of a fresh created Objectstore at a safe place to be able to compare at a later time.
Over time it gets difficult to understand whether the product, a fixpack, a performance & tuning exercise or a manual dba intervention have changed database indices.
There are some recommendations around on various levels: (product, fixpacks, p&t)

Create performance enhancing indexes	https://www.ibm.com/support/knowledgecenter/SSNW2F_5.5.0/com.ibm.p8.performance.doc/p8ppt152.htm	
Indexing for IBM FileNet P8 Content Engine Searches	http://www-01.ibm.com/support/docview.wss?uid=swg21502886&wv=1	
Managing workflow indexes	https://www.ibm.com/support/knowledgecenter/en/SSNW2F_5.5.0/com.ibm.p8.ce.admin.tasks.doc/p8pcc301.htm	
DBCREATEPEVIEWS workflow system property	https://www.ibm.com/support/knowledgecenter/SSNW2F_5.5.0/com.ibm.p8.ce.admin.tasks.doc/bpfad034.htm	
Performance tuning for IBM Content Navigator	https://www.ibm.com/support/knowledgecenter/SSEUEX_3.0.8/com.ibm.installingeuc.doc/eucpt007.htm	
Setting the log size for Oracle case history databases	https://www.ibm.com/support/knowledgecenter/SSCTJ4_5.3.3/com.ibm.casemgmt.design.doc/acmtn205.htm	

2.3.1 Report existing indices per Objectstore

Following SQL Script helps to get a good understanding of the current available indices:

```
select
B.table_name, B.index_name, B.col, A.column_expression
from all_ind_expressions A,
(
select table_name , index_name, listagg(column_name, ',')
within group (order by column_position) as col
from all_ind_columns
where index_name not like 'BIN$%'
--and table_name like 'DOCVERSION'
group by table_name , index_name
order by table_name , index_name
)
B
where A.index_name (+) = B.index_name
```

DOCVERSION	AI_UDEC8_ESTV_PARTNERID_D	SYS_NC00179$	LOWER("UDEC8_ESTV_PARTNERID")
DOCVERSION	BER_HOME_ID_D	HOME_ID	
DOCVERSION	BER_RECOVERY_ITEM_ID_D	RECOVERY_ITEM_ID	
DOCVERSION	DOCTITLE_IDX	SYS_NC00173$,OBJECT_ID	LOWER("U1708_DOCUMENTTITLE")
DOCVERSION	I_DOCVERSION_DOSSIERNR	U39C6_EFIMD_DOSSIERNR,OBJECT_ID	
DOCVERSION	I_DOCVERSION_ISDOCID	UAC84_EFIMD_ISDOCID,OBJECT_ID	
DOCVERSION	I_DOCVERSION_MIMETYPE	MIME_TYPE	
DOCVERSION	I_DOCVERSION_OBJECT_CLASS_ID	OBJECT_CLASS_ID	
DOCVERSION	I_DOCVERSION_OBJ_VERS_SER_ID	OBJECT_ID,VERSION_SERIES_ID	
DOCVERSION	I_DOCVERSION22	VERSION_SERIES_ID,SYS_NC00153$,SYS_NC00154$	"MAJOR_VERSION_NUMBER"
DOCVERSION	I_DOCVERSION22	VERSION_SERIES_ID,SYS_NC00153$,SYS_NC00154$	"MINOR_VERSION_NUMBER"
DOCVERSION	I_DOCVERSION73	SECURITY_FOLDER_ID	
DOCVERSION	IDX$$_BARCODE	SYS_NC00155$,OBJECT_CLASS_ID,HOME_ID,RECOVER	LOWER("UE5E8_EFIMD_BARCODE")
DOCVERSION	IDX_CMP	OBJECT_CLASS_ID,SYS_NC00179$,OBJECT_ID	LOWER("UDEC8_ESTV_PARTNERID")
DOCVERSION	IDX_STA_1	UF176_EFIMD_FORMID,OBJECT_CLASS_ID,UB926_SOURCESYSTEM,HOME_ID,RECOVERY_ITEM_ID,CREATE_DATE	
DOCVERSION	IDX_STA_2	CREATE_DATE	
DOCVERSION	IDX_STA_3	U4838_PAGINATOR	
DOCVERSION	IDX$$_14070001	IS_CURRENT,OBJECT_CLASS_ID,HOME_ID,RECOVERY_ITEM_ID	
DOCVERSION	IPDOS_U3988_ESTV_DOK_GRUPPE_D	SYS_NC00181$	LOWER("U3988_ESTV_DOKUMENTENGRUPPE")
DOCVERSION	IPDOS_U44F8_ESTV_DOCTYPE_D	SYS_NC00180$	LOWER("U44F8_ESTV_DOCTYPE")
DOCVERSION	SYS_C008148	OBJECT_ID	
DOCVERSION	UI_U69D7_CMACMASSOCIATEDCASE	U69D7_CMACMASSOCIATEDCASE	
DOCVERSION	U8616_EFIMD_BPRID	U8616_EFIMD_BPRID	

FIGURE 1: TABLE SHOWS OUTPUT FROM THE ABOVE SQL

The above report shows the table name, the index name and the order of attributes in the index. For functional indices you see a cryptic name instead of the expression. But you get the expression in the column right to it.

From a perspective of the application, it would be great to understand which type of class has which property templates exposed and how do they be called from the perspective of the database:

```
select g.symbolic_name , c.column_name, c.dbg_table_name, c.column_datatype,
       from columndefinition c
       , globalpropertydef g
where c.prop_id = g.object_id
```

	A	B	C	D
1	SYMBOLIC_NAME	COLUMN_NAME	DBG_TABLE_NAME	COLUMN_DATAT
151	Comment	ue638_comment	DocVersion	8
152	CommunicationType	ue1d6_communicationtype	DocVersion	6
153	ComponentBindingLabel	u2e28_componentbindinglabel	DocVersion	8
154	ContainerType	u2618_containertype	Container	8
155	CustomObjectType	u26e8_customobjecttype	Generic	8
156	Description	ue0c8_description	Container	8
157	Description	ue4b8_description	DocVersion	8
158	Description	u4258_description	Event	8
159	Description	u3628_description	Generic	8
160	Description	u5c58_description	Link	8
161	DisableEH	u3fa6_disableeh	DocVersion	6
162	DocGroup	u53e6_docgroup	DocVersion	6
163	DocState	u10a6_docstate	DocVersion	6
164	DocType	ubb16_doctype	DocVersion	6
165	DocTypeVersion	u9396_doctypeversion	DocVersion	6
166	DocumentLanguage	u5446_documentlanguage	DocVersion	6
167	DocumentTitle	uc1f8_documenttitle	Annotation	8

FIGURE 2: SEARCHING A LIST OF SYMBOLIC_NAMES, IT'S COLUMN NAME AND THE RELATED DATABASE TABLE

The report above shows multiple attributes (property_templates) which are used in different contexts.
Document Classes are mapped into the table "DocVersion", Folders and Case Folders are mapped in "Container", Custom Objects in "Generic", RootClasses in "UT_..." etc.

Regardless whether you have one or multiple document classes, all are sharing a huge database table "Docversion".
Saying this, the table may contain many attributes and for search reasons you need multiple indices.
So it makes sense to minimize the number of database indices if possible to reduce the overhead.

2.4 Objectstore - Workflow System part

With the usage of Workflows and in particular with Case Manager the combination of both area (content and workflow) must be tuned as well
The workflow system follows a bit a different approach.

How to gather the mapping between Roster, Queues and EventLogs and the Database Views: Use vwtool (see
https://www.ibm.com/support/knowledgecenter/en/SSNW2F_5.5.0/com.ibm.p8.pe.vw.doc/bpfvl051.htm

For tracing the JDBC calls against the VW-database objects, use as well vwtool
https://www.ibm.com/support/knowledgecenter/en/SSNW2F_5.5.0/com.ibm.p8.pe.vw.doc/bpfvl049.htm

Switch on:

Database access	This option outputs the SQL statements that are used by workflow system, along with the values of the substitution variables. Setting this option automatically sets the database time option.
Database outputs	Trace the field values and outputs of database statements, such as SELECT, UPDATE, and so on.
Database time	This prompt appears only if you answer no to the database access prompt.

Typically, the workflow system out of the box is already set reasonably well leveraging database indexing. You should mainly check whether you are applying heavily filters and try to avoid designs where one queue is holding millions of work items.

There is no way to run case insensitive searches using the Workflow system other than what has been configured on the database side. Typically, a Workflow System is using status meta data and is typically not really sensitive to cases. There is no such feature like on the content system where the queries would apply a lower() function.

If you are considering using the workflow subsystem, maybe case management then think about an approach which makes the database case insensitive!