

Ablauf von „startdupindmass“ beim Erstaufbau der Strukturierten Titeldaten; sisis-DB; DBS=syb; BVBX8:

(18.05.2005)

Information von SISIS nach Rückfrage wegen Protokollierungsstatus:

...

die Programme brauchen das Logging schon, ansonsten können diese nicht ausgeführt werden.

Sie dürfen bei Informix nicht das Logging komplett ausschalten, sondern müssen nach /dev/null loggen, d.h. die logs werden mitgeschrieben und anschließend gleich gelöscht.

Das ist in der onconfig einzutragen und der Informix-Server muß neu gestartet werden.

Für sybase muss trunc. on chkpt true eingeschaltet werden.

Außerdem ist das startdupindmass NICHT im laufenden Betrieb zu starten. Alle Anwendungen (insbesondere solche über die Katalogdaten geändert werden können (z.B. CATSERVER, SIKOM) müssen gestoppt werden.

Aus Sicherheitsgründen (da das Datenbank-Logging ja deaktiviert wird) sollten generell alle Anwendungen, insbesondere auch die INDEX-Prozesse, gestoppt werden.

Einstellen der zu verwendenden Felder:

Standardmässig werden von SISIS [folgende Kategorien](#) als Dup-Daten vorgeschlagen (Auszug aus \$SISISHOME/sql_syb/sikis/titel_dupdaten.sql):

```
create table titel_dupdaten (  
  katkey INTEGER NULL , /* Katalogschluessel */  
  verfasser VARCHAR (101) NULL , /* Kateg 100: Verfasser */  
  autor_avs VARCHAR (161) NULL , /* Kateg 100 200: Verf Urh  
aufbereitet */  
  urheber VARCHAR (101) NULL , /* Kateg 200: */  
  titel VARCHAR (101) NULL , /* Kateg 331: */  
  titel_avs VARCHAR (161) NULL , /* Kateg 331: Titel  
aufbereitet */  
  zusatz VARCHAR (91) NULL , /* Kateg 335: */  
  bandangabe VARCHAR (41) NULL , /* Kateg 089: */
```

```

verlagsort VARCHAR (21) NULL , /* Kateg 410: */
verlag VARCHAR (41) NULL , /* Kateg 412: */
erschjahr VARCHAR (11) NULL , /* Kateg 425: */
isbn VARCHAR (14) NULL , /* Kateg 540: */
issn VARCHAR (9) NULL , /* Kateg 543: */
preis VARCHAR (11) NULL , /* Kateg 542: */
datumaufn datetime NULL , /* Kateg 002: */
datumaend datetime NULL , /* Kateg 003: */
schlagwort VARCHAR (41) NULL , /* Kateg 902: */

```

Sollte diese Auswahl nicht ausreichend sein, dann können vor Ort noch bis zu zehn Kategorien eingetragen werden. Dies geschieht über die Hilfs-Tabelle „prof_dupdaten“. Hier werden von Hand noch gewünschte Kategorien aufgenommen und dann in die „titel_dupdaten“ geladen.

Hierzu ist es notwendig diese prof_dupdaten zu füllen per:

```
insert into prof_dupdaten (setnr, <kategorie>) values (1,<wert>)
```

Beispiel:

```
insert into prof_dupdaten (setnr, kateg) values (1,304)
```

⇒ dann wird die Kategorie 304 als zusätzliche Kategorie für die Ablage der Daten in strukturierter Form herangezogen.

Oder man erzeugt eine Ladedatei, welche man per load (Informix) oder bcp (Sybase) in die Tabelle einliest.

Danach noch mal die Tabelle titel_dupdaten neu anlegen per sql-Tabellengenerierung (Damit die zusätzlichen Felder auch übernommen werden):

```
isql -Usisis -Pxxxxxx -Dsisis
```

```
1> drop table titel_dupdaten
```

```
2> go
```

```
1> exit
```

```
$ cd $SISISHOME/sql_syb/sikis
```

```
$ isql -Usisis -Pxxxxxx -Dsisis < titel_dupdaten.sql
```

Danach das Skript **startdupindmass** gem. untenstehender Anleitung starten.

Es empfiehlt sich, vorher bereits die Datenmenge abzufragen, um dann den laufenden Zähler besser einschätzen zu können.

```
1> select count (*) from titel_daten where aktion!=4
2> go
```

```
-----
      139953
```

(1 row affected)

```
$ dbinfo
```

```
name          db_size      owner          dbid
  created
  status
```

```
-----
-----
-----
sisis          2150.0 MB sa          5
  Nov 07, 2000
  select into/bulkcopy/pllsort, Auto-Identity
```

(1 row affected)

```
device_fragments      size          usage          free kbytes
-----
logdbs                 200.0 MB      Nur Log        203360
rootdbs                1950.0 MB     Nur Daten      632256
```

(return status = 0)

```
$
```

```
$ startdupindmass
```

```
04.06.02 14:27:27:
      Dupdaten-Tabelle angelegt und leer?
```

```
04.06.02 14:27:27:
      Entladen der Katalogschlüssel in die Datei
/tmp/dupindmass.keys...
```

```
04.06.02 14:27:42:
      Merken des Tablespace der Indices der Dupdaten-
Tabelle
```

```
04.06.02 14:27:43:
      Indexe sind in default
```

```
04.06.02 14:27:44:
      Coredaten in den Systemparametern auf 'J' setzen
```

```
04.06.02 14:27:44:
      DUPINDMASS laedt mit /tmp/dupindmass.keys ...
Aufbau der Titeldaten in strukturierter Form läuft
Start : 04.06.2002 , 14:27
Bearbeitete Saetze:      139950
Es wurden <139953> Sätze bearbeitet
```

```

04.06.02 16:33:20:
    Indices der Duplicate-Daten wiederherstellen ...

04.06.02 16:33:36:
    Update statistics ...

04.06.02 16:33:47:
    Das Laden der Duplicate-Daten in titel_dupdaten ist
beendet

```

Bitte beachten Sie die Datei startdupindmass.prot.25689

In /var/mail/sisis liegen Mails für Sie vor

\$ more startdupindmass.prot.25689

```

04.06.02 14:27:27:
    Dupdaten-Tabelle angelegt und leer?

04.06.02 14:27:27:
    Entladen der Katalogschlüssel in die Datei
/tmp/dupindmass.keys...

04.06.02 14:27:42:
    Merken des Tablespace der Indices der Dupdaten-
Tabelle

04.06.02 14:27:43:
    Indexe sind in default

04.06.02 14:27:44:
    Coredaten in den Systemparametern auf 'J' setzen

04.06.02 14:27:44:
    DUPINDMASS laedt mit /tmp/dupindmass.keys ...

```

Es wurden <139953> Sätze bearbeitet

```

04.06.02 16:33:20:
    Indices der Duplicate-Daten wiederherstellen ...

04.06.02 16:33:36:
    Update statistics ...

04.06.02 16:33:47:
    Das Laden der Duplicate-Daten in titel_dupdaten ist
beendet

```

\$

\$ dbinfo

name	created	db_size	owner	dbid
sisis	Nov 07, 2000	2150.0 MB	sa	5
select into/bulkcopy/pllsort, Auto-Identity				

(1 row affected)

device_fragments	size	usage	free kbytes
logdbs	200.0 MB	Nur Log	187584

```
rootdbs                1950.0 MB    Nur Daten                585392
(return status = 0)
$
```

```
1> sp_spaceused titel_dupdaten
```

```
2> go
```

name	rowtotal	reserved	data	index_size	unused
titel_dupdaten	139953	46718 KB	42564 KB	4112 KB	42 KB

```
(1 row affected)
(return status = 0)
```

→ Die Tabelle titel_dupdaten wurde bei knapp 140.000 Datenbankeinträgen ca. 47 MB groß und dauerte ca. 2 Stunden.

Andreas Friedel
Verbundzentrale BVB/A