SCRIPTS FOR RETRIEVING DATASETS

SCRIPTS

ADVANTAGES

- less client/server transactions
- procedures/functions can be called with parameters by the client
- stored procedures are compiled within database server, and therefore queries are executed faster
- stored procedures also separates query logic from the code logic.

DISADVANTAGE

- database specific

TEST ENVIRONMENTS

- windows XP

- 1G RAM

- Pentium4 3GHz

MYSQL

- version 5.0

POSTGRESQL

- version 8.2

RETRIEVING GENOTYPE AND PHENOTYPE STUDIES

- Get_dataset(phenotype study)
- Get_dataset(genotype study)
- Create_array_dataset(genotype table)

GET_DATASET

Description:

- Creates a new table in the database. The created table format is the same as the observation sheet.
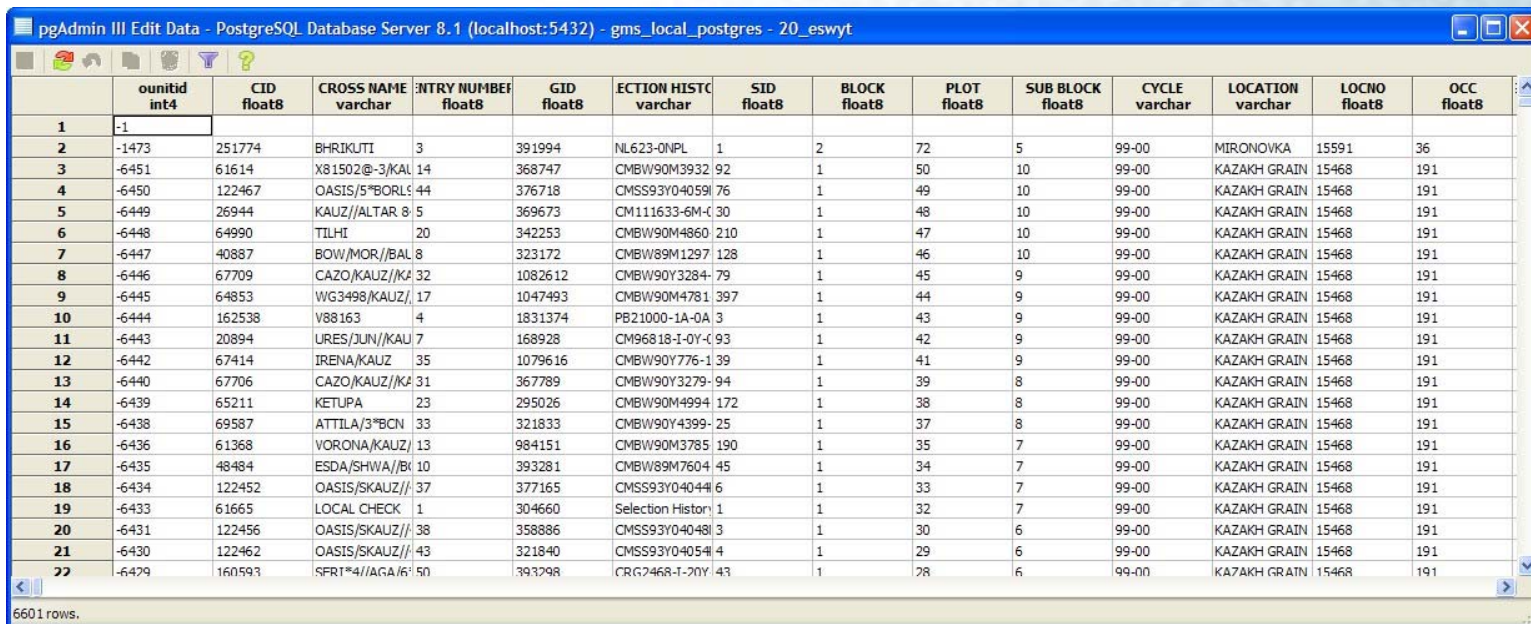
Input Parameter/s:

newTableName – Name of the table to be created. Study name is used to properly identify the table

studyName – Name of the Study

GET_DATASET

- ## POSTGRESQL:

  SELECT get_dataset('20_eswyt','20 ESWYT');
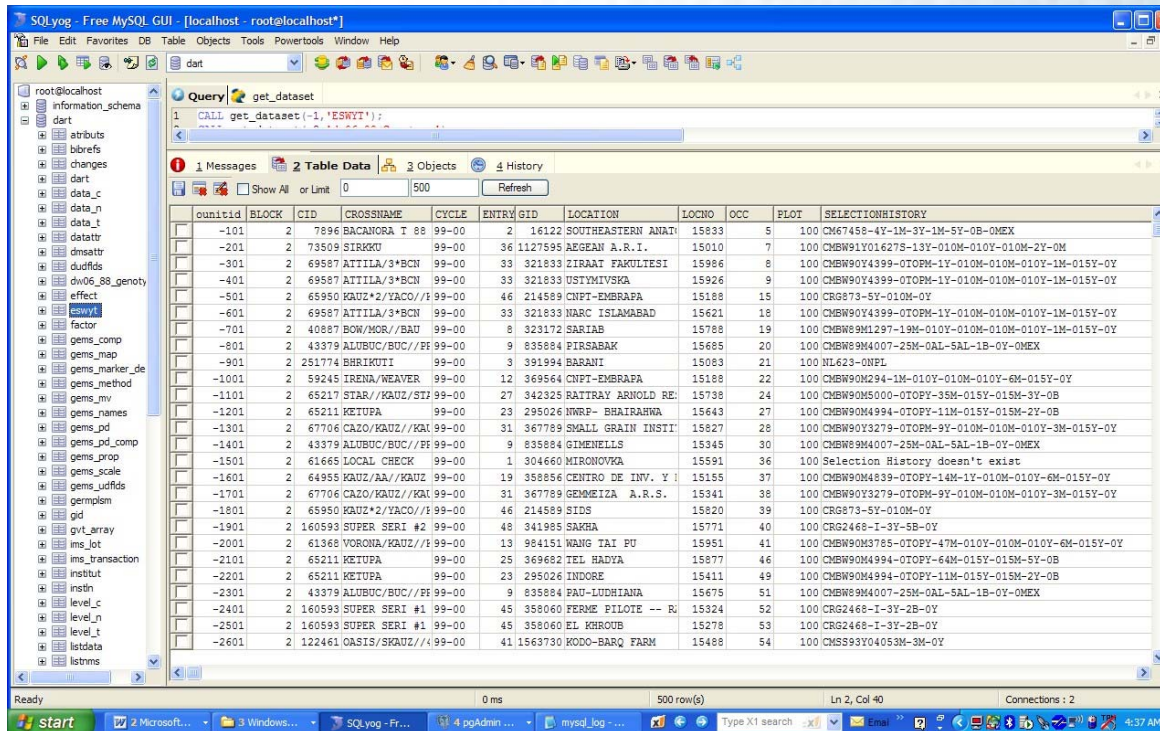 **Total query runtime: 11157 ms.**

GET_DATASET

- ## POSTGRESQL:

SELECT get_dataset('dart','DArT_Genotype');   (33,667 rows)
    Total query runtime: 6984 ms. ~ 7 secs

SELECT get_dataset('DW06_88_Genotyp','DW06_88_Genotyp');
    Total query runtime: 14781 ms. ~ 15 secs

GET_DATASET

- ## MySQL:

   /*[4:28:25 AM][45828 ms]*/ CALL get_dataset(-1,'ESWYT')
       ~ 45 secs

GET_DATASET

- **MySQL:**

Other Example :
Retrieval of a genotyping dataset with 33,667 rows
/*[2:08:10 PM][75984ms]*/ CALL get_dataset(-2,'dart')

CALL create_array_dataset('dw06_88_Genotyp')
/*[1:03:06 AM][78000 ms]*/ CALL get_dataset(-2,
'dw06_88_Genotyp')

GET_DATASET

- Workbook:
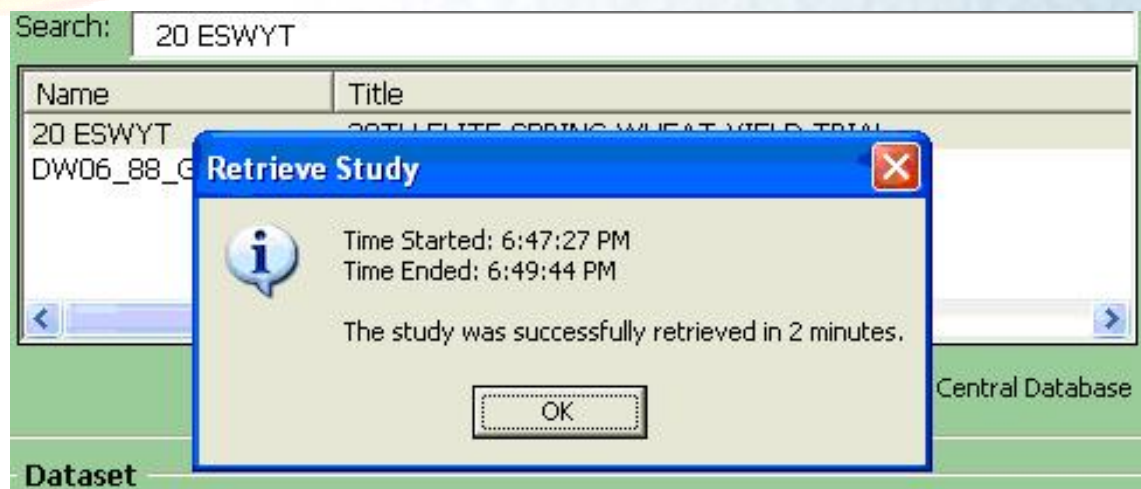
Another Example :
Retrieval of a genotyping dataset with 33,667 rows
(it takes 3mins in workbook)

CREATE_ARRAY_DATASET

- **DESCRIPTION:**

  creates a new table with two columns – GID and marker. The marker column is type text. It contains the concatenated allele values for each marker tested on the GID. Essentially, transforms the retrieved serial dataset into a parallel dataset.

- **INPUT PARAMETERS:**

  genotyping table with GID,MARKER, ALLELE as columns

  - output from the get_dataset(genotype)

CREATE_ARRAY_DATASET

ounitid	CLONE_ID	GID	MARKERNAI	MARKER_ID	ALLELE	ALLELEID
-47987	119455	3826179	wPt-9310	445	0	666
-47542	119455	3826178	wPt-9310	445	0	666
-47097	119455	3825285	wPt-9310	445	X	1043
-46652	119455	3822784	wPt-9310	445	0	666
-46207	119455	3826177	wPt-9310	445	0	666
-45762	119455	3825284	wPt-9310	445	0	666
-45317	119455	2621166	wPt-9310	445	0	666
-44872	119455	2621174	wPt-9310	445	0	666
-44427	119455	2621109	wPt-9310	445	0	666
-43982	119455	2621094	wPt-9310	445	0	666
-43537	119455	2620837	wPt-9310	445	0	666
-43092	119455	2485920	wPt-9310	445	0	666
-42647	119455	2485890	wPt-9310	445	0	666
-42202	119455	2485260	wPt-9310	445	0	666
-41757	119455	2485148	wPt-9310	445	0	666
-41312	119455	2485043	wPt-9310	445	0	666
-40867	119455	2485050	wPt-9310	445	0	666
-40422	119455	2484320	wPt-9310	445	0	666
-39977	119455	2484301	wPt-9310	445	0	666
-39532	119455	2484313	wPt-9310	445	0	666
-39087	119455	4062993	wPt-9310	445	0	666
-38642	119455	4057155	wPt-9310	445	0	666
-38197	119455	1673085	wPt-9310	445	0	666
-37752	119455	1671946	wPt-9310	445	0	666
-37307	119455	1649008	wPt-9310	445	0	666
-36862	119455	4069854	wPt-9310	445	0	666

CREATE_ARRAY_DATASET

MYSQL:
68391 ms

POSTGRESQL:
2500ms