

# 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:

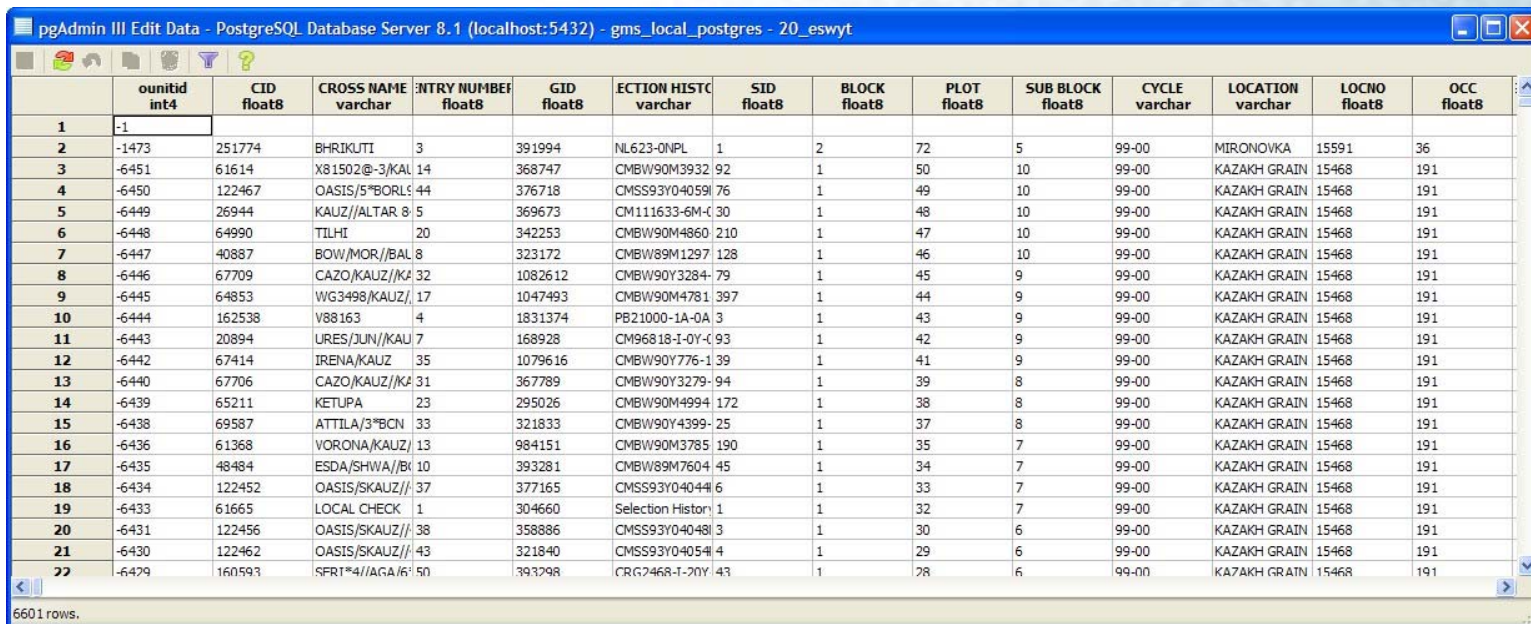
newTableName – Name of the table to be created. Study name is used to properly identify the table

studyName – Name of the Study

# GET\_DATASET

- **POSTGRESQL:**

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



	ounitid int4	CID float8	CROSS NAME varchar	ENTRY NUMBER float8	GID float8	SECTION HISTO varchar	SID float8	BLOCK float8	PLOT float8	SUB BLOCK float8	CYCLE varchar	LOCATION varchar	LOCNO float8	OCC float8
1	-1													
2	-1473	251774	BHRIKUTI	3	391994	NL623-ONPL	1	2	72	5	99-00	MIRONOVKA	15591	36
3	-6451	61614	X81502@-3/KAL	14	368747	CMBW90M3932	92	1	50	10	99-00	KAZAKH GRAIN	15468	191
4	-6450	122467	OASIS/5*BORLS	44	376718	CMSS93Y04059I	76	1	49	10	99-00	KAZAKH GRAIN	15468	191
5	-6449	26944	KAUZ//ALTAR	8	369673	CM111633-6M-C	30	1	48	10	99-00	KAZAKH GRAIN	15468	191
6	-6448	64990	TILHI	20	342253	CMBW90M4860	210	1	47	10	99-00	KAZAKH GRAIN	15468	191
7	-6447	40887	BOW/MOR//BAL	8	323172	CMBW89M1297	128	1	46	10	99-00	KAZAKH GRAIN	15468	191
8	-6446	67709	CAZO//KAUZ//K2	32	1082612	CMBW90Y3284	79	1	45	9	99-00	KAZAKH GRAIN	15468	191
9	-6445	64853	WG3498//KAUZ//	17	1047493	CMBW90M4781	397	1	44	9	99-00	KAZAKH GRAIN	15468	191
10	-6444	162538	V88163	4	1831374	PB21000-1A-0A	3	1	43	9	99-00	KAZAKH GRAIN	15468	191
11	-6443	20894	URES//JUN//KAU	7	168928	CM96818-1-0Y-C	93	1	42	9	99-00	KAZAKH GRAIN	15468	191
12	-6442	67414	IRENA//KAUZ	35	1079616	CMBW90Y776-1	39	1	41	9	99-00	KAZAKH GRAIN	15468	191
13	-6440	67706	CAZO//KAUZ//K2	31	367789	CMBW90Y3279	94	1	39	8	99-00	KAZAKH GRAIN	15468	191
14	-6439	65211	KETUPA	23	295026	CMBW90M4994	172	1	38	8	99-00	KAZAKH GRAIN	15468	191
15	-6438	69587	ATTILA//3*BCN	33	321833	CMBW90Y4399	25	1	37	8	99-00	KAZAKH GRAIN	15468	191
16	-6436	61368	VORONA//KAUZ	13	984151	CMBW90M3785	190	1	35	7	99-00	KAZAKH GRAIN	15468	191
17	-6435	48484	ESDA//SHWA//B	10	393281	CMBW89M7604	45	1	34	7	99-00	KAZAKH GRAIN	15468	191
18	-6434	122452	OASIS//SKAUZ//	37	377165	CMSS93Y04044	6	1	33	7	99-00	KAZAKH GRAIN	15468	191
19	-6433	61665	LOCAL CHECK	1	304660	Selection Histor	1	1	32	7	99-00	KAZAKH GRAIN	15468	191
20	-6431	122456	OASIS//SKAUZ//	38	358886	CMSS93Y04048	3	1	30	6	99-00	KAZAKH GRAIN	15468	191
21	-6430	122462	OASIS//SKAUZ//	43	321840	CMSS93Y04054	4	1	29	6	99-00	KAZAKH GRAIN	15468	191
22	-6429	160593	SFR1*4//AGA/6	50	393298	CRG2468-1-20Y	43	1	28	6	99-00	KAZAKH GRAIN	15468	191

# GET\_DATASET

- **POSTGRESQL:**

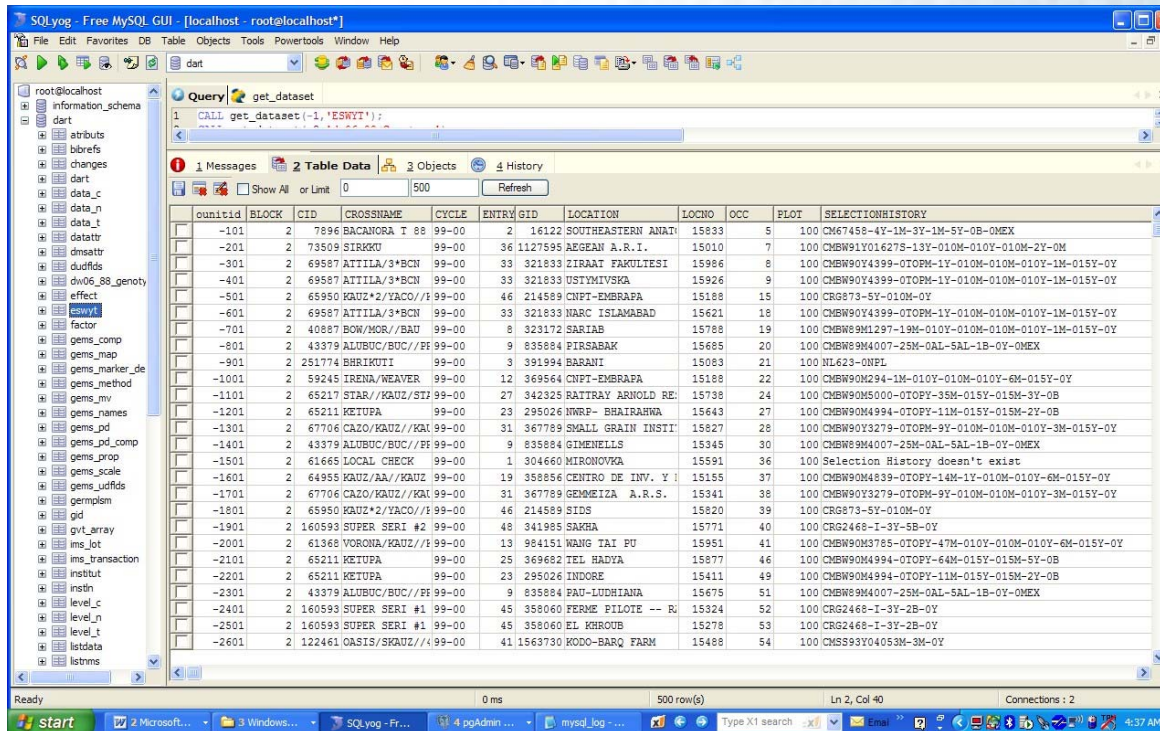
```
SELECT get_dataset('dart','DArT_Genotype');    (33,667 rows)  
Total query runtime: 6984 ms. ~ 7 secs
```

```
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**



The screenshot shows the SQLyog GUI with a query window displaying the following query: `CALL get_dataset(-1,'ESWYT');` The results pane shows a table with columns: `countid`, `BLOCK`, `CID`, `CROSSNAME`, `CYCLE`, `ENTRY_GID`, `LOCATION`, `LOCNO`, `OCC`, `PLOT`, and `SELECTIONHISTORY`. The table contains 26 rows of data, including entries for locations like BACANORA, SIRAOKU, ATILLA, KAUZ, and KETUPA.

countid	BLOCK	CID	CROSSNAME	CYCLE	ENTRY_GID	LOCATION	LOCNO	OCC	PLOT	SELECTIONHISTORY
-101	2	7896	BACANORA T 88	99-00	2	16122 SOUTHEASTERN ANAT	15833	5	100	CM67458-4Y-1M-3Y-1M-5Y-0B-0MEX
-201	2	73509	SIRAOKU	99-00	36	1127595 AEGEAN A.R.I.	15010	7	100	CMBW91Y01627S-13Y-010M-010Y-010M-2Y-0M
-301	2	69587	ATILLA/3*BCN	99-00	33	321833 ZIRJAT FAKULTESI	15986	8	100	CMBW90Y4399-OTOPM-1Y-010M-010M-010Y-1M-015Y-0Y
-401	2	69587	ATILLA/3*BCN	99-00	33	321833 USTYMIVSKA	15926	9	100	CMBW90Y4399-OTOPM-1Y-010M-010M-010Y-1M-015Y-0Y
-501	2	65950	KAUZ*2/YACO//I	99-00	46	214589 CNPT-EMBRAPA	15188	15	100	CRG873-5Y-010M-0Y
-601	2	69587	ATILLA/3*BCN	99-00	33	321833 NARC ISLAMABAD	15621	18	100	CMBW90Y4399-OTOPM-1Y-010M-010M-010Y-1M-015Y-0Y
-701	2	40887	BOW/MOR//BAU	99-00	8	323172 SARIAB	15788	19	100	CMBW9M1297-19M-010Y-010M-010M-010Y-1M-015Y-0Y
-801	2	43379	ALUBUC/BUC//FH	99-00	9	835884 PIRABAK	15685	20	100	CMBW9M4007-25M-0AL-SAL-1B-0Y-0MEX
-901	2	251774	BHRIKUTI	99-00	3	391994 BARANI	15083	21	100	NL623-0NPL
-1001	2	59245	IRENA/WEAVER	99-00	12	369564 CNPT-EMBRAPA	15188	22	100	CMBW9M294-1M-010Y-010M-010Y-6M-015Y-0Y
-1101	2	65217	STAR//KAUZ/STJ	99-00	27	342325 RATIRAY ARNOLD RE	15738	24	100	CMBW9M5000-OTOPY-35M-015Y-015M-3Y-0B
-1201	2	65211	KETUPA	99-00	23	295026 MWRF- BHAIRAHWA	15643	27	100	CMBW9M4994-OTOPY-11M-015Y-015M-2Y-0B
-1301	2	67706	CAZO//KAU//KAI	99-00	31	367789 SMALL GRAIN INSTI	15827	28	100	CMBW90Y3279-OTOPM-9Y-010M-010M-010Y-3M-015Y-0Y
-1401	2	43379	ALUBUC/BUC//FH	99-00	9	835884 GIMENELLS	15345	30	100	CMBW9M4007-25M-0AL-SAL-1B-0Y-0MEX
-1501	2	61665	LOCAL CHECK	99-00	1	304660 MIRONOVKA	15591	36	100	Selection History doesn't exist
-1601	2	64955	KAUZ/AA//KAUZ	99-00	19	359856 CENTRO DE INV. Y	15155	37	100	CMBW9M4839-OTOPY-14M-1Y-010M-010Y-6M-015Y-0Y
-1701	2	67706	CAZO//KAU//KAI	99-00	31	367789 GEMMEIZA A.R.S.	15341	38	100	CMBW90Y3279-OTOPM-9Y-010M-010M-010Y-3M-015Y-0Y
-1801	2	65950	KAUZ*2/YACO//I	99-00	46	214589 SIDS	15820	39	100	CRG873-5Y-010M-0Y
-1901	2	160593	SUPER SERI #2	99-00	48	341985 SARHA	15771	40	100	CRG2468-I-3Y-5B-0Y
-2001	2	61368	VORONKA//KAUZ//H	99-00	13	984151 WANG TAI FU	15951	41	100	CMBW9M3785-OTOPY-47M-010Y-010M-010Y-6M-015Y-0Y
-2101	2	65211	KETUPA	99-00	25	369682 TEL HADVA	15877	46	100	CMBW9M4994-OTOPY-64M-015Y-015M-5Y-0B
-2201	2	65211	KETUPA	99-00	23	295026 INDORE	15411	49	100	CMBW9M4994-OTOPY-11M-015Y-015M-2Y-0B
-2301	2	43379	ALUBUC/BUC//FH	99-00	9	835884 PAU-LUDHIANA	15675	51	100	CMBW9M4007-25M-0AL-SAL-1B-0Y-0MEX
-2401	2	160593	SUPER SERI #1	99-00	45	358060 FERME PILOTE -- RU	15324	52	100	CRG2468-I-3Y-2B-0Y
-2501	2	160593	SUPER SERI #1	99-00	45	358060 EL KHROUB	15278	53	100	CRG2468-I-3Y-2B-0Y
-2601	2	122461	OASIS/SKADZ//I	99-00	41	1563730 KODO-BARQ FARM	15488	54	100	CMSS93Y04053M-3M-0Y



# GET\_DATASET

- **MySQL:**

Other Example :

Retrieval of a genotyping dataset with 33,667 rows

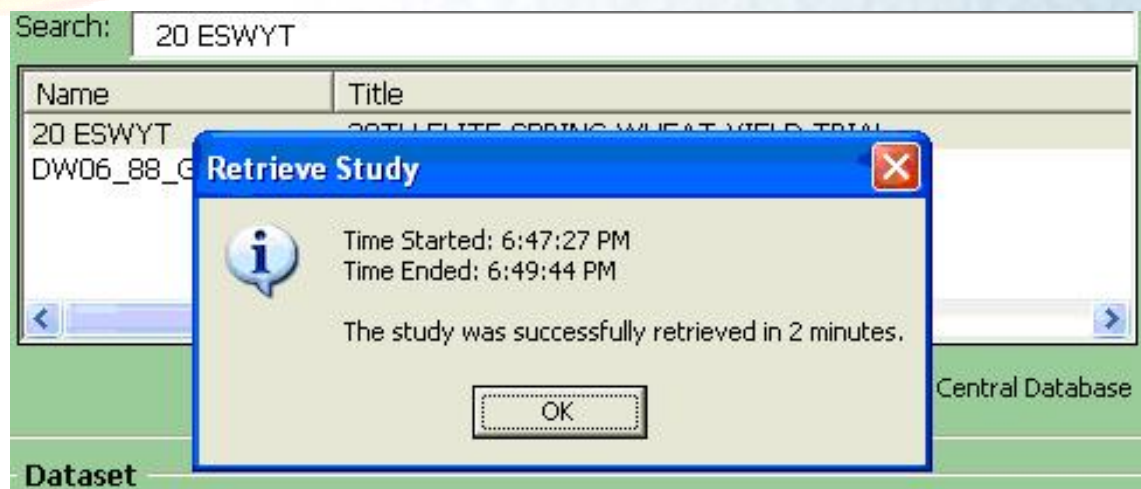
```
/*[2:08:10 PM][75984ms]*/ CALL get_dataset(-2,'dart')
```

```
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

gid	m		
<input type="checkbox"/>	390604	0:1:0:0:0:1:1:1:1:0:0:0:0:1:0:1:0:1:0:1:0:1:0:0:0:1:0:1:1...	889 b...
<input type="checkbox"/>	358886	0:1:0:0:0:1:1:1:1:0:0:0:0:1:0:1:0:1:0:1:0:1:0:0:0:1:0:1:1...	889 b...
<input type="checkbox"/>	377165	1:1:0:0:0:1:0:1:1:0:0:1:1:X:1:0:1:1:0:0:1:1:1:X:0:0:1:0:0...	889 b...
<input type="checkbox"/>	1127595	0:1:0:0:0:1:1:1:1:0:1:0:X:1:0:1:0:1:0:1:0:1:0:0:0:1:0:1:1...	889 b...
<input type="checkbox"/>	1079616	0:X:1:0:X:1:1:1:1:1:0:X:X:X:1:X:1:1:X:1:1:1:0:X:0:1:X:1:1...	889 b...
<input type="checkbox"/>	1101307	0:1:0:0:0:1:1:1:1:X:0:X:X:1:0:X:0:1:0:X:0:1:0:X:0:X:0:1:1...	889 b...
<input type="checkbox"/>	321833	0:1:X:0:0:1:1:1:1:0:0:0:0:1:1:1:1:1:0:1:1:1:1:0:0:1:0:1:1...	889 b...
<input type="checkbox"/>	1082612	0:1:0:1:0:1:0:1:0:0:1:0:0:0:0:1:0:1:0:1:0:1:0:0:0:0:0:1:1...	889 b...
<input type="checkbox"/>	367789	0:0:0:0:0:1:1:1:1:X:X:X:1:1:X:1:X:1:0:1:X:1:X:X:0:1:0:1:1...	889 b...
<input type="checkbox"/>	321822	0:1:0:0:0:1:0:1:1:X:1:0:0:1:0:1:0:1:0:0:0:0:1:0:0:0:0:X:1...	889 b...
<input type="checkbox"/>	1081265	0:1:0:0:0:1:0:1:1:0:1:0:1:0:0:1:0:1:0:1:0:0:0:0:1:0:0:X:0:0:1...	889 b...
<input type="checkbox"/>	295146	0:1:0:0:0:1:0:1:1:0:0:1:1:0:1:0:1:1:0:0:1:1:0:0:0:1:1:1:1...	889 b...
<input type="checkbox"/>	342325	0:1:0:0:0:1:0:1:1:1:0:1:1:0:1:0:1:0:X:1:0:0:X:1:1:X:0:1:1:1:0...	889 b...
<input type="checkbox"/>	295091	1:1:0:0:0:1:0:1:1:1:0:1:0:0:0:0:0:1:0:0:0:0:1:1:0:0:1:0:1:0...	889 b...
<input type="checkbox"/>	369682	0:1:0:0:0:1:1:1:1:1:0:0:0:1:0:1:0:1:0:1:0:1:0:0:0:1:0:1:1...	889 b...
<input type="checkbox"/>	292717	X:1:0:0:0:1:1:1:1:1:0:0:1:1:1:1:1:1:0:1:1:1:0:X:0:1:1:1:1...	889 b...
<input type="checkbox"/>	295026	1:1:0:0:0:1:1:1:1:1:0:0:0:1:X:1:X:1:0:1:X:1:1:0:0:1:0:1:X...	889 b...
<input type="checkbox"/>	369679	1:1:0:0:0:1:0:1:1:1:0:1:1:1:1:0:1:1:0:1:1:1:1:X:0:1:1:1:1...	889 b...
<input type="checkbox"/>	367751	0:0:0:0:0:0:1:0:1:0:0:0:0:0:0:X:0:X:1:0:0:X:1:1:1:0:1:0:0:0...	889 b...
<input type="checkbox"/>	342253	0:0:0:0:0:X:1:0:1:0:0:0:0:0:0:1:0:1:1:0:0:1:1:1:1:0:1:0:0:0...	889 b...
<input type="checkbox"/>	358856	0:X:0:0:0:1:1:1:1:0:0:0:0:1:0:1:0:1:0:1:0:1:X:X:0:1:0:1:1...	889 b...
<input type="checkbox"/>	1064007	0:0:X:X:X:1:1:1:1:1:X:X:X:1:0:1:0:1:X:1:0:X:X:X:0:1:0:1:1...	889 b...
<input type="checkbox"/>	1047493	0:X:0:0:0:1:1:1:1:0:0:0:0:1:0:1:0:1:0:1:0:1:0:0:0:1:0:1:1...	889 b...
<input type="checkbox"/>	934019	1:1:0:0:1:0:0:1:1:1:0:1:1:0:1:0:1:0:0:1:1:1:1:0:0:0:0:0...	889 b...
<input type="checkbox"/>	342202	1:1:1:1:1:1:0:1:1:1:0:1:1:1:1:1:1:1:1:1:1:1:1:1:0:1:0:1:1...	889 b...
<input type="checkbox"/>	368747	0:1:0:0:0:1:1:0:1:0:0:0:0:1:0:1:0:0:0:1:0:1:0:0:0:1:0:1:1...	889 b...