



ENHANCED DIII-D DATA MANAGEMENT THROUGH A RELATIONAL DATABASE

Presented by

J. Burruss

for the DIII-D National Team

Presented to

42nd Annual Meeting of the Division of Plasma Physics

October 23–27, 2000

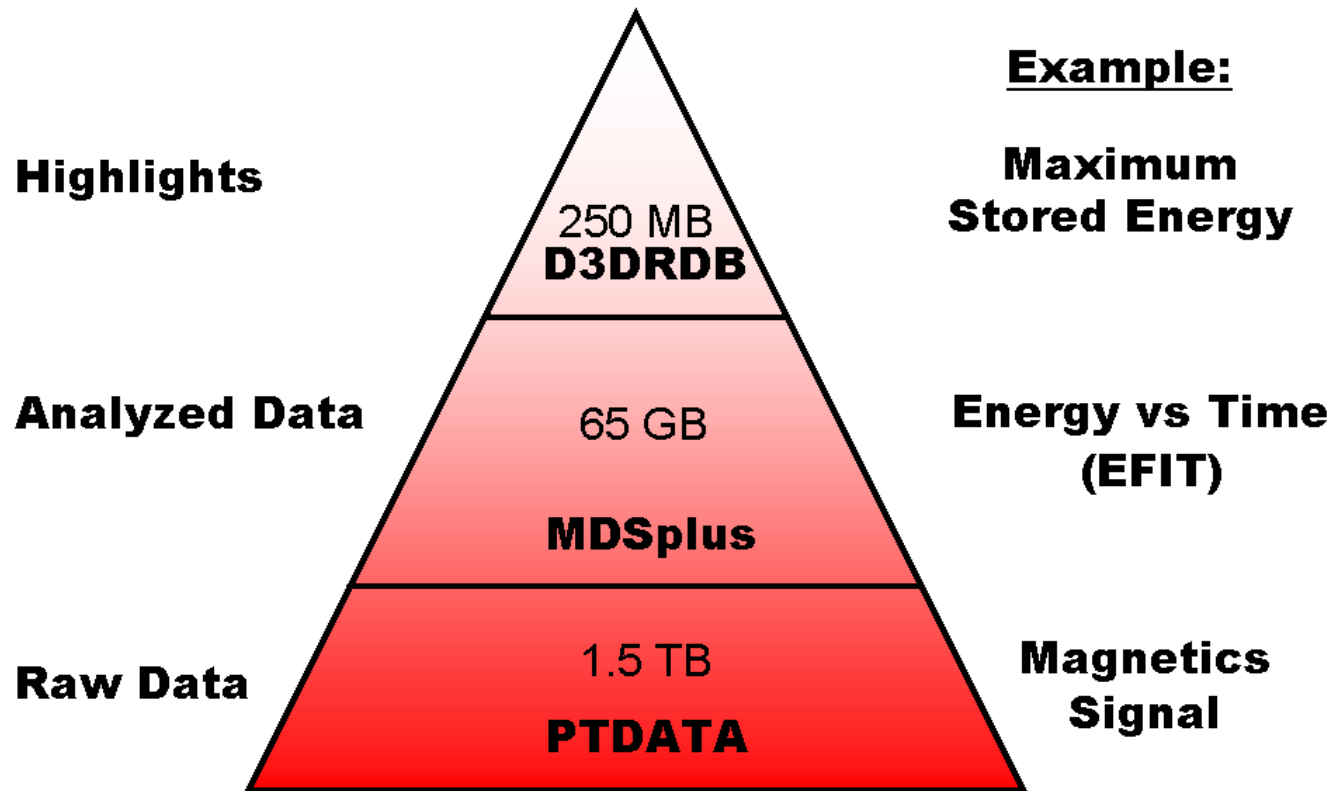
Quebec City, Canada

SUMMARY

- **The DIII-D Relational Database is being used in concert with MDSplus and PTDATA to manage experimental data at DIII-D.**
- **The relational database stores highlights of the experimental data.**
 - The complete multi-Terabyte data registry is summarized by a 255 Megabyte relational database.
 - Summary shot quantities, comments, and reports are stored.
- **The database is also used to monitor the DIII-D computational infrastructure.**
 - Disk usage, code usage, and CPU usage are stored in the database.
 - Yields data for making intelligent choices about upgrading CPUs, buying new disks, etc.
- **Data is only as useful as it is accessible, so many ways to view the data are provided:**
 - IDL
 - Electronic logbook
 - Web access (including simple data plots)
 - DBPlot (for more complex data plots)
 - Any ODBC-compliant commercial application (such as Excel or Access)

THE D3DRDB DATABASE CONTAINS DATA HIGHLIGHTS

- It's a good place to start your data exploration since you have less data to sift through.
- You can drill down to more detailed information.



SUMMARY PHYSICS QUANTITIES ARE STORED IN D3DRDB

- **The summaries table stores summary quantities for every plasma shot at DIII-D.**
 - New shots are automatically loaded 10-15 minutes after the shot.
 - Older shots are loaded manually.
 - There are currently 49564 shots in the summaries table.
 - Many of these shots were taken from an existing database¹.
- **The entries table stores comments and reports.**
 - 38829 shot comments, the oldest of which date back to 1986.
 - 3859 reports (e.g. session leader summaries, physics operator reports).
- **Other tables keep track of experimental runs and personnel.**

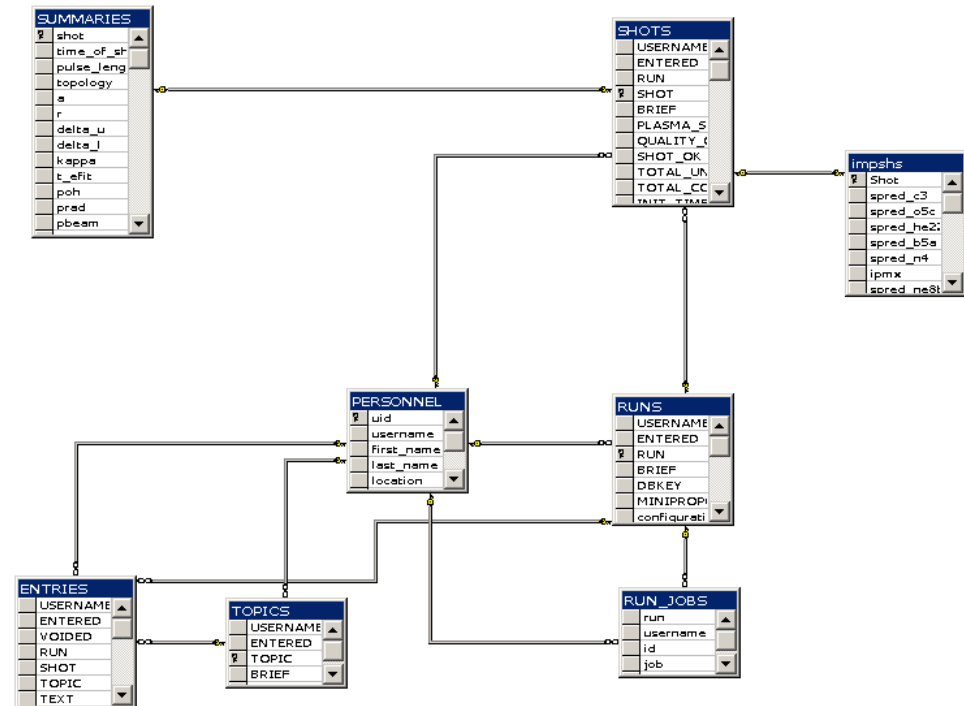
¹ Maintained by Pete Taylor, this database contained many of the quantities currently recorded in the summaries table.

ADDITIONAL INFORMATION MAY BE STORED IN USER TABLES

- **Users can submit their own tables.**
- **Usually related to other tables by shot.**
- **There are strong benefits to user tables:**
 - Tables may contain information useful to many DIII-D personnel
 - Users may immediately relate their data to other data in the database
 - Regular backups with no user effort
- **Example user tables:**
 - Thrust9: contains gyrotron data; indexed by shot and gyrotron
 - Ornl_pellets: pellet information; indexed by shot and pellet number
 - Impshs: impurity data for Tokamak shots

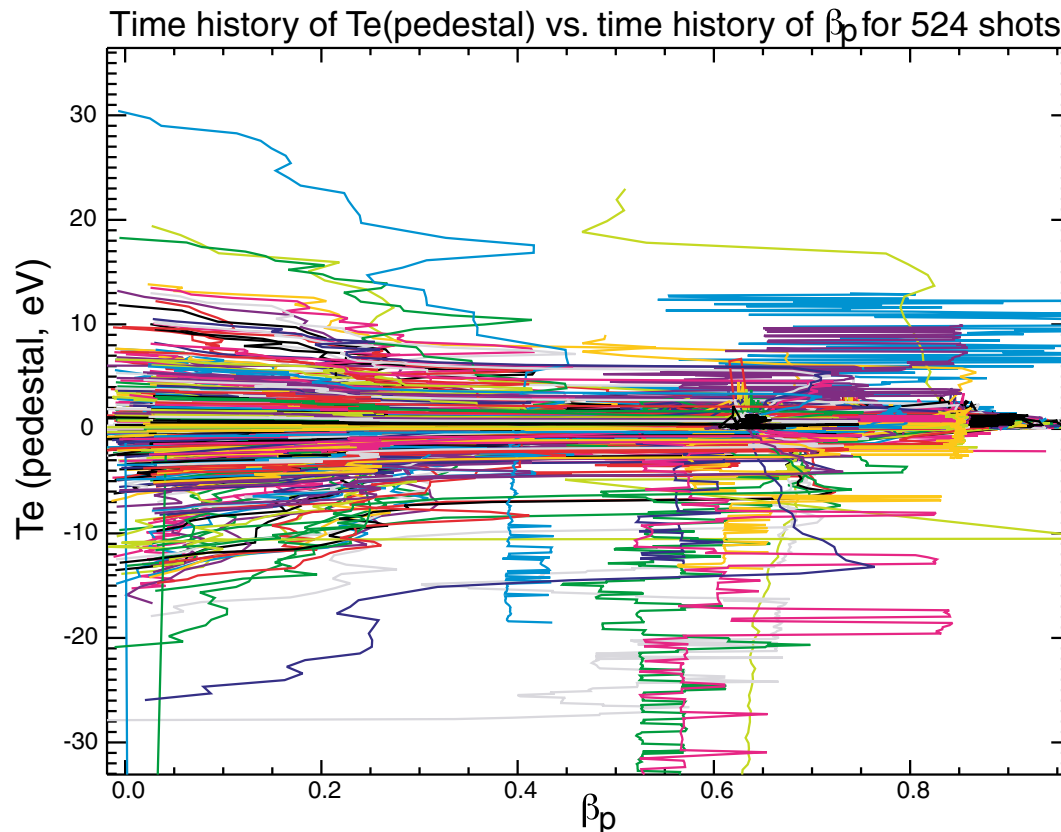
D3DRDB RELATES THESE DIFFERENT EXPERIMENTS AND DATASETS

- **MDSplus stores data for a given shot hierarchically.**
 - Not efficient for queries across multiple shots.
- **D3DRDB has a different organization.**
 - Tables may have multiple relationships with other tables.
 - Optimized for queries across multiple shots.
- **Different datasets are related to each other.**
- **You can quickly answer questions like “which shots from 1999 had a maximum plasma current of at least 2.0 MA?”**



RELATIONAL DATABASE MAKES SURVEYING SHOTS FROM MDSplus FASTER

- Use SQL query to quickly identify a small number (524) of shots of interest
 - Maximum $\beta_p > 1.2$
- Then go to MDSplus for time history data and further windowing
 - $\beta_p > 1.2$ for a one second & then plot pedestal temperature versus β_p
- Scanning every shot in MDSplus not required



THERE ARE MANY WAYS TO ACCESS D3DRDB

- You can access the database through any ODBC compliant application.

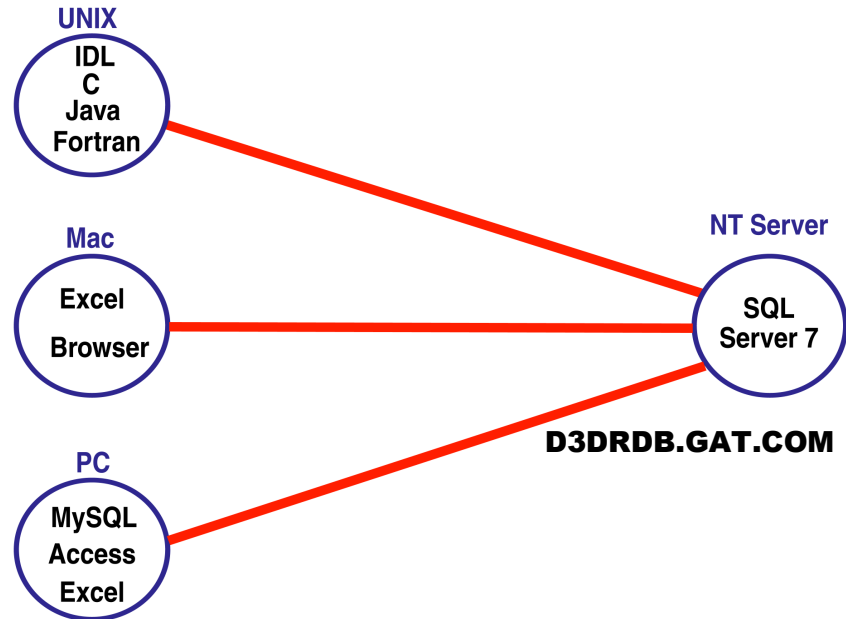
- Excel
- Access
- JMP
- MySQL and other ODBC databases

- You can connect from many different operating systems.

- Web interfaces let users view data through a web browser.

- You can access the database using programming languages.

- Java
- IDL
- Other languages



THE MOST BASIC WAY TO ACCESS THE DATABASE IS THROUGH STRUCTURED QUERY LANGUAGE (SQL)

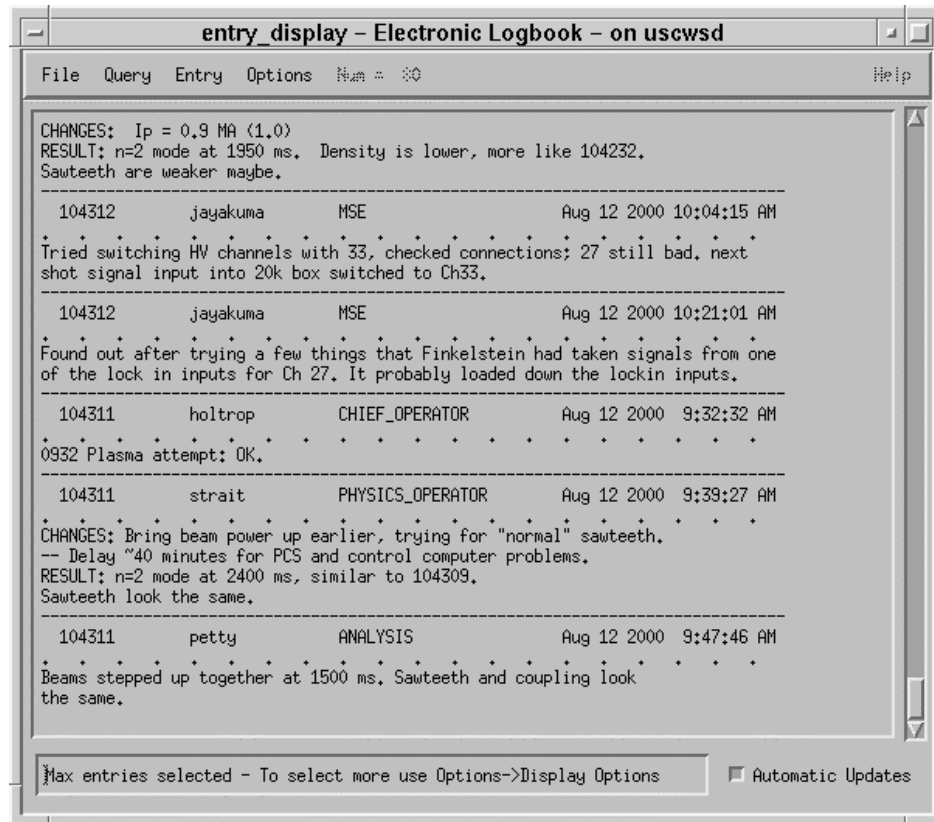
- SQL is the database language used by all major database vendors.

- SQL is easy to understand. An example query:

```
SELECT shot  
FROM summaries  
WHERE kappa BETWEEN 1.8 AND 1.9
```

- This query retrieves a list of shots where kappa is between 1.8 and 1.9.
- You can submit SQL queries through IDL, the web, and some ODBC-compliant applications.

YOU CAN ACCESS THE DATABASE THROUGH THE ELECTRONIC LOGBOOK²



```
entry_display - Electronic Logbook - on usewsd
File Query Entry Options Num = 00 Help
CHANGES: Ip = 0,9 MA (1,0)
RESULT: n=2 mode at 1950 ms. Density is lower, more like 104232.
Sawteeth are weaker maybe.
-----
104312      jayakuma      MSE              Aug 12 2000 10:04:15 AM
-----
Tried switching HV channels with 33, checked connections; 27 still bad, next
shot signal input into 20k box switched to Ch33.
-----
104312      jayakuma      MSE              Aug 12 2000 10:21:01 AM
-----
Found out after trying a few things that Finkelstein had taken signals from one
of the lock in inputs for Ch 27. It probably loaded down the lockin inputs.
-----
104311      holtrop       CHIEF_OPERATOR   Aug 12 2000  9:32:32 AM
-----
0932 Plasma attempt: OK.
-----
104311      strait        PHYSICS_OPERATOR Aug 12 2000  9:39:27 AM
-----
CHANGES: Bring beam power up earlier, trying for "normal" sawteeth.
-- Delay 40 minutes for PCS and control computer problems.
RESULT: n=2 mode at 2400 ms, similar to 104309.
Sawteeth look the same.
-----
104311      petty         ANALYSIS         Aug 12 2000  9:47:46 AM
-----
Beams stepped up together at 1500 ms. Sawteeth and coupling look
the same.
-----
Max entries selected - To select more use Options->Display Options  Automatic Updates
```

- Used both for viewing and entering shot comments and run summaries.

- An open logbook updates itself automatically (no pushing the refresh button).

- All entries are stored in the relational database.

- Past entries can be rapidly searched.

² Developed at MIT.

EXPERIMENTAL SUMMARIES ARE AVAILABLE FROM THE WEB

- The Experimental Summaries page generates reports from data stored in D3DRDB.
- Users can search for summaries based on any of several criteria.

Summary Information on DIII-D Physics Experiments

Click here for most recent summary

Browse by calendar Year: Month:

Search by date of experiment (in ISO format e.g. 19990219 = February 19, 1999 and 200003 = March 2000)
 Normal Quick Search

Search by shot number

Search by miniproposal number

Search by name (1993-present)

Search by string in summary.
 Normal Quick Search

Submit SQL query (the sample query finds the most recent shot)

```
SELECT shot, time_of_shot
FROM summaries
WHERE shot = ( SELECT MAX(shot) FROM summaries )
```

Summary Information on DIII-D Physics Experiments
 Experimental Summary (20000510)

High Density H-mode in Closed Divertor
 Run: 20000510 Shot Range: 102805 - 102836
 Session Leader: petrie Assistant Session Leader: fenster
 Physics Operator: leonard Assistant Physics Operator: hyatt
Miniproposal: 534

Summaries Found: SESSION_LEADER
 PHYSICS_OPERATOR
 CHIEF_OPERATOR

Electronic Logbook Entries Found: 85

(click here for executive view)

25 plasma shots found (click any shot number for more information)

SHOT	TIME	IP (MA)	BTOR	PBEAM (MW)	A	R	KAPPA	NEUTRONS	WTOTMAX (MJ)
102807	08:26	0.03	0.86					1.0 E8	
102808	08:42	0.04	0.86					1.0 E8	
102809	08:55	0.04	0.95					1.0 E8	
102810	09:39	0.04	0.86					1.0 E8	
102815	12:06	0.04	0.86					1.0 E8	
102816	12:19	1.37	2.04	4.69	0.60	1.73	1.72	2.2 E13	0.34
102817	12:33	1.37	2.04	6.98	0.60	1.73	1.73	5.0 E13	0.32

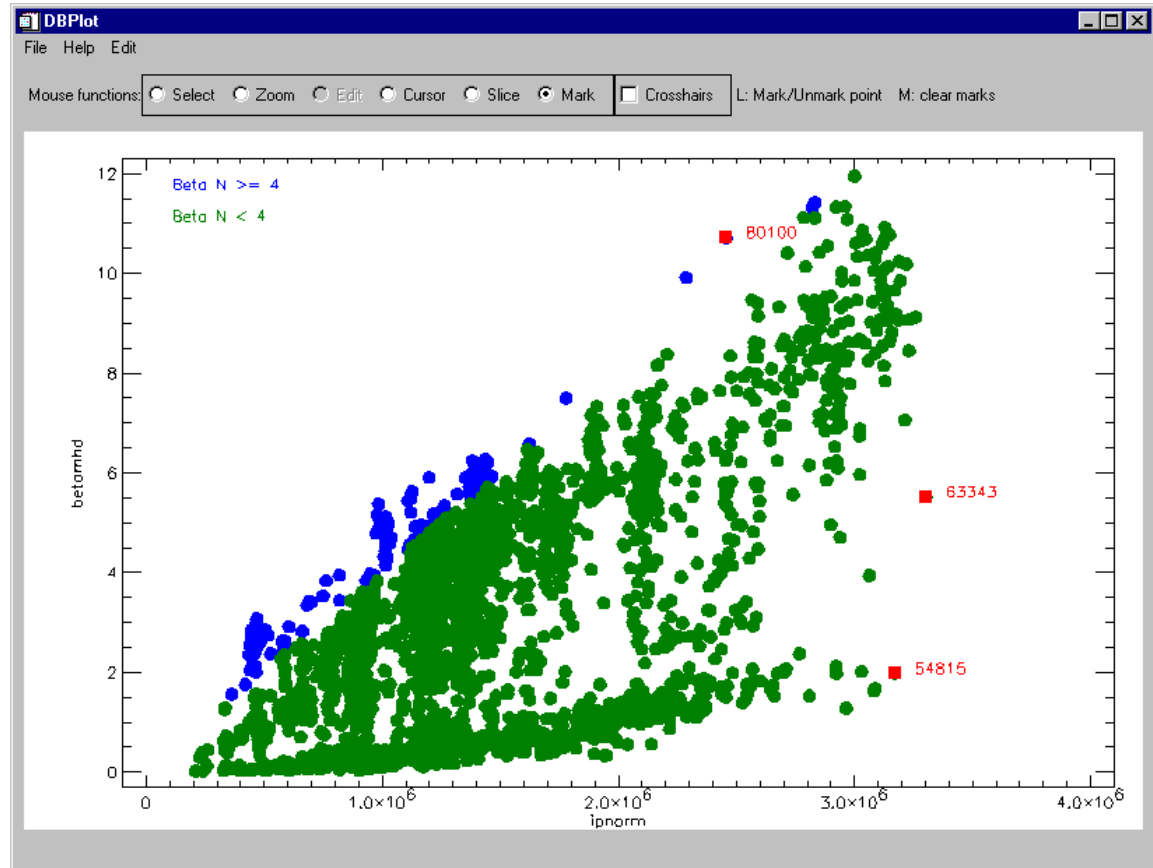
- Summaries contain experiment titles, miniproposal numbers, and personnel information.

- Shot highlights are also displayed.



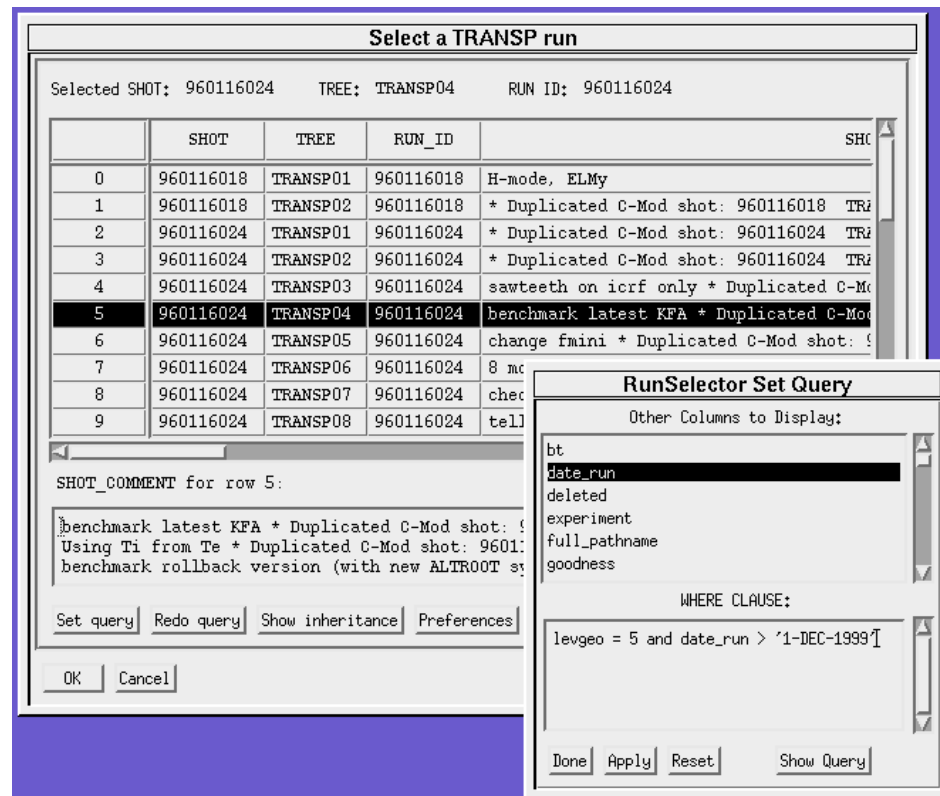
DBPLOT MAY BE USED TO PLOT SUMMARY QUANTITIES

- Users can type in a query and plot the results.
- Uses GAPIPlotObj
 - Provides useful features such as zoom, scale, and mark points.
 - Used in other tools (EFITViewer, ReviewPlus) so users do not need to learn a new user interface.
- Marked points from one plot show up as marked points in another plot.
 - Useful for identifying outlying points.



RUNSELECTOR LETS USERS TRACK PHYSICS ANALYSIS CODE RUNS

- RunSelector³ is a graphical program for displaying information in the code run database⁴.
- Provides a formalized method for tracking analysis code runs.
- Hierarchical organization of runs.
- Will be integrated into existing applications.



³ Developed by Jeff Schachter.

⁴ Based on a schema developed at MIT.

DATA DOCUMENTATION STORED IN THE DATABASE MAY BE ACCESSED THROUGH THE WEB

DIII-D Data Documentation

Browse EFIT signals

Search by name

D3DRDB documentation PTDATA and MDSPlus Signals Both

Search by string in description

D3DRDB documentation PTDATA and MDSPlus Signals Both

Browse by diagnostic

D3DRDB documentation PTDATA and MDSPlus Signals Both

Browse by responsible officer

D3DRDB documentation PTDATA and MDSPlus Signals Both

Submit SQL Query: (This sample query lists the number of each type of document in the database)

```
SELECT 'D3DRDB Documents' 'Document Type', COUNT(column_name) 'Documents Found'
FROM db_doc
UNION
SELECT 'Signals', COUNT(name)
FROM signal_names
```

Feel free to send any questions, suggestions, comments, or concerns about this web page and the D3DRDB database to Justin Burruss (burruss@fusion.gat.com).

- The Data Documentation page contains information on database quantities.
- The page also documents PTDATA and MDSplus pointnames.
- It is also searchable using any of several criteria.
- Help documents are available from tools via Netscape.

D3DRDB IS ALSO USED FOR INFRASTRUCTURE MONITORING

- **Disk usage**

- Usage information is automatically collected each morning.
- Keeps histories of disk usage.
- Can bring to light trends in disk usage.
- Indicates whether disks are approaching maximum capacity.

- **Code usage**

- Shows how analysis tools are being used.
- Records when a tool was used and who used it.
- Low usage may indicate a tool that is no longer needed, or perhaps just under-advertised.

- **Usage of Analysis Workstations (LSF Usage)**

- Monitors usage of machines in the LSF cluster; updated every 5 minutes.
- Tracks benchmarks such as CPU utilization, memory, and swap space.

DISK USE IS SUMMARIZED AT THE DISK USAGE PAGE

- Provides a clear picture of disk use.
 - Allows purchase of more disk space before it becomes critical.
- Calls attention to users that need to move files to an alternative storage location.
 - Tape backup for old files.
 - MDSplus for standard data files.

- In this example disk D is almost full, and about 1/3 of the disk is being misused.

- You can zoom in on disk D for details.

Total Misused	Directory	Owner	Efit	Extra	Cache	Old
1803.22	nondim	luce	86.28	43.30	988.93	684.71
637.40	invq	rice	333.01	74.26	49.74	180.39
456.39	fwcd	forest	178.95	18.09	99.30	160.05

- It looks like 989 MB may be freed up by asking one user to move his Cache files off the filer.

- It may be time to increase disk capacity.

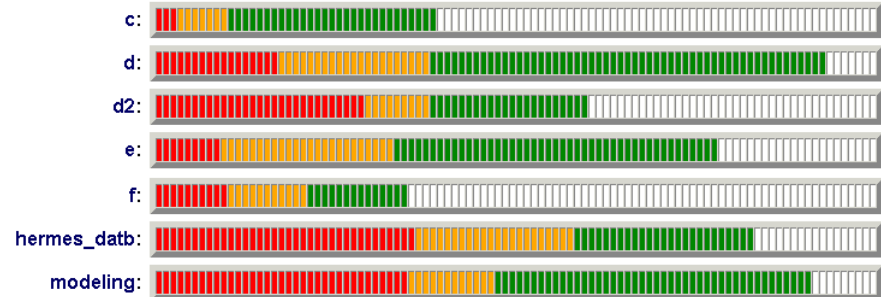
Disk Usage

Username	Total Misused	Share of Total Disk Misuse
luce	14671 MB	23%
murakami	7143 MB	11%
root	4309 MB	7%
west	3032 MB	4%
stjohn	1954 MB	3%

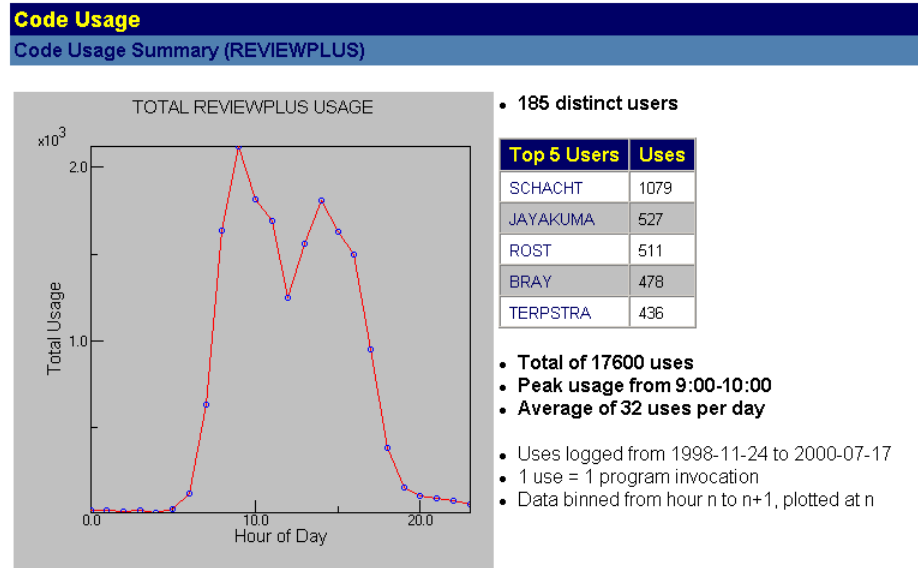
Files that should **not** be stored on the Filer are:

- EFIT Output files (A, G, and M files)
- Extraneous EFIT files
- GAPfiles cache files
- Old files and directories

Disk Usage Chart



THE CODE USAGE DATABASE SHOWS HOW TOOLS ARE BEING USED



● Usage reports are available from the web.

● Programmers can see how their programs are being used.

— Does a tool meet user needs?

● Identify who uses a tool the most.

— Make sure any changes to a tool are acceptable to these 'power users.'

● Find out who uses a tool the least.

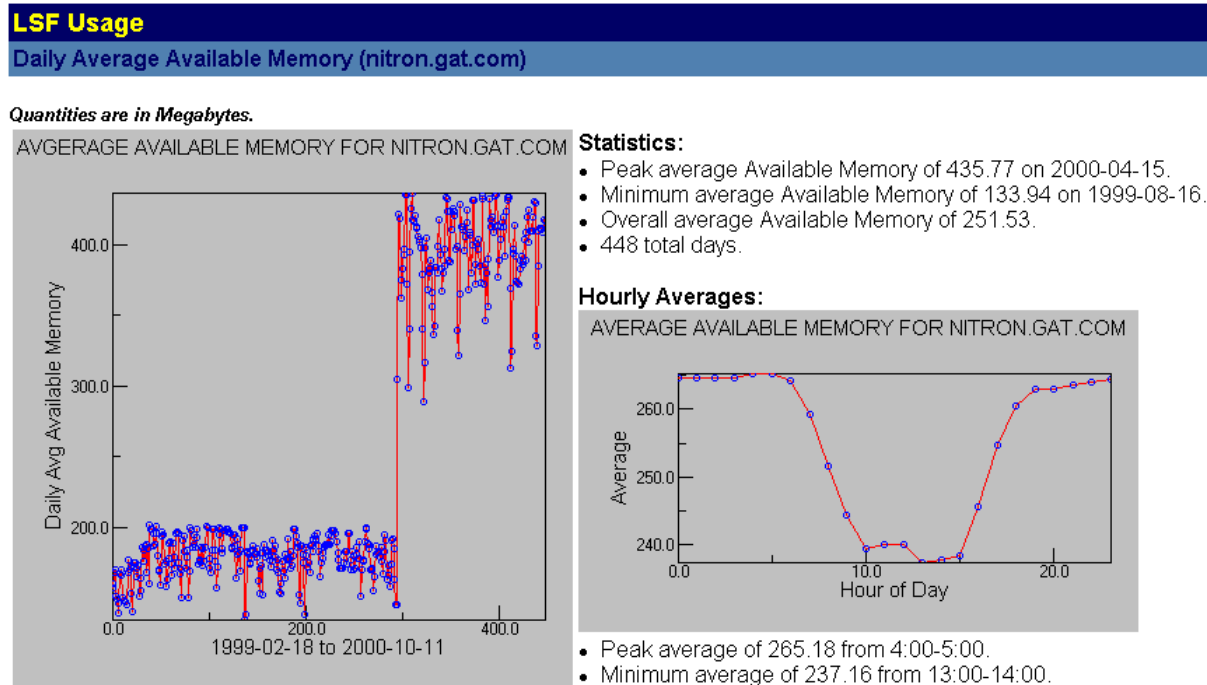
- Do these users know about the tool?
- Do they find the program awkward to use?
- Does the tool fail to meet their needs?

● Find out if a tool is not being used.

- Does the tool fail to meet a need, or is it just under-advertised?
- Is it time to stop supporting an old tool?

LSF COMPUTER CLUSTER USAGE IS MONITORED

- Decision makers can monitor LSF cluster usage to identify problems.
- Rapid, simple plots of benchmarks such as CPU usage and available memory.
- Highlight trends in usage.
- This example shows available memory on one machine before and after an upgrade.
 - The upgrade appears to have met user needs.
- Available memory remains high after the upgrade.



FUTURE PROJECTS

- **Continue encouraging users to add tables to D3DRDB.**
 - There are already a few user tables.
 - There is room for more.
- **Move legacy databases (Yoka) to D3DRDB.**
 - Confinement, stability, threshold, and edge databases must be moved.
- **Continue the Integration of Netscape web browsers into tools.**
 - Netscape sessions will be served from a new Linux server.
- **Populate summaries table with old shots.**
 - Old shot data is currently stored on magnetic tape.
- **Create a code modification database.**
 - Users get automatic notification of program updates.
 - No need to send broadcast emails whenever code is updated.
 - Users are notified of modifications when they use a tool.
- **Explore object-relational databases.**
 - Support complex objects rather than simple data types.
 - Allow queries such as “find all similar temperature profiles.”