



ENHANCED COMPUTATIONAL INFRASTRUCTURE FOR DATA ANALYSIS AT THE DIII-D NATIONAL FUSION FACILITY

Presented by

D.P. Schissel

for the DIII-D National Team

Presented to

26th EPS Conference on Controlled Fusion & Plasma Physics

June 14–18, 1999

Maastricht, The Netherlands

ACKNOWLEDGMENTS

- **The Data Analysis Applications Group**
 - Qian Peng, Jeff Schachter, Dave Schissel from GA
 - Ted Terpstra from PPPL
- **DIII-D Computer Staff**
 - J. Freeman, K. Keith, B. McHarg, C. Parker, T. Warner
- **The DIII-D User Community**
- **From LLNL — Tom Casper, Bill Meyer, and Jeff Moller**
- **From MIT/C-Mod — Martin Greenwald, Tom Fredian, and Josh Stillerman**
- **Supported by U.S. DOE Contracts DE-AC03-99ER54463 and DE-AC02-76CH03073**

DIII-D PROGRAM COLLABORATORS

National Laboratories

ANL
INEL
LANL
LLNL
ORNL
PNL
PPPL
SNLA
SNLL

Industry Collabs

CompX
CPI (Varian)
GA
Gycom
Orincon

Universities

Alaska
Alberta
Cal Tech
Chalmers U.
Columbia U.
Hampton U.
Helsinki
Johns Hopkins U.
Lehigh
MIT
Moscow State U.
RPI
U. Maryland
U. Texas
U. Wales
U. Washington
U. Wisconsin
UCB
UCI
UCLA
UCSD

International Laboratories

Academia Sinica (China)
Cadarache (France)
CCFM (Canada)
Culham (England)
FOM (Netherlands)
Frascati (Italy)
Ioffe (Russia)
IPP (Germany)
JAERI (Japan)
JET (EC)
KAIST (Korea)
Keldysh Inst. (Russia)
KFA (Germany)
Kurchatov (Russia)
Lausanne (Switzerland)
NIFS (Japan)
Troitsk (Russia)
Southwestern Inst. (China)

Tsukuba U. (Japan)

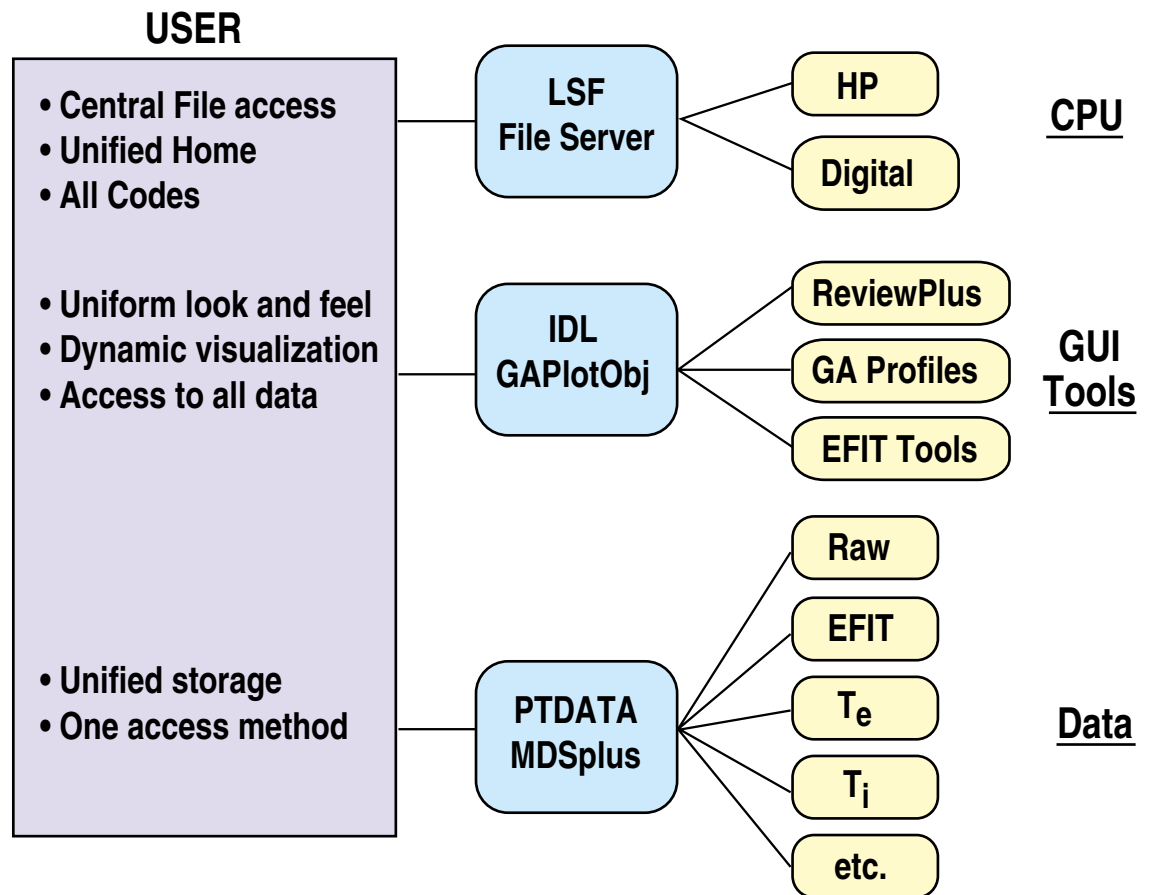
COLLABORATOR SUPPORT IS FUNDAMENTAL TO EFFICIENT UTILIZATION OF THE DIII-D FACILITY

- Two-thirds of onsite scientists are collaborators
- 150 of 300 users off-site collaborators
- Increasing number of collaborators places more demand on computing infrastructure
- In this presentation:
 - The analysis needs of the DIII-D Research Team
 - What we are doing to meet them
 - Future plans

INCREASE DIII-D DATA ANALYSIS THROUGHPUT AND DATA RETRIEVAL RATE BY EASE OF USE

- Underlying philosophy is uniformity

- Look and feel of GUI tools
- Access methods to analyzed datasets
- Access to existing computer power



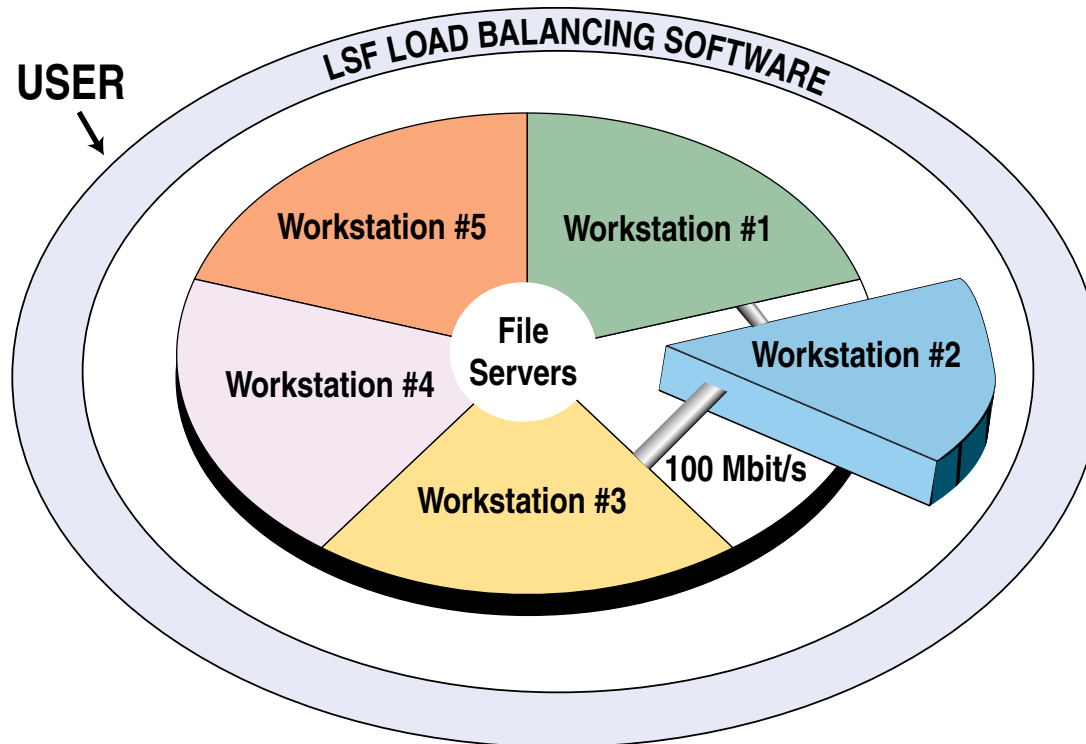
COMPUTATIONAL SUPPORT OF THE DIII-D RESEARCH TEAM

- **Computer**
 - Efficiently load balance onsite CPUs
 - Leverage use of remote CPUs
- **Access to data**
 - 7 day-24 hour (24/7) remote access via MDSplus, PTDATA
 - And a relational database
- **Tools to view and analyze data**
 - Easy-to-use GUIs (IDL) with ample documentation
 - Tools run from remote CPU
- **Communication with other scientists**
 - Remote meetings and experimental operations

Much of this plan is in place now. Future work will build on current activity

LOAD BALANCING SOFTWARE PROVIDES SUBSTANTIAL COMPUTATIONAL POWER TO ONSITE USERS

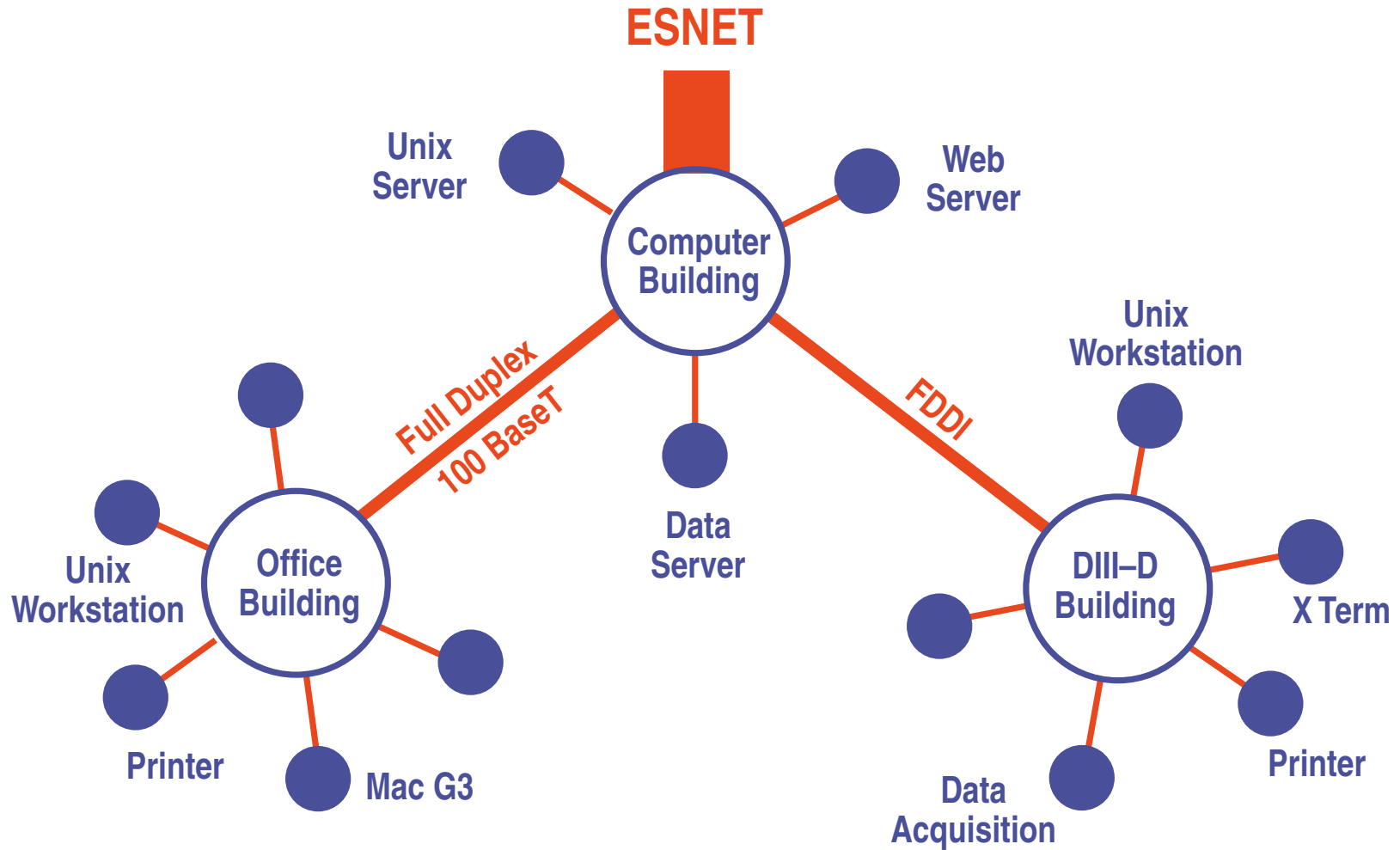
- Heterogeneous Unix environment with a large server and numerous workstations
- Platform Computing's LSF Suite performs interactive load sharing
- Cost effective for sharing CPUs (GA, LLNL, ORNL, PPPL) and commercial software
- Central file servers for data and user files with fast network access



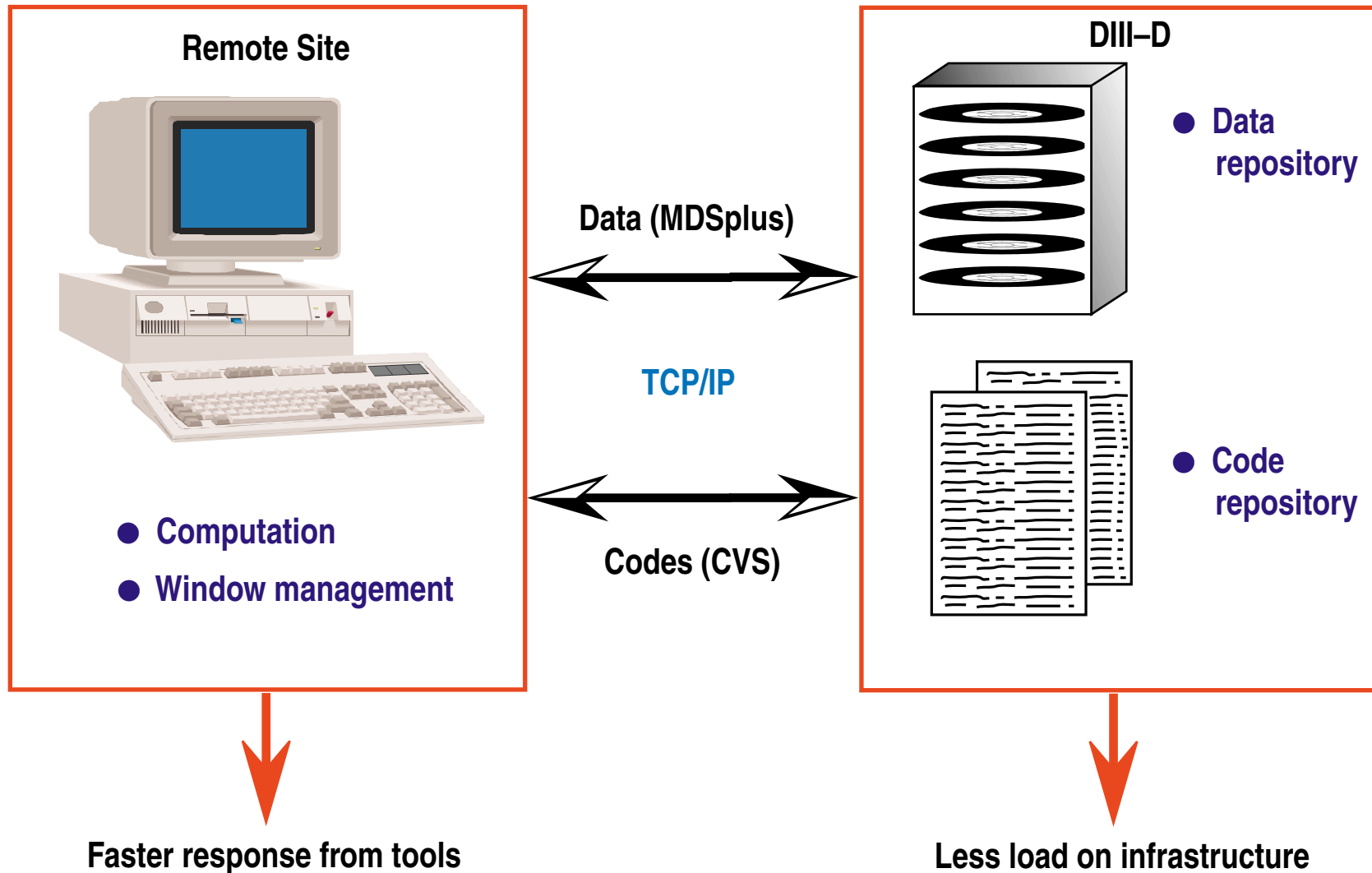
COMPUTER HARDWARE FOR DATA ANALYSIS

- **Large Unix Server – HP 9000 T600 Enterprise Server**
 - 3 CPUs with a total of 1GB memory to a total of 44.7 SPECfp95
 - Over 350 accounts including onsite and off-site collaborators
- **HP 9000 based Unix workstations**
 - C180 and a 735/125 dedicated to tokamak operations
 - Another C180 and numerous 735s available
- **DEC 433au Unix personal workstations**
 - 5 machines placed on the desktop but in a shared environment
 - Each workstation has a speed of 18 SPECfp95
- **DEC AlphaServer 4000 5/300 running OpenVMS**
 - 13 SPECfp95 available to all onsite and off-site collaborators
 - An OpenVMS Personal Workstation 433au running OpenVMS for legacy code

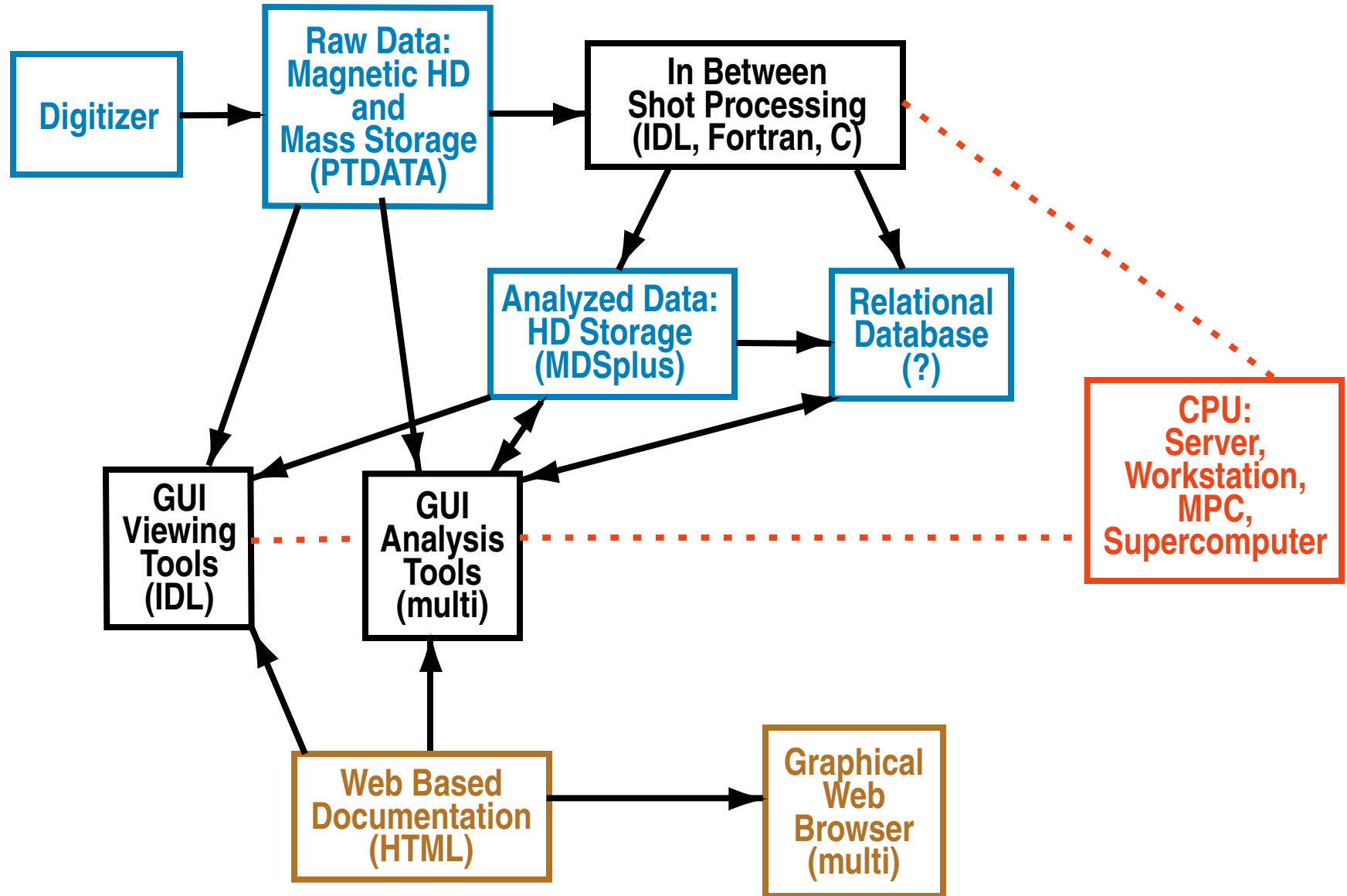
THE UPGRADED STAR NETWORK TOPOLOGY HAS AIDED UNIFICATION OF THE COMPUTING ENVIRONMENT



REMOTE COMPUTING BENEFITS OFF-SITE AND ONSITE RESEARCH TEAM MEMBERS

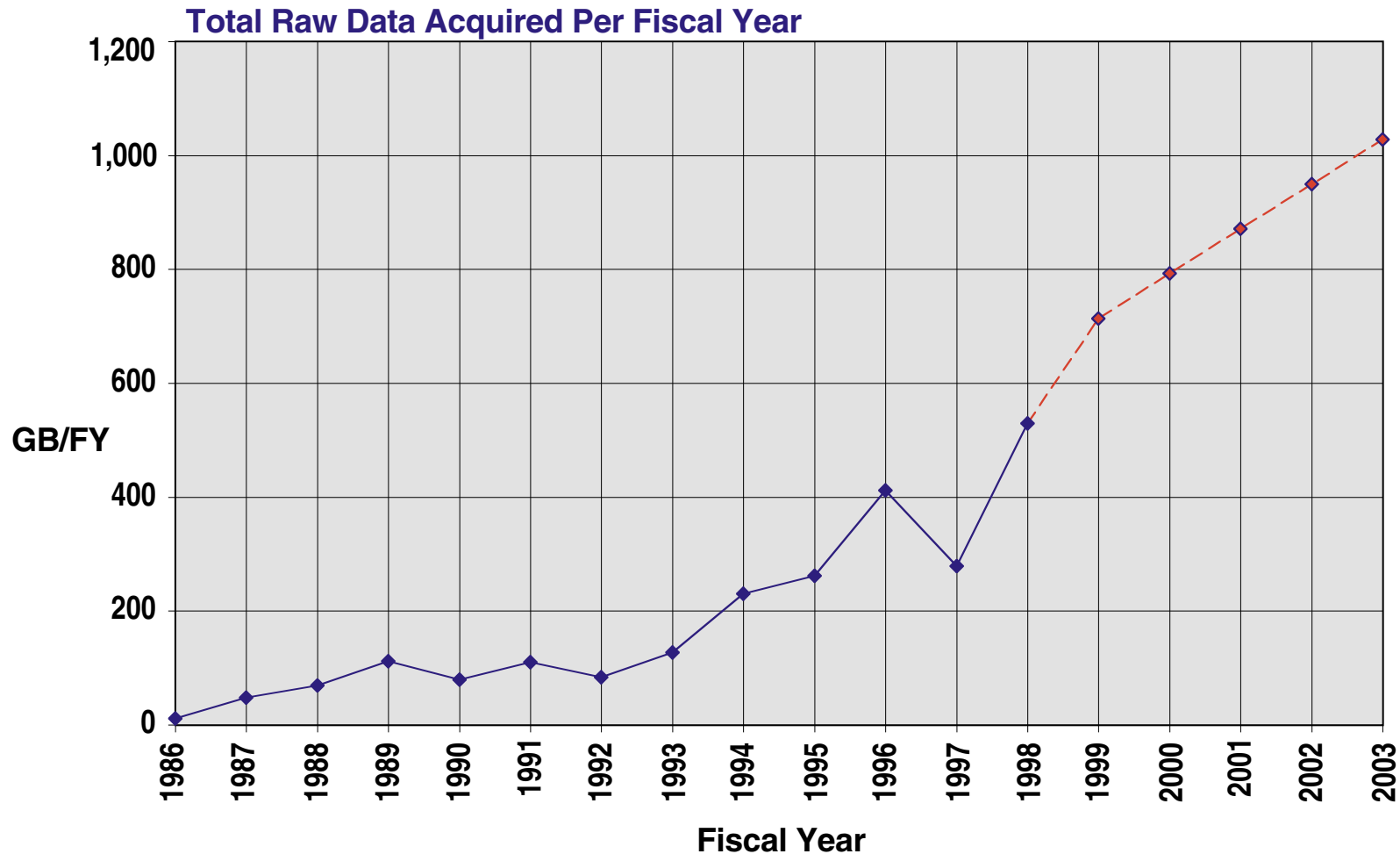


DIII-D DATA FLOW CHART



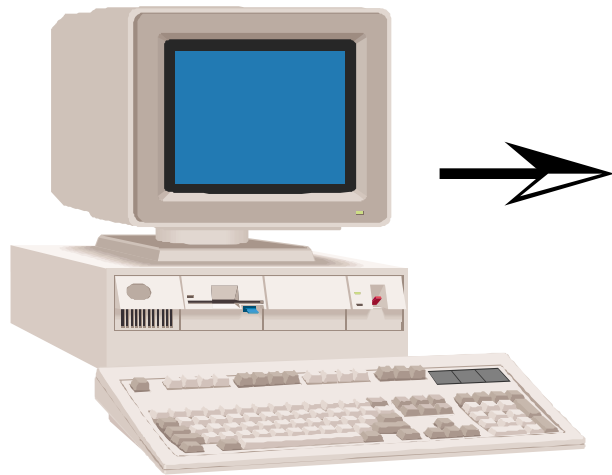
RAW SHOT DATA FILE SIZE CONTINUES TO INCREASE

- More diagnostics, faster digitization rates, potentially longer pulse lengths
- Assume 15 weeks in FY98 and 18 weeks thereafter



ALL DIII-D DATA AVAILABLE AROUND THE CLOCK

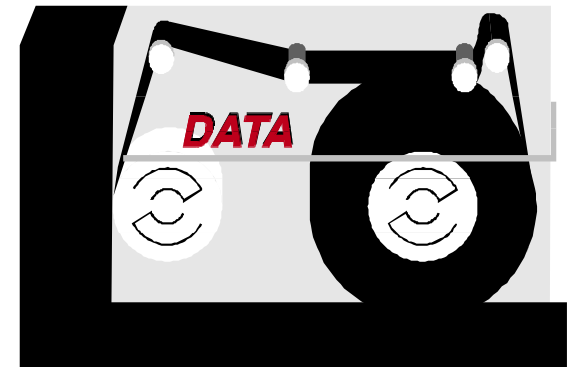
- All raw shot data (PTDATA) automatically available (24/7) for data mining
 - 720 GB interactively (15–30 sec minimum), ≈9000 raw compressed shots
 - 2.4 TB on a 2–5 minute minimum time scale per request
 - Present needs are 1 TB with about 0.3 TB added per year
- All analyzed shot data (MDSplus) on magnetic hard drive



Sun Ultra-1 Clone
100 GB Magnetic-Raid 5



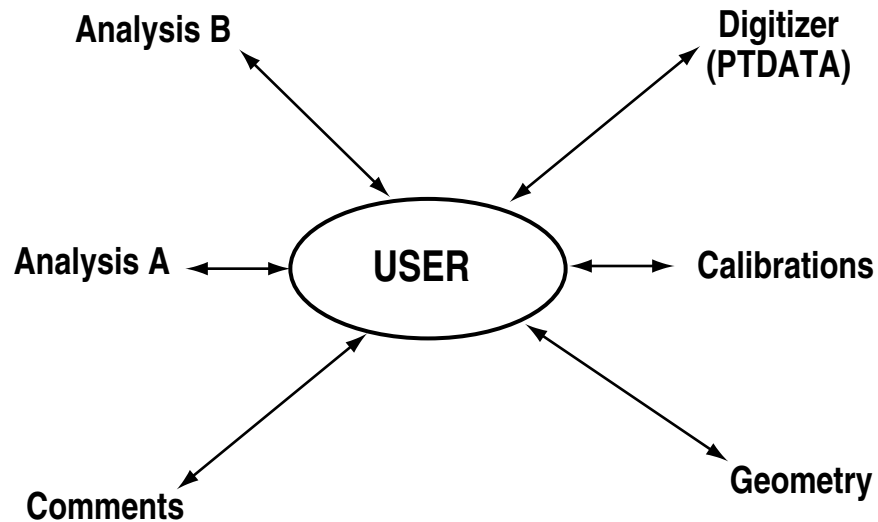
HP Magneto-optical 600fx
620 GB from 238 platters
2.6 GB/platter, 6 drives



ATL DLT 7000 drives
2.4 TB from 68 bays
35 GB/tape, 4 drives

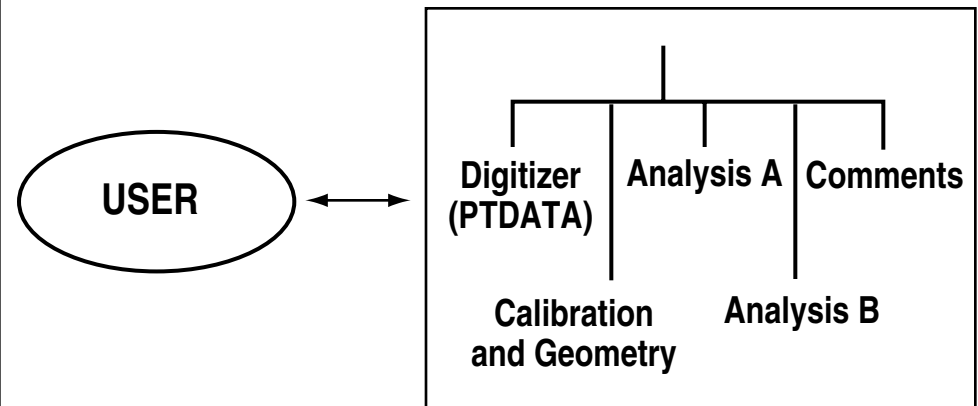
MDSplus UNIFIES DATA ACCESS

Conventional Storage



- Separate interface for each data type
- Must know data format and file location
- Data and context stored separately
- **Hard to share results**

MDSplus



- One interface to many data types
- Only need location of data in tree
- Store all relevant information
- **Remote exploration of data productive**

MDSplus HAS PROVIDED A UNIFIED DIII-D ENVIRONMENT FOR BOTH RAW AND ANALYZED DATA

- Utilized to unify data storage and retrieval
 - All 1998 shots and selected older ones for a total of 4730
 - PTDATA can be accessed through MDSplus
 - MDSplus data retrieval integrated into existing and new tools
- A wide variety of data has been stored
 - Control room EFITs
 - CER and Thomson scattering data only available through MDSplus
 - Confinement related quantities, impurity density, auxiliary heating
- DIII-D now operating with Unix MDSplus server
 - Approximately 38 GB of data is being served
 - Digital Unix AlphaServer 800 with 100 GB of RAID 5
 - Unix port completed March 1999 (GA, LLNL, MIT, PPPL)
 - Will be exercised fully during next June operations period

DATA USAGE AND PUBLICATION POLICY HAS BEEN CREATED TO CLARIFY A COLLABORATOR'S RESPONSIBILITY

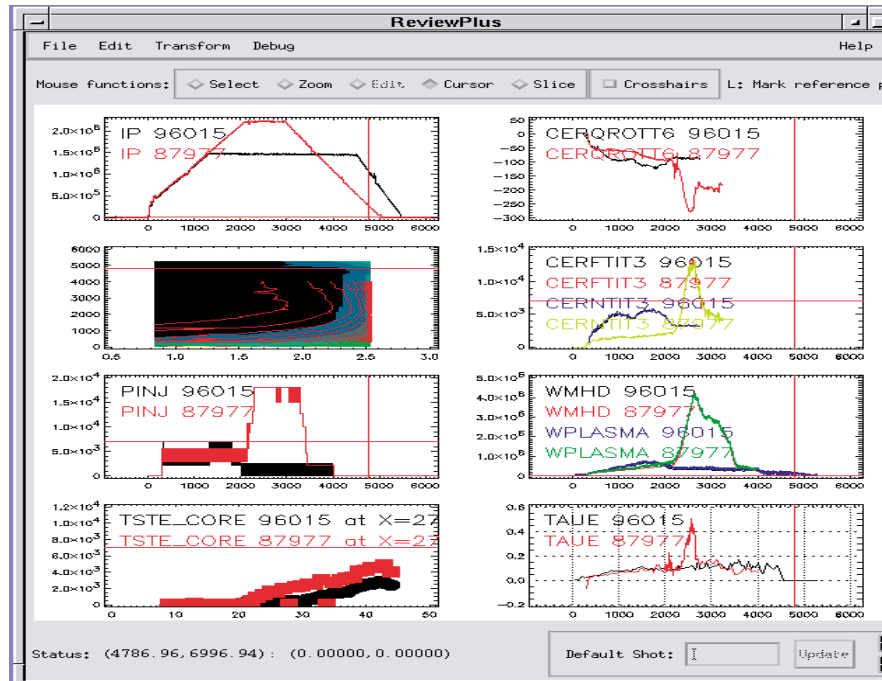
- **DIII-D raw and analysis data available without a local DIII-D computer account**
 - For PTDATA access DIII-D can supply the software
 - For MDSplus access the MIT group can supply the software
 - MDSplus data access security is by username and Internet address
- **Collaborators are offered full access to DIII-D data as it is collected and analyzed**
 - No data will be supplied without signing the policy agreement
- **Presentations of DIII-D data at conferences and workshops**
 - Require approval of DIII-D Director
 - For major conferences a rehearsal presentation is expected
- **The DIII-D program peer technical review is required for all papers**
- **Papers will be posted on the Web either at DIII-D or on the collaborator's site**

A NEW RELATIONAL DATABASE ALLOWS THE SCIENTIFIC COMMUNITY TO MINE FUSION DATA

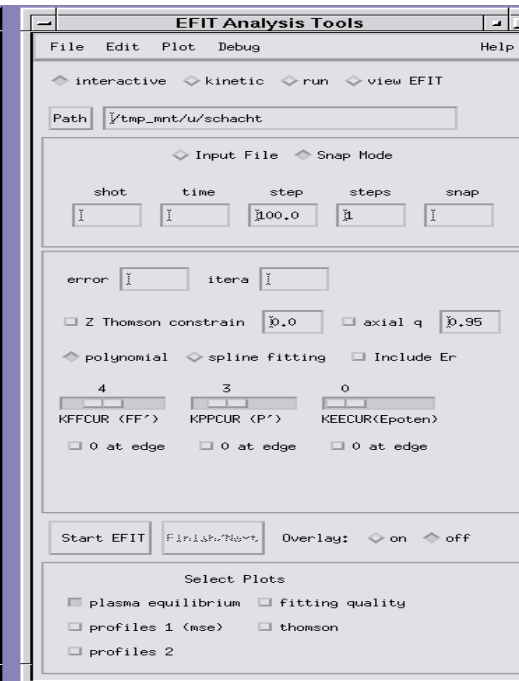
- **Desired user functionality**
 - “Broad & shallow” containing summary data from all tokamak pulses
 - “Highly Analyzed” containing publication quality data
 - “Special Purpose” containing user specified information
- **GUI tools connected to database engine**
 - Relational queries from the Web, Fortran, C or IDL
 - Multi-platform clients
 - Independent of database engine
- **Combined analysis of different tokamak databases will be straightforward**
 - Similar to the sharing of our viewing and analysis tools
 - The development of these tools must be a coordinated effort

GRAPHICAL USER INTERFACES SIMPLIFY DATA VIEWING AND ANALYSIS

ReviewPlus



EFITtools



- Same "look and feel" in all GUIs
 - Use GA-developed object oriented IDL based graphics library (GAPlotObj)
- GUI data viewing/analysis tools: ReviewPlus, EFITtools, GAprofiles
- GUIs simplify use of FORTRAN analysis codes
 - EFIT, ONETWO
- Easier for collaborators to learn tools

WEB BASED DOCUMENTATION BRINGS CRITICAL INFORMATION TO THE ONSITE AND REMOTE COLLABORATOR

- **Computer code documentation**
 - By the scientific research staff
 - By the computer scientist staff
- **Data documentation**
 - Includes both raw digitizer data and analyzed data repositories
- **One source, multiple paths to reach the source**
 - Direct from Web browser
 - From within a tool activating the web browser

TOOLS RUN ON REMOTE CPUs

- **Can now run ReviewPlus, EFITtools, GProfiles at LLNL, MIT, PPPL**
 - Retrieve DIII-D data via MDSplus
 - Any site with IDL and MDSplus client can use
 - Some tools will run on MacOS, Win95/NT
- **EFIT analysis between pulses is distributed for faster processing**
 - EFIT analysis on one shot shared by computers at DIII-D and LLNL

IDL IS LANGUAGE OF CHOICE FOR GUI TOOLS REQUIRING GRAPHICS

- **Benefits of IDL**
 - Extensive, powerful interactive graphics capabilities
 - Rich, easy to use language
 - Platform independent
- **Drawback: remote collaborators must purchase a license**
- **Public domain software (Java, Python) does not have required graphics**
 - Cost of implementing interactive features outweighs cost of IDL license
- **Adapt as software technology evolves**
 - e.g., A more robust Java with required graphics

COLLABORATIVE ANALYSIS CODE DEVELOPMENT EFFICIENTLY UTILIZES EXISTING COMPUTER SCIENCE RESOURCES

- **Good history of collaborative**
 - TRANSP from PPPL
 - MDSplus led by MIT
 - EFIT from GA
 - Distributed EFIT computing from LLNL

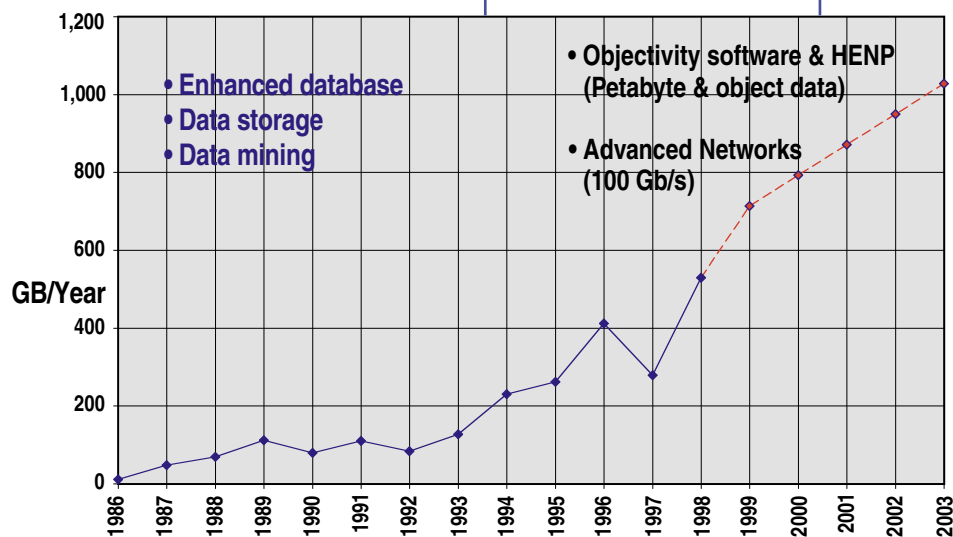
- **Exploring new areas of collaboration (GA, MIT, LLNL, ORNL, and PPPL)**
 - IDL-based transport analysis results display tool
 - IDL-based data preparation software
 - Name translation service
 - Run management database
 - Database analysis applications and database tool kit

DIII-D DATA ANALYSIS COLLABORATION IS ADVANCING SCIENTIFIC SIMULATION

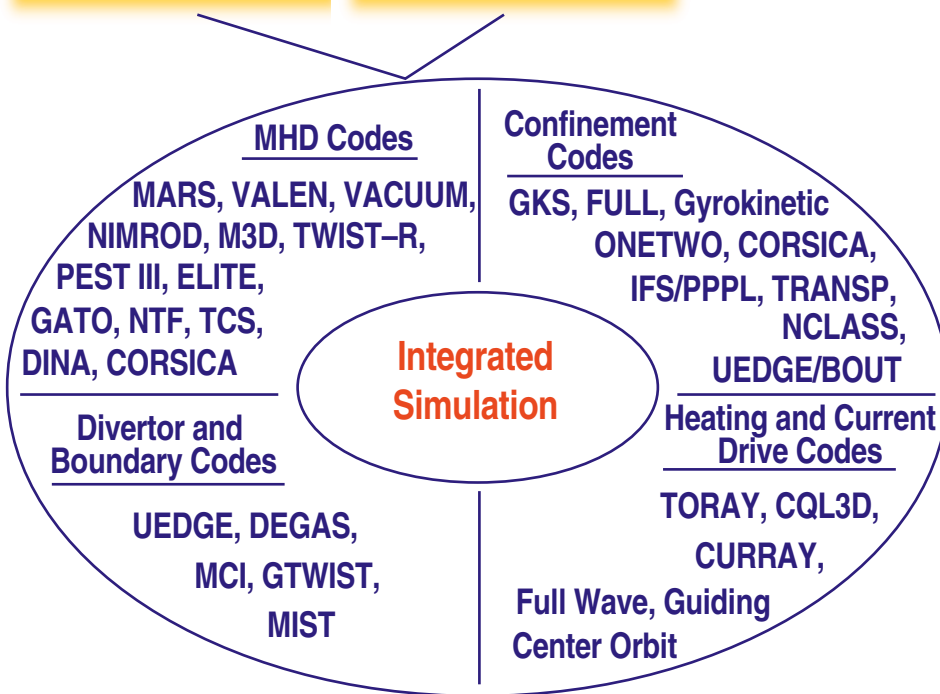
- MIT MDSPlus: a common language database management system (Alcator C-Mod, DIII-D, NSTX)

Scientist
Visualization
Tools

- Transparent data and code sharing
- Aim is Predictive capability



DIII-D Raw Data Size per Year



Collaborative Codes Used to Model DIII-D Data



REAL-TIME REMOTE COMMUNICATION AT DIII-D

- **Support meetings and tokamak operation**
- **Remote meetings**
 - ShowStation Ip for viewgraphs
 - Conference calls for audio
 - Video conferencing for broad view video
- **DIII-D and C-Mod operated from LLNL — software**
 - Physics and shape control
 - Audio/video with MBone
- **Tokamak operations**
 - How does the off-site physicist communicate with the session leader?
- **Present tokamak operations**
 - C-Mod electronic logbook — well received
 - Telephone

SUMMARY: DIII-D NATIONAL FUSION FACILITY IS AN ACTIVE, FRUITFUL COLLABORATION

- **Onsite and remote computing supported**
 - Efficiently utilize onsite and remote CPUs
 - 24/7 remote data access via MDSplus, PTDATA
 - Easy-to-use GUIs for data viewing and analysis
 - Remote real-time communication with onsite DIII-D staff
- **Future will be built on current success**
 - Remotely accessible relational database and tools
 - Cooperative use and continual development of tools
 - More effective real-time interaction

