

DATWeb

An Archive-Backup Management Tool at Leibniz Rechenzentrum

Werner Baur, Bill West

I. Background

Once upon a time, the central TSM administrators at the Leibniz Rechenzentrum (LRZ) used a collection of Perl scripts and a large text file manipulated with vi to keep an overview of the status of TSM elements distributed over the 10 servers running on two host systems. Here is an extraction from this configuration text file representing the information about one local TSM administrator and his domain:

=====

Ansprechpartner

Name: Rudolf Gabler
Telefon: 2180-6019
Email: rug@usm.uni-muenchen.de
Uni: LMU
Institut: Sternwarte
Kennzeichen am LRZ:

Technisches

Nodename(n): hal1, hal2, hal3, hal4, hal5, hal6, hal7, hal8, hal9, hal10,
hal11, hal12, hal13, hal14, hal15, hal16, usms01, usms02,
usmu01, usmu02, usmu03, xm00, hal20, icc, hal21, hal22, kutepov,
bib, loge, adt.usm, majestix.usm, miraculix.usm, hal23, musca.usm,
volans.usm, apus.usm, vela.usm, carina.usm, crux.usm, fox.usm,
fourier.usm, capella.usm, dwarf.usm, pollux.usm, castor.usm, spica.usm,
riegel.usm, schimansky.usm, wotan.usm, mizar.usm, nekkar.usm,
sirius.usm, polaris.usm, algol.usm, arktur.usm, vega.usm, regulus.usm,
birnbaum.usm, rubin.usm

Plattform(en): ULTRIX, SunOS, Win3.11, Win95, DEC UNIX, Mac, Linux

Schedule ja/nein: ja
Geschaetzter Datenumfang: 40GB, 20GB
Geschaetzte durchschnittl. Filesgroesse:
Verwendung:
Server: s1, s3, s2, s5, s7, s10, s10
Besonderheiten: grossenteils von s3 auf s1 exportiert!
History:

LRZ:

Eingetragen von: Baur (Strunz) (Hufnagl) (Bender) (Hufnagl) (Hufnagl) (Hufnagl) (Hufnagl) (Hufnagl)
Eingetragen am: 21.09.95 (26.03.97) (17.06.97) (18.12.97) (30.06.98) (04.12.98)
26.01.99) (12.05.99) (11.11.99) (17.11.99) (27.04.00) (17.05.00)
(06.06.00) (13.10.00) (21.12.00) (21.03.01) (30.4.01)

Betreuer:

=====

Actually the most important tool used was the extremely good memory of our operators who kept track of which nodes were on which one of the ten TSM servers. But at the beginning of the year 2001 the number of nodes was about twelve hundred, the number of local TSM administrators was over two hundred and fifty and the organizations using TSM were then was around two hundred. The correlation of information was "loose", not "linked"!

II. DATWeb's Beginning

The first efforts to aid our operators were to build databases where information was linked and maintained only once where possible. The correlation between the TSM database views of organization had drifted from that list of organizations maintained by the LRZ user administration group's view. The file of node and administration information was primarily used as a basis for cut and paste into TSM commands issued in command line form.

The maintenance of this 260,000-character file was maintained with ironhanded discipline but still it developed many different representation of the same object.

The first effort resulted in breaking down this configuration file data into a node list, and administrator list managed as linked web databases by the central TSM administrators and the organizational data from the database maintained by the user administration group in another department.

These lists were displayed and maintained through a web browser interface, which we named DATWeb. The data and supporting CGI Perl scripts were placed on a separate, run-of-the-mill PC, not one of the TSM hosts.

Most of the statistically oriented Perl scripts used activity and occupancy data that is created daily by CRONTAB initiated batch TSM query jobs and stored into files under afs and accessible without subsequent impact of the two TSM hosts. These occupancy and activity files are later massaged into a compact comma-separated-values (csv) form, commonly used as an interchange format used by other applications such as EXCEL, and are stored in the DATWeb's local LINUX file system.

The primary objective of this effort was simply to provide a more consistent view for the central administration of TSM and ease the effort to access this information but this did little for the local TSM administrators. There were special processes for saving the administrator data in the contact field of the TSM node database, a sort of back up for DATWeb's administrator list. There were procedures to check the consistency between the DATWeb database and the server databases. Lists of nodes, which had not been used within the last six months, which had expired passwords, or which had not been accessed with the last six months, are easily maintained to give a degree of control over the actuality of our environment.

The first extension to DATWeb was to use the information about occupancy and activity to inform the local TSM administrators, who's numbers were steadily increasing, about their systems. This included warnings about questionable nodes which had not been accessed for some time, or which had expired or nearly expired passwords. Here the objective was to "indirectly" encourage local TSM administrators to inform the central operators of nodes that were no longer needed thus improving the quality of the backup/archive storage. It also "indirectly" highlighted the inactive local administrators. In the highly distributed and loosely connected source of information about the status of user organization, this e-mail based newsletter resulted in better control. Non-existent administrators were exposed.

The second extension was to allow the local administrators to inform themselves about the status of their systems in a more comfortable overview. Basically the information contained in the monthly e-mail was available interactively with a browser interface and did not require knowledge of the TSM commands or server - node topology. A consistent registration of administrators and nodes was also included in the first revision. This reduced the need for the central operators to check for such things as uniqueness of node names or scheduler options. In order to allow for greater flexibility of moving nodes from one server to another, we require unique node names. This has never lead to a naming problem yet, even with now over 2000 nodes and an increase of nearly 80 nodes per month in the last year. This only reflects the individuality and inventiveness of our users. Occasionally, a request to provide a special report or overview of the collected data was made, such as ranking organizations by usage and activity, displaying growth of backup/archive files and space and of nodes supported. These reports have been integrated and are available on an as needed basis. that is, administrators see only those elements which they "own".

DATWeb allows four levels of access, guest, local administrators, central administrators, and the DATWeb administrator:

1. GUESTS - The guests may access DATWeb to get an overview of the TSM environment at LRZ and may request to be registered as a local administrator. No password is required for access.

2. LOCAL ADMINISTRATORS -The local administrator are required to login with their full name and their LRZ master account number and may view data dealing exclusively with their managed nodes and their servers. This interface allows for a simplified registration of new nodes. The LRZ master account number is maintained by the user administration at LRZ, which is primarily responsible for controlled access to the LRZ computing resources.

3. CENTRAL ADMINISTRATORS - The central administrators have an overview of all data in relation to the TSM environment. This level provides a simplified interface to register new administrators and new nodes. Additional procedures are provided to check for consistency between the TSM view of the environment and the DATWeb maintained view. A simple monitoring of the servers' status, status of the databases, recovery logs, and volume (cartridges) usage is displayed. The access at this level is restricted by password maintained by the DATWeb administrator.

4. DATWeb ADMINISTRATOR - The DATWeb administrator is also restricted by password. This level is primarily concerned with software maintenance, and control of the periodic background jobs required. To test new code or corrections to existing modules, a version of DATWeb runs on a separate system from the production DATWeb machine. Once the DATWeb administrator is satisfied with a particular new implementation or correction set, the software is saved in the distributed file system and later installed on the production DATWeb. Periodic activities are monitored from this level and each job can be re-initiated from here should it be necessary.

All levels of access are maintained as sessions with DATWeb as a file indexed by the IP address of the user. Details of the current session are retained in this session file that remains active until the user explicitly logs out or is cleared after midnight or after a defined period of inactivity. The details kept in this session file are primarily there for control of the access information and for ease of use, for example, to remember the parameters or options last selected by the user.

The DATWeb implementation is strictly server-side oriented. Browser testing has been primarily directed to the broadly used browsers such as Internet Explorer 6.0 and Netscape 7.02 /Mozilla 5.0. Some testing has been done with Opera 7.11.

III. Current Status

DATWeb consists of 237 scripts totalling 30,000 lines of Perl/HTML code. DATWeb produces over 70 reports. These reports and means for maintaining the support data reflect the particular requirements of the Leibniz Rechenzentrum DAT group. An important goal for the DATWeb implementation is to make the support of this product simple, to minimize support for future maintenance. The use of Perl, Apache, with simple database functionality reflects the goal to minimize the support requirements. The general application of DATWeb to other sites was considered for the sake of discipline, but not considered essential. Still, there are a number of localized solutions and shortcuts taken in the implementation.

I.V. Future

There may always be the need to add functionality, such as new reports, but the goal is to stabilize the future development of DATWeb, that is, allow it to approach zero. LRZ is project oriented, but maintenance of local developments is not one of these projects. DATWeb is reaching a degree of complexity that may require some consolidation and restructuring of its modules. The acceptability of this interface by the local administrators is both positive and encouraging. The initial goal of increased quality of information for operational and administrative purposes has justified the effort, but we will continue to explore other tools available for their application in our environment. Future developments needed to encourage effective use of our back up and archiving environment, through local accountability and consequent monitoring and control, in the words of Valdimir Ilyich Lenin, "trust is good, but control is better", are most likely to be the grounds for further development of DATWeb or related tools.

Appendix 1 List of Displays in DATWeb

Access Levels*	Object	List of Major Displays/Reports
----------------	--------	--------------------------------

* All refers to all levels (guests, local and central TSM administrators)

Administrators refers local and central TSM administrators)

Central refers to central TSM administrators

DATWeb administrator reports are not listed here

1	All	Overview	Overview of Archiving/Backup System at Leibniz Rechenzentrum
2	All	Help	Help for the DATWeb Interface
3	All	DATWeb session	Current session parameters and their settings
4	All	Overview	Operational Overview
5	All	Overview	Recent Activity
6	All	Overview	Monthly Development of ABS Activity at LRZ in the Year 2002
7	All	Interrupts	Display all interrupts for the year 2002!
8	All	Tape	Overview of the nnn,nnn Tape Mounts in the Year 2002
9	All	Tape	Tape Drive Utilization in the Year 2002
10	All	Tape	Details of Parallel Tape Usage per Server at LRZ for year 2002
11	All	Configuration	Backup and Archiving Hardware Configuration of 2 Hosts at LRZ
12	All	Configuration	Graphic Diagram of ABS Hardware Configuration
13	All	Interrupts	Display all interrupts for server Sx since January 200x until now! (Graphically)
14	All	Interrupts	Display all interrupts for server Sx since January 200x until now! (Details in table format)
15	Administrators	Administrator	Administrator: Xxxxx Yyyy
16	Administrators	Administrator	Overview of this month's (xxxx) ABS Activity for the x nodes administered by Xxxxx Yyyy
17	Administrators	Administrator	ABS Growth for Administrator Xxxxx Yyyy with account x999
18	Administrators	Node	Node: ZZZZZ
19	Administrators	Node	Overview of this month's (xxxx) ABS Activity for the node ZZZZZ
20	Administrators	Node	Distribution of Activities (for node ZZZZZ)
21	Administrators	Node	ABS Growth for node ZZZZZ
22	Administrators	Node	Latest session data for node ZZZZZ
23	Administrators	Node	TSM's detailed information about node ZZZZZ on server Sx
24	Administrators	Server	Display of sessions for server Sx on Tuesday, July xxth, 200x
25	Administrators	Sessions	xxx sessions for server Sx on Tuesday, July xxth, 200x
26	Central	Client versions	Distribution of the xx TSM client versions over all xxxx nodes!
27	Central	Client versions	Overview of distribution of the xxx TSM client versions over all xxxx nodes!
28	Central	Volumes	Volume Statistics (ABS Volume Usage at LRZ for all xx,xxx Volumes, Distribution over Servers)
29	Central	Overview	Distribution of System vs. Non-system Files
30	Central	Overview	Distribution of the TSM clients by license type!
31	Central	File space type	Distribution of Node's Backup File Space by File System Type
32	Central	Overview	Organizations Ranked by Archive Space Used
33	Central	Overview	Organizations Ranked by Total Space Used
34	Central	Overview	Organizations Ranked by Backup Space Used
35	Central	Overview	Organizations Ranked by Number of Nodes Used
36	Central	Overview	Organizations Ranked by Data Transferred
37	Central	Overview	Organizations Ranked by Number of Objects Transferred
38	Central	Overview	Organizations Ranked by Number of Sessions
39	Central	Overview	List of Administrators with TSM Usage to Review
40	Central	Overview	List of xxx nodes not accessed within the last 180 days!
41	Central	Overview	List of the xxx expired nodes and their space usage!
42	Central	Overview	List of all xxx administrators
43	Central	Node	Register a new node based on existing node's data!
44	Central	Organizations	List of all xxx organizations using DAT
45	Central	Organizations	List of all xxx organizations using DAT sorted by their LRZ master account number
46	Central	Organizations	List of yyy keywords within organization names using DAT!
47	Central	Servers	Status of Servers
48	Central	Nodes	Overview of resources used by expired, expiring, idle and zombie nodes
49	Central	Overview	Definition Deviations in the TSM Servers and DATWeb Node Database
50	Central	Overview	Deviations between the TSM node definitions and the DATWeb node database
51	Central	Overview	List of xx nodes that have neither backup nor archived data ... and have not been accessed for 180 days!

52	Central	Overview	List of the xx expired nodes and their space usage!
53	Central	Overview	List of xx nodes with a password that will expire within the next 90 days (sorted by urgency!)
54	Central	Overview	Nodes found with an unknown local TSM administrator!
55	Central	Overview	List of the xx local TSM administrators without any nodes and who are also not auxiliary administrators
56	Central	Overview	Nodes found with an unknown organization!
57	Central	Servers	Servers' Data base Usage
58	Central	Server	TSM definitions for server Sx
59	Central	Servers	Storage Pool Migration within the last 7 Days on all Servers!
60	Central	Server	Display of all recent migrations of Storage Pool XXX.YYY.Z on Server Sx!
61	Central	Server	Display of details of recent migrations of Storage Pool XXX.YYY.Z on Server Sx
62	Central	Server	Overview of Migrations of Storage Pool XXX.YYY.Z on Server Sx over a 6-week period
63	Central	Server	Display xxx mounts for server Sx on Tuesday, July 29th, 2003
64	Central	Server	Display xxx mounts on server Sx on Tuesday, July 29th, 2003
65	Central	Server	Display overlap of the xxx mounts for server Sx on Tuesday, July xxth, 200x
66	Central	Server	Overview of ABS Activity for server over the week ending on Saturday, July xth, 200x
67	Central	Client versions	Overview of distribution of the xx TSM client versions over all xxxx nodes!
68	Central	Client versions	Distribution of the xx TSM client versions over all xxxx nodes!
69	Administrators	Activity	x Sessions for n Nodes on Server Sx over 1 Day, 2Days, 1 Week ending Wednesday, xth September 200x
70	Administrators	Events	x events for all nodes on server Sx since Monday, xth September 200x