To: Distribution

From: T. H. Van Vleck

Date: March 21, 1975

Subject: New Storage System Long Range Plans

INTRODUCTION

This memorandum describes the sequence of milestones which are expected to occur during the implementation of the new Multics storage system. The order in which some of these events occur may, of course, be different from that described here, and additional activities may be required as our insight into the problem develops.

OVERVIEW

Although the implementation strategy described in MTR-068 is still our plan, the list of phases described there has been expanded and the decision to release the new storage system with release 4.0 in June 1976 has been shown explicitly.

The following table shows the major phases of the implementation of the new storage system.

Phase		Date
I	Command Level One user at command level	May 75
II	Prototype Running Several users, old backup, one logical vol	June 75
III	Design Review Salvager, backup, mount/demount	Sept 75
IV	Installable System Run mini-service at CISL	Nov 75
V	Initial Installation at MIT One logical vol, no mount/demount	Feb 76
VI	Follow-up Installation at MIT Operational enhancements	March 76
VII	MR 4.0 Installed at MIT With mount/demount	April 76
VIII	Release MR 4.0	June 76
IX	Further Enhancements Very large configurations Administrative improvements	-

WORK COMPLETED

1. Statement of Problem.

   See MTB-017, dated November 73.

2. Preliminary Design.

   MTB-055, April 74
   MTB-060, May 74
   MTB-065, April 74
   MTB-095, June 74
   MTB-110, August 74
   MTB-167, February 75

3. Preliminary Task Schedule.

   See MTR-068, dated October 74, and this document.

CURRENT TASKS

4. Phase I: Command Level.

When this benchmark is reached, the system can be bootloaded from a Multics system tape, either cold or warm, come up to Initializer command level, and shut down. Only one disk need be used; but it will have a standard label, VTOC, and volume map. Paged I/O will be used for the VTOC. A new version of BOS is required to support the new configuration deck.

Target date: May 75.

FUTURE TASKS

5. Phase II: Demonstrable System.

When this benchmark is reached almost all functions of the current Multics work in the new system, although performance may be unacceptable. Since the VTOC is still accessed by means of paged I/O, 1K per volume of page table for VTOC image is wired, plus the 512 words per volume of volume map. No backup or salvager will be implemented. Although much more interesting in terms of function, this stage is not very difficult to accomplish because the system initialization path checks out almost all of the storage system.

Target date: June 75.

6. Study and Definition of Backup Problem

This activity will define the key variables to be optimized in subsequent backup design.

7. Design for Salvagers.

There will be three salvagers: the single-pack salvager, the directory tree structure salvager, and the tree-to-segment connection salvager.

8. Design for Backup.

This delsgn will present the long-term plan for the evolution of the system's backup capabilities.

9. Specifications for Salvagers.

It is very important that the salvager be will specified before coding is begun. This set of specifications may be in

the form of a PLM, or part of a PLM.

10. Specifications for Backup.

Once again this may be one or more PLM's. However backup is likely to consist of several different phases.

11. Implementation of New Backup.

The implementation of the new backup facilities is listed here as a task but may actually extend for an extremely long time, depending on the difficulty of implementing the functions specified in the new design.

12. 64-word I/O Facility.

This facility will be used to transport VTOC information and volume map data between core and the disk. Changes must be made to the disk DIM, and a sort of "mini page control" must be written for the management of the memory devoted to 64-word data.

13. New VTOC Manager.

The new VTOC manager will use the 64-word I/O facility. This change frees 1K per volume of wired core, and decrease the I/O channel time and latency time for requests for data in the VTOC.

14. Smaller VTOC Entry.

The interim VTOC entry in use up to this point has 256 words instead of 192, in order to simplify the code for the deactivation of 256K segments. This stage complicates the code but reduces the size of the VTOC by 25%.

15. Design for Volume Mount/Demount.

This task adapts the solution which will be used to manage tape reels to the management of disk packs.

16. Phase III: Design Review.

This review covers the design of the salvagers, backup, and the volume mounting and demounting modules.

Target date: September 75.

17. New Directory Locking.

This step creates a wired hardcore table with one entry per active directory. The directory lock and other per-directory information are kept in this table instead of in the ASTE.

This will allow the shrinking of the ASTE.

18. Performance Measurements

At this stage performance should meet or exceed the performance of the installed Multics. If it does not, we will have to find out why and fix it.

19. BOS Utilities for Pack Maintenance.

Programs to format the pack, write labels, set up VTOC, handle bad-track data.

20. Multics Utilities for Pack Maintenance.

Same functions as above, but done through the unstructured removable pack facility.

21. Specifications for Command Changes.

Many minor changes will have to be made to the command system. Quite a few of these can be done ahead of time if a document setting forth the standards for system commands and subroutines is published.

For example, it will be possible in the new storage system to have a segment which exists, and to which the user has access, but which is not currently on line. Some commands should treat this case as if the segment did not exist. Others should behave just as if the segment did exist. Still other commands must take a new action in this case: and this action should be specified well in advance.

22. Error and Exception Handling Improvements.

This task provides sensible system action for the cases of "no more VTOC entries on the volume" and "no more pages on the volume." The second case may be quite complicated if we decide that we must move the segment to another volume in the volume group which has sufficient room.

23. Single-Volume Salvager.

Makes the volume map and the file maps in VTOC entries be consistent. Validates the whole VTOC on a pack.

24. Directory Tree Salvager.

Cleans up the directory hierarchy, like the current salvager, but need only look at logical organization.

25. Tree-to-Segment Connection Salvager.

Checks that every segment in the tree actually exists and has a VTOC entry which points back to the branch. All unique ID's must match. This salvager may be able to splint damaged sections of the hierarchy, by deducing the names of lost directories and rebuilding them.

26. Shutdown and Emergency Shutdown Improvement.

Changes to the locking strategy will make the disk copy of the hierarchy always consistent. ESD must not wipe these pages out with inconsistent pages.

27. Interim Backup.

Modify current backup programs to dump and reload the logical volume ID for a directory. This simple change allows the current incremental/catchup/complete dumper to be used for backup until a new version is designed and built.

28. Run Mini-Service at CISL.

Until we actually run a "service" of some sort, we will not know what the performance is really like and what operational improvements are required. These learning activities, once begun, continue for the life of the system.

Target start date: Nov 75.

The availability of the salvager is likely to be the determining factor on this date.

29 Phase IV: Make System Installable.

Once system reliability and performance are acceptable, we will attempt to convince MIT management that the new storage system should be installed at MIT.

Target date: Jan 76.

30. Formalities of Submission.

This step covers filling out submission forms, auditing of all programs, running final performance runs, fixing last-minute problems, etc.

31. Phase V: First Installation at MIT.

Target date: Feb 76

32. Phase VI: Follow-on Installation at MIT.

Operational experience will lead us to make many improvements to the interface and behavior of the storage system. Performance measurements under actual load may also show use where to concentrate our programming effort in order to speed the system up; if these improvements are possible we will install them soon.

Target date: March 76.

33. Dynamic Mounting.

For this stage, some of the code in initialization which connects a pack to the storage system will be moved to permanent residence. Two highly privileged entries will be provided for the mounting of a pack. One will request the mounting of a single volume, while another will check the volume lable and make the volume usable.

34. Dynamic Demounting.

For the demounting of a volume, a privileged gate entry must be provided which will lock out further segment activations on the volume and then flush core and bulk store of all pages which belong on the volume by forced deactivations.

35. Command System Changes.

These changes are the ones specified in paragraph 21. In addition to the changes to handle new error and state conditions, the create_dir command must accept and check the new parameter which specifes the logical volume in which storage will reside, and the list and status commands must be modified to show this attribute.

36. Ring 1 Volume Mount Module.

Ring 1 programs must be installed to allow users to request the mounting and dismounting of a volume. These programs will call the privileged entries defined above.

37. Resource Control Module.

The ring-1 programs must be modified to interface with the system control process so that the operator may reply "yes" or "no" to the request to mount a volume, and so that the system can account for volume usage.

38. Volume Registration Module.

The solution adopted for tape reels must be adapted for control of access to disk packs.

39. Phase VII: Install MR 4.0 at MIT.

    Target date: April 76.

40. Phase VIII: Release MR 4.0

    Target Date: June 76.

FURTHER ENHANCEMENTS

41. Pageable Volume Maps.

Until this point the volume maps have remained wired in core while the volume is online. The cost of this strategy becomes prohibitive when very large configurations are used. The volume map for a DSU191 volume is 512 words, while the map for a MSU450 will be about 2K. This stage modifies the system to use the 64-word I/O facility to transport volume maps to and from core on demand. Careful design must be used so that performance is not degraded severely.

42. Keep Duplicate Copies of Selected Volumes.

Crucial volumes in the system should be maintained in duplicate; all modified pages will be written out to both devices. In a configuration which places the secondary copy on a different disk subsystem from the primary copy, the cost of maintaining two copies should be very low.

43. Automatic Use of Secondary Volume on Error.

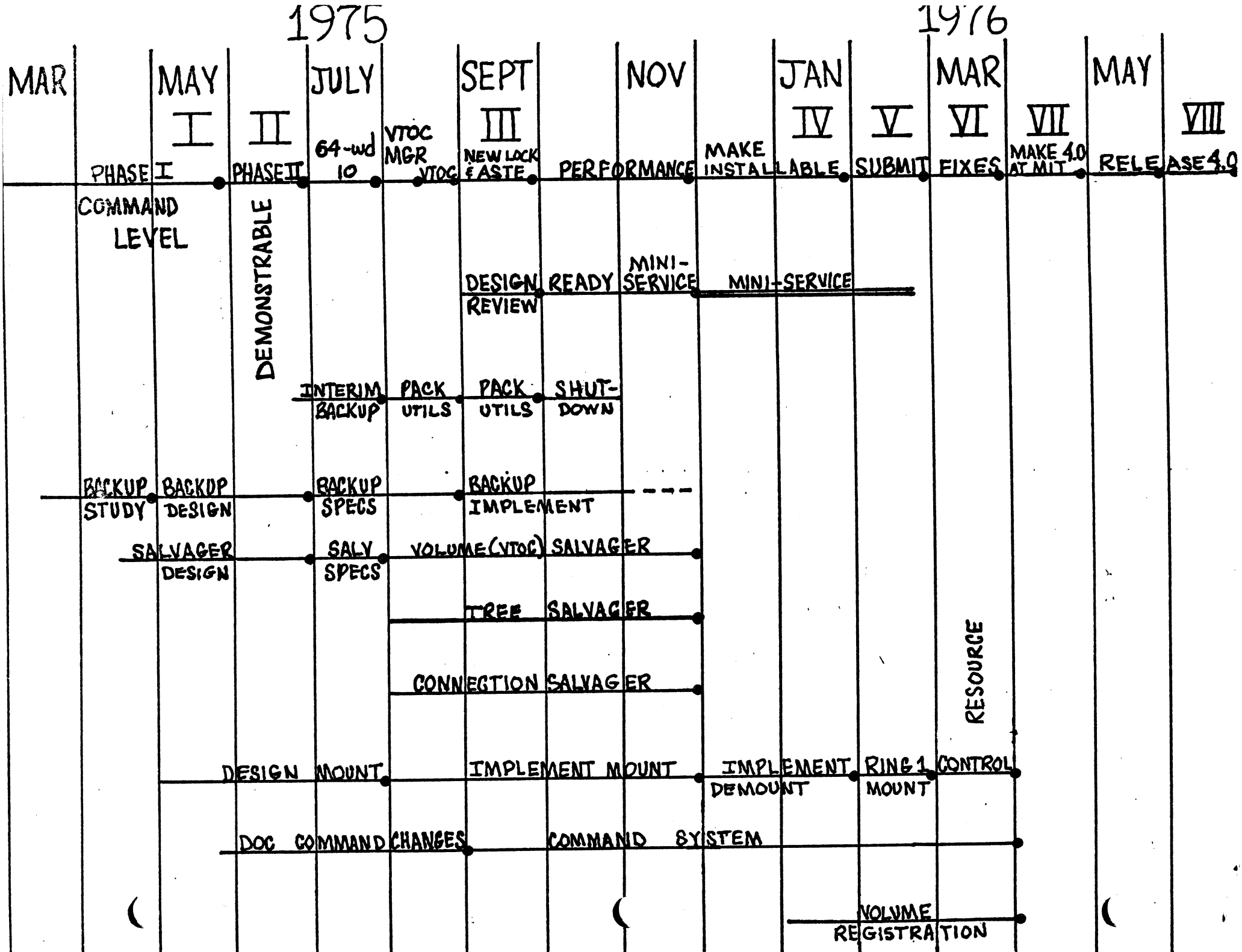
Once the duplicate copy facility is available, the system can be modified so that when a disk record is unreadable, the system automatically switches to the use of the secondary copy.

44. Disk DIM Error Handling Improvements.

Further improvements are possible to the disk error handling programs. It should be possible for the operator to move a pack which is encountering read errors from one drive to another, without crashing the system.

45. Billing.

Modifications must be made to the administrative and billing package to enhance the administrator's ability to manage the system resources. Some of these improvements cannot be specified until we have obtained some operational experience.


1975
1976
MAR
MAY
JULY
SEPT
NOV
JAN
MAR
MAY
I
II
III
IV
V
VI
VII
VIII
PHASE I
PHASE II
64-wd IO
VTOC MGR
VTOC
NEW LOCK & ASTE
PERFORMANCE
MAKE INSTALLABLE
SUBMIT
FIXES
MAKE 4.0 AT MIT
RELEASE 4.0
COMMAND LEVEL
DEMONSTRABLE
DESIGN REVIEW
READY
MINI-SERVICE
MINI-SERVICE
INTERIM BACKUP
PACK UTILS
PACK UTILS
SHUT-DOWN
BACKUP STUDY
BACKUP DESIGN
BACKUP SPECS
BACKUP IMPLEMENT
SALVAGER DESIGN
SALV SPECS
VOLUME (VTOC) SALVAGER
TREE SALVAGER
CONNECTION SALVAGER
RESOURCE
DESIGN MOUNT
IMPLEMENT MOUNT
IMPLEMENT DEMOUNT
RING 1 MOUNT
CONTROL
DOC COMMAND CHANGES
COMMAND SYSTEM
VOLUME REGISTRATION