Fujitsu

ETERNUS CD10000
Software Version V2.0
October 2015

Release Notice

All rights reserved, including intellectual property rights. Technical data subject to modifications and delivery subject to availability. Any liability that the data and illustrations are complete, actual or correct is excluded. Designations may be trademarks and/or copyrights of the respective manufacturer, the use of which by third parties for their own purposes may infringe the rights of such owner.

© 2015 Fujitsu Technology Solutions GmbH
1 General

This Release Notice is a summary of the major extensions, dependencies and operating information with respect to FUJITSU Storage ETERNUS CD10000 S2 hardware and the software V2.0. The release level is that of: October 2015.

A collection of hints and information which are necessary for professional and field service can be found in the appropriate and subsequent Support Bulletins, if applicable.

The names used in this Release Notice may be trademarks whose use by third parties for their own purposes may violate the rights of the owner.

1.1 Ordering

ETERNUS CD10000 can be ordered from your local distributors.

ETERNUS CD10000 is supplied subject to a single payment.

1.2 Delivery

The ETERNUS CD10000 files are supplied on DVD.

Delivery of files for ETERNUS CD10000 on DVD as part of hardware delivery:

- ETERNUS CD10000 2.0 Release : 2015-10-08

1.3 Documentation

The following manuals are available:

- ETERNUS CD10000 V2.0 User Guide

The documentation is available in the form of online manuals at http://manuals.ts.fujitsu.com.
2 Brief Product Description

The Fujitsu Storage ETERNUS CD10000 is a hyperscale storage system that addresses next generation storage requirements imposed by current developments in mobile broadband, social networks, big data, and cloud infrastructures. The appliance is implemented on the basis of a storage node cluster and built with standard IT components (network access, server, local hard drives). This grid of storage nodes presents itself as a single large storage repository. From a hardware perspective data access is granted via customer network facing 10GbEthernet interfaces, while the storage nodes among each other communicate via a private network which is also based on 10 Gb Ethernet.

From a software perspective storage access is provided via two major interfaces
- Via a Ceph block device interface
- Via a Ceph object storage interface (access by e.g. Ceph object library ("librados"), S3 or Open StackSwift)

For the sake of product quality the previously available (and still experimental) Ceph file system access (featuring one global namespace) is not part of the S2 generation.

All storage is managed as a single large entity, which greatly eases capacity and workload management.

A central technology element of the appliance is the intelligent data distribution across the storage nodes. In contrast to other systems the ETERNUS CD10000 manages read and write operations with practically no access to a central instance. A special data distribution algorithm called CRUSH (Controlled Replication under Scalable Hashing) enables all storage clients to exactly calculate where to write and read data within the storage cluster. This pseudo-random but still deterministic data distribution method enables an equal utilization of all involved storage nodes in the cluster. In addition all data can be written in a redundant way (via replicas or via a RAID-like algorithm called erasure coding). Using this method the system can tolerate the loss of hard drives as well as the loss of complete storage nodes without losing data. In addition, lost data copies are recreated in the background and the original redundancy level is restored automatically. Because data copies of a single hard disk drive are distributed across multiple nodes rather than copied to just another hard disk, more resources are involved in regenerating the old redundancy level and therefore the process of restoring the original redundancy level is very fast.

New nodes can be added easily while the system is running and in order to avoid hot spots data is reshuffled to create free space on all involved storage nodes, i.e. existing as well as new ones. As a result the cluster continues to benefit from the aggregated bandwidth of all storage nodes while writing new data to it. The 10GbEthernet backend offers the necessary bandwidth for
- creating redundant copies
- regeneration of the redundancy levels after a data loss
- reshuffling of data after new nodes are added.

As a consequence the 10GB Ethernet frontend network is fully available for reading and writing data.
On the basis of this technology the added value is:

- basically unlimited capacity. However configurations larger than 40 storage nodes are supported on so-called Special Release Request to ensure proper configuration.
- no performance hot spots
- parallel access over object, block and file interfaces by consistent storage management at the same time
- full redundancy for all data in the cluster
- automatic error detection and recovery of lost data at the same time
- possibility to transparently phase out old nodes and add new ones
- decouple management effort from storage growth

The ETERNUS CD10000 is a storage system that integrates hardware and software in one appliance including an end to end service concept. The hardware was chosen with care to reach a perfect balance between performance and price. The software allows managing all storage and networking components centrally. This includes monitoring, reporting, and the deployment / update of new nodes.
3 Software and Hardware Extensions

Only the new features and functional enhancements available with ETERNUS CD10000 S2 Hardware and the V2.0 software are described in the following section. The ETERNUS CD10000 User Guide provides detailed information about the features.

3.1 Software

3.1.1 General Enhancements

- RedHat Ceph Storage 1.3 based on the Ceph Hammer release
- Support of the new ETERNUS CD10000 S2 hardware
- The ETERNUS CD10000 V2.0 software supports the ETERNUS CD10000 S1 and S2 hardware.
- Support of the new optional Ceph Object Gateway and Monitor node types. The new types are installed, managed and monitored from the management node.
- Change of node’s hostnames convention (nodeX instead of storageX) to reflect new node types.
- Easy MAC address takeover in case of NIC and IB-board replacement
- Automatic firmware update for switches and RAID controller Additional to the existing function for the update of the iRMC and BIOS firmware, update of the RAID controller and cluster switches is supported.
- CD10000 CLI Enhancements
  - The values for DNS server, MTU size and Gateway can be specified when the public network is configured. The values can be modified at any time with the available command options.

3.1.2 vsm_cli Enhancements

The following new options are added to the vsm_cli command:

- set-ceph-parameters
  Ceph-specific parameters can be set or modified for an running Ceph cluster.
  The command updates the ceph.conf file, saves it in the VSM database, and restarts the cluster.

- Disk-location
  Shows the physical location (controller/slot) of an OSD or system disk. With the “-l on” option the appropriate LED start blinking. This can be switches off with the “-l off” option.

- maint-status
  reports whether a node is currently in maintenance mode

- add-storage-group
  Adding new drives grouped in a new storage group into an existing cluster.
  The command must be executed prior to adding a new node, or adding a JBOD to an existing node.

- snmptrap-off
  To avoid unnecessary SNMP traps while a node is in maintenance mode this suboption is available. If specified with option maint-on, the SNMP trap daemon is stopped prior to executing maint-on. SNMP traps are automatically enabled when maint-off is successfully executed.
3.1.3 Basking GUI

- Management Node Network IP
  Changing the IP address of the Management Node’s IP address in the customer network is now possible via the In the “Management network” section.

- Management node customer address redundancy
  For redundancy purposes it is now possible to bond the eth0 and eth1 ports. Select “Manage settings” and “Network” tab and choose “Management IP setup”.

- Show Actual OSD Location
  The new page “Disk/OSD Location” was added to “Information” section. See the description “vsm_cli disk-location”.

- Log Collections
  The collected log files can be downloaded to the local computer.

- Public network configuration
  The values for DNS server, MTU size and Gateway can be added or modified for the public network of the cluster.

- LDAP/AD Setup and modification
  To add user accounts for access to the management node on Linux level you can use LDAP or Active Directory (AD). Before adding user accounts you first have to configure how to access the LDAP or Active Directory server. You can do this with the section LDAP/AD configuration on the Cluster tab.

3.1.4 VSM (Virtual Storage Manager) GUI Enhancements

- New version V1.1
  The new version 1.1 is integrated and the version is displayed on the Cluster Status page.
  The version provides bug fixes mainly.

3.2 Hardware

The new ETERNUS CD10000 S2 hardware is focused on reducing the cost of managing exponential data growth, ensuring business continuity when processing mass data, and the provisioning of validated solutions for faster productive deployment of customer systems.

- Storage Nodes
  The Storage Nodes are based on the PRIMERGY RX2540 Rack Server, which can be enhanced with up to three ETERNUS JX40 S2 systems and opens up flexible options for scaling performance and capacity. Upgrades with capacity enhancements using ETERNUS JX40 S2 can be conducted on-site. Up to two PCIe SSDs can be ordered for High performance storage. They are not intended to be upgraded on-site.

- Network infrastructure
  An Ethernet backend network based on a 10Gb technology offers the bandwidth for backend operations (e.g., creating redundant copies) and a 10 Gb Ethernet frontend network for reading and writing data. In addition a 1 Gb Ethernet fabric is installed for storage cluster management.

- Management Node
  This node is responsible for the management of the appliance.

- Optional Monitor Node
  Per default the instances of the Ceph monitors run on the storage nodes. In heavily used environments it can make sense to run the monitor instances on dedicated nodes. So called monitor nodes can be configured for this purpose.

- Optional Ceph Object Gateway Node
  For configurations that require RESTful object access via S3 or OpenStack Swift, one or more gateway nodes are required to translate S3 or Swift object commands into the native Ceph object access. These optional nodes can be configured for this purpose on customer side.
4 Technical information

4.1 Resource requirements

ETERNUS CD10000 provides all required resources on its own hardware. Connection cables for the connection to the customer network have to be present or have to be ordered separately by the customer.

4.2 Software configuration

The ETERNUS CD10000 User Guide provides basic configuration instructions for the ETERNUS CD10000 system.

4.3 Product installation

4.3.1 Software Upgrade

All ETERNUS CD10000 running the software version V1.0SP02 can be upgraded to the software V2.0.

Proceed as follows to upgrade to the new software version:

- Insert the ETERNUS CD10000 V2.0 DVD into the slot of the management node. An automatically mount will take place to ABSOLUTE_PATH_TO_DVD.
- Upload the provided RPM packages to the repository:
  
  ```
  update-manager upload -d ABSOLUTE_PATH_TO_DVD --iso
  ```
- Publish repository
  
  ```
  update-manager publish
  ```
- Update the cluster manifest partials using Update Manager (manifest partials contain defined cluster state)
  
  ```
  update-manager updatepartials
  ```
- Unmount the DVD to avoid reinstallation from scratch after rebooting.
  
  ```
  umount ABSOLUTE_PATH_TO_DVD
  ```
- Run the migration script, to execute hostname migration. Wait until the management node prompt (pmgmt#) appears:
  
  ```
  python /opt/fujitsu/fpm/scripts/migrate_hostnames.py
  ```
- Merge the custom/changed manifests with the updated (this step can be omitted when there are no custom changes or no custom manifests).
- Upgrade the management node only using Update Manager. Executing this command, the management node will be rebooted. After rebooting wait about five minutes before going to the next step.
  
  ```
  update-manager upgrade -n pmgmt
  ```
- Upgrade the cluster using Update Manager (this function upgrades only storage nodes).
  
  ```
  update-manager upgrade
  ```

All nodes are upgraded one by one. If the upgrade of a node fails, the process exits with an error message. For a more detailed description of the upgrade process see ETERNUS CD10000 User Guide, chapter 9.
4.4 Product use

4.4.1 Rules Regarding the Number and Mixture of Storage Nodes
The minimum number of storage nodes is four.
Regarding the number and mixture of storage node types the following rules apply: It is possible to build the ETERNUS CD10000 appliance with different capacities of storage nodes. The minimum configuration consists of four storage nodes in one cluster. With less than 6 storage nodes in one cluster you have to use the same amount of capacity for each node to ensure proper work of the cluster.
In general smaller configurations below 8 storage nodes are not recommended and if less than 8 storage nodes are necessary the usage scenario and potential implications on operating the cluster should be considered very carefully. In any case please note that configurations with less than 4 storage nodes are not supported under any condition.

4.4.2 Backup of the Management Node
ETERNUS CD10000 is equipped with one management node. If this management node must be replaced, a backup of the system should be available.
For this purpose the Fujitsu Backup and Restoration Tool (fbart) is created. It provides backup and restoration of the management node of the Fujitsu ETERNUS CD10000. Multiple backups according to a predefined schedule are created automatically.
The backups include all the configuration data of the management node which are needed for a complete restoration.
During the “Getting started” process the backup has to be activated manually.

4.4.3 Replacing defect Hard Disks
If you have to substitute a new disk for a failed/retired disk keeping the overall storage cluster configuration intact this must be done with the vsm_cli replace disk commands (see User Guide for detailed information).
The web-based GUI VSM does not support this process!

4.4.4 Web Browser
VSM and Basking require Google Chrome, Firefox or Internet Explorer 10, or another modern Web-browser (tested with Google Chrome and Internet Explorer 10 with compatibility mode turned off).

4.5 Discontinued functions (and those to be discontinued)

4.5.1 Ceph MDS (Metadata Server Daemon) and Ceph File System
Ceph MDS and the Ceph file system are not supported by RedHat Ceph Storage 1.3.

4.6 Incompatibilities

None
4.7 Restrictions

4.7.1 Ceph Cluster creation
The ETERNUS CD10000 storage cluster provides the web-based user interface Virtual Storage Manager for Ceph (VSM) and the command line vsm_cli to manage and monitor the storage cluster. You can perform many configuration, administration, and monitoring functions with VSM. Ceph commands are installed on the system and can be used for some purposes.
It is highly recommended to create the Ceph cluster with VSM or vsm_cli and not with Ceph commands. If you are using Ceph to create the cluster it is not possible to manage or monitor the cluster with VSM afterwards.

4.7.2 Fbart ignores values in configuration file
If the default values for the nodes used as backup target are changed these values are ignored. The backup is created on node1 and node2.

4.7.3 Vsm_cli SSH keys exchange
When certain vsm_cli commands are performed, they might fail due to SSH keys not being properly exchanged between the nodes, this is a known bug. Use the following work around to fix the problem:

Execute the following command:

```
# /opt/fcephadm/ex-key.sh pmgmt node1 node2 ... nodeN
```

Execute this command in the following scenarios:
- For V1.0 S1/SP2 systems upgrading to V2.0: execute the command before the upgrade process.
- For fresh installed V2.0 systems: execute the command after you have configured the Management Node at customer site.
- If vsm_cli command does not work properly.

4.7.4 Other restrictions
- When the IP address of the Management node is changed on eth0, the page is not reloaded automatically. The user has to reload the page manually (the address in the address bar is changed successfully) and log in once again.
- Parallel execution of vsm_cli commands is not supported.
- Multi-location configurations are not supported.
- The optional Monitor Node will always be shown as possible node to be added in VSM with “Add Servers”, even if it is already part of the cluster.
- A broken storage or monitor node has to be removed from the cluster with the VSM GUI.
- The command vsm_cli disk-location and the corresponding Basking function do not support the Ceph Object Gateway nodes.
- If system disks are selected in the Basking GUI or with vms_cli disk-location and the “LED blink” option is specified, the location is displayed but the LED does not blink.
4.8 Procedure in the event of errors

If an error occurs, first consult the troubleshooting chapter in the ETERNUS CD10000 User Guide.

The following error documents are needed for diagnostics:

A detailed description of the error condition, indicating whether and how the error can be reproduced.

The ETERNUS CD10000 serial number can be obtained with the following command on the management node:

```
# cd10000 ip cluster show
```

```
+---------------------+---------------------+---------------------+---------------------+---------------------|
| Cluster SN          | Management node SN  | Administration net  |       |
| YM2D0010034         | YLNT04583           | 192.168.20.0/23     |       |
```

The entry in column “Cluster SN” is the relevant serial number.

The error must be reported to the appropriate service provider.

An AIS Connect access is essential for diagnostics.

5 Hardware requirements

None