

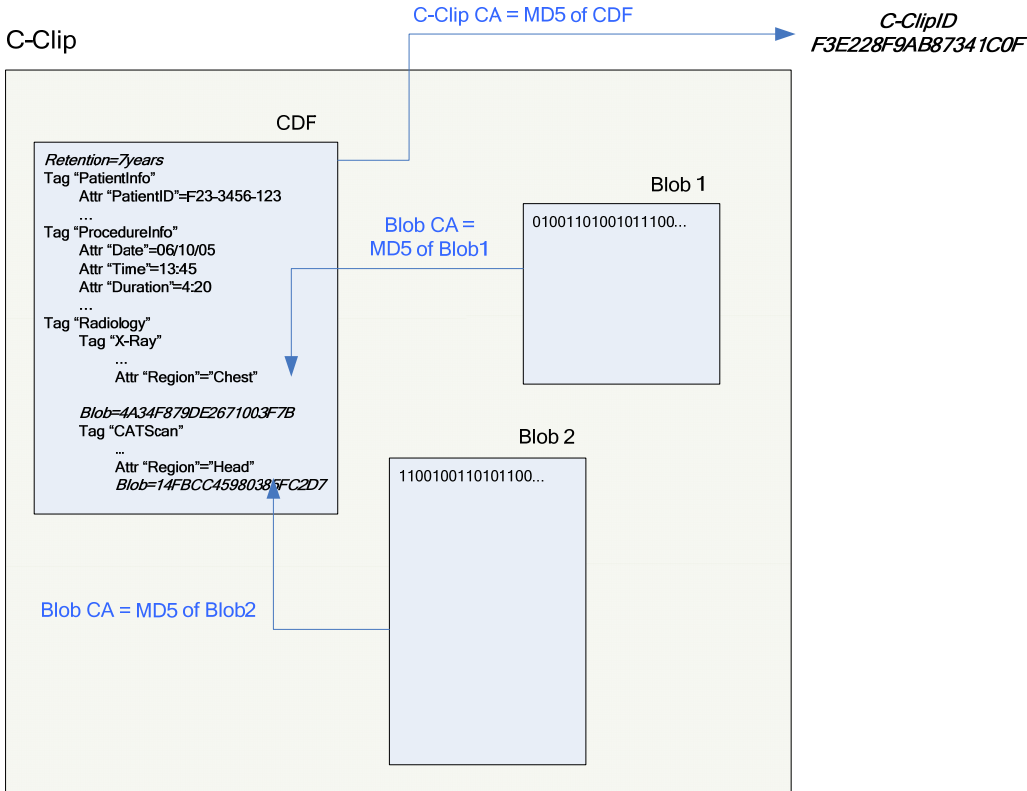
MANAGE DATA SMARTER™

StorFirst Altus Compliance
Technical Summary

Centera API Primer

StorFirst Altus is tightly integrated with EMC's Centera through the Centera API and it uses specific features within that API in order to provide application compliance. Since Centera compliance is a key element for the applications that write content to Centera, the API was designed early on to support third-party replication and recovery applications that adhere to, and bolster the compliance story.

In order to understand how Altus works, it is necessary to understand what a C-Clip is, and how it uses the MD5 hash algorithm to create Content Addresses (CA). The following diagram shows the relationships between the objects on the Centera that make up the C-Clip:



The CDF is stored as an XML file on the Centera, and the Blobs are stored as separate files. The C-Clip is therefore a logical object linked by the hierarchy of Content Addresses.

Typical Centera Applications

In practice, a typical Centera application uses a few Centera API calls. After initializing the API, it will normally only need to use the FPclip_Create API call to create a C-Clip. Once all metadata (Tags and Attributes, including Retention period) and data (Blobs) associated with the Clip have been written, the FPclip_Write API call is used to obtain the Content Address of the Clip, known as the C-Clip ID. Only after all input to the C-Clip has been recorded can the MD5 of the finalized CDF be calculated, and the C-Clip ID be communicated back to the application. The application will usually store the returned C-Clip ID in a database. In order to read data, the application will use the FPclip_Open call to open the C-Clip ID, and then FPTag_BlobRead calls to obtain data, and finally, FPclip_Close when no more data is needed.

StorFirst Altus Replication

Altus operates independently of the applications that are reading and writing data to the Centera. In order to obtain a list of C-Clips created since its last replication, it uses FPQuery_XXX calls which return the C-Clip IDs in chronological order of creation. Unlike a “typical” application, Altus uses the *FPClip_RawRead* API call for each C-Clip ID to read its CDF, which is then stored directly to tape/VTL exactly as received from the API, together with the C-Clip ID. Altus then identifies the blob identifiers specified within the CDF, and uses the *FPTag_BlobRead* function to transfer the raw data in each blob directly to tape/VTL. It is at this point where the components that make up the C-Clip are stored on tape/VTL in *exactly the same format* as they are on the Centera.

StorFirst Altus Restore

At restore time, Altus reads the C-Clip ID and the CDF from the tape/VTL. In order to start the re-creation of the C-Clip on the Centera, it uses the *FPClip_RawOpen* API call which specifies that the content for the C-Clip is already known. As a parameter to the call, Altus provides the CDF from the tape/VTL. Altus then identifies the blobs recorded in the CDF and writes the blob contents to the Centera using the *FPTag_BlobWrite* API calls, supplying the raw blob data from the tape/VTL. At this point, the Centera is overwriting the blob CAs in the CDF with new ones calculated from the data being written. If the tape/VTL has integrity, the CAs will be identical. Finally, and this is the key step, Altus attempts to finalize the C-Clip using the *FPClip_Write* API command. The call will fail if – after re-computing the CA of the CDF – it does not match the C-Clip ID passed to the *FPClip_RawOpen* command. Clearly, if the content of the blobs has not changed, and the metadata in the CDF has not changed (again, including retention), the CA for the CDF will exactly match the CA which had been computed when the C-Clip was first created by an application. Otherwise, it will not, and the entire C-Clip will be rejected by the Centera.

StorFirst Altus Media Management

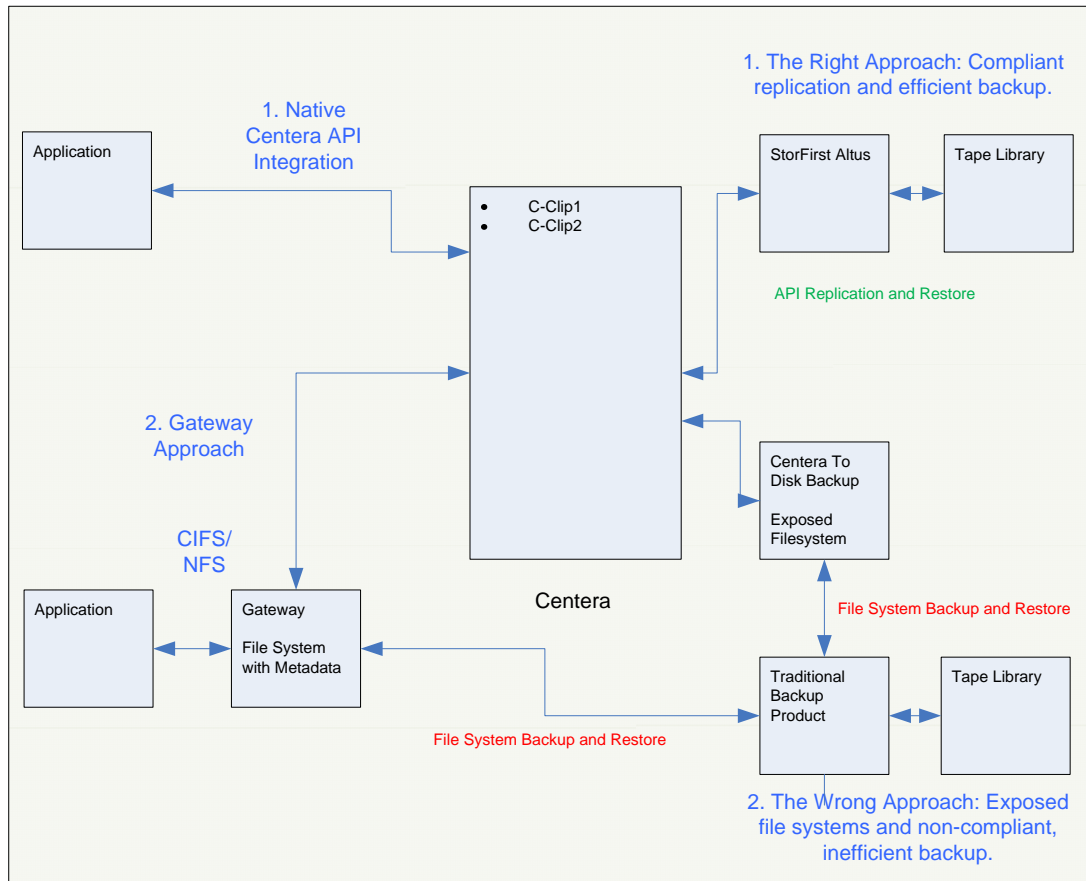
Compliance can only be asserted by a storage management product if it is in absolute control of the data from origin to destination. Altus fully integrates support for tape/VTL devices – both libraries and drives - and communicates with them at the SCSI level using its own device drivers. The tape/VTL format is proprietary to Altus and is designed specifically for the requirement of storing native Centera data. The Altus solution does not involve any third-party software at any stage; it does not hand-off data to any middleware product; and it tracks all near-line tapes (within the library) as well as all off-line tapes (outside the library) within its own media management database. Therefore, from the point at which a C-Clip is retrieved from Centera, to the point at which it is stored on tape/VTL at replication time (and recalled back from tape/VTL to the Centera during a restore), it remains completely under Altus control.

Competitive Solutions

There are a few other solutions for backing up Centera data to other media, but Seven10 contends that none are as compliant as StorFirst Altus. To understand where these applications fail, we must identify the two methods of archiving data to Centera:

1. **Gateway Access** – This puts an appliance between the applications creating content and the Centera itself. The appliance will export normal file-system-like share capabilities such as CIFS and NFS.
2. **Native Centera API Integration** – This means an application writes directly to Centera with no third party tools.

In either scenario, StorFirst Altus is fully compliant with the features and functionality of Centera. The following diagram shows two different approaches for archiving data *to* Centera and two different approaches for replicating data *from* Centera to other media such as tape.



On the right side of the diagram, we see the two different approaches for replicating and restoring Centera data:

1. **StorFirst Altus** – The software uses the RawRead and RawOpen commands provided by the Centera API to export C-Clips in 100% compliant format. All data and metadata are preserved through this approach. No matter which archiving approach is used, StorFirst Altus is fully compliant.
2. **Traditional Backup Product** – The traditional backup solution has two methods for replicating customer data, but both of them rely on a file system in between the Centera and the backup media.
 - a. The first is to backup the data directly from the gateway. This has a **major flaw** in terms of compliance. During restore, a new file, albeit with the same name and path, will be created in the gateway, not on Centera, and the gateway will then archive that file to Centera. By virtue of the calls that the gateway uses, this will result in a new C-Clip with a different ID than before. Because of this, there is no way to guarantee that the file's data has not been corrupted at some path between the backup product and the media.
 - b. The second method is to use a **Centera-to-Disk Backup** product as a middleware between the Centera and the backup product. The first part of this process uses similar calls to StorFirst Altus, but places C-Clips in an **emulated file system** so that the traditional backup product can then backup those files to tape. The problem with this approach is the emulated file system cannot be guaranteed to be tamper proof. There is an exposure window, where the emulated file system is not actually backed up to tape until the traditional backup product touches it.

In both cases, the traditional backup product must scan these file systems to find objects which have changed. This is less efficient than StorFirst Altus' "incremental forever" approach, which was adopted due to Seven10's intimate knowledge of the behavior of the Centera platform.

Summary

As demonstrated above, the immutability of data provided by the EMC Centera CAS storage paradigm is preserved when Altus is used as a replication and recovery vehicle for Centera data. It does this by adhering to two basic principles: (1) let the Centera client API and the Centera CentraStar operating system remain in control of data integrity and validation; and (2) ensure that there is an untainted data path between the Centera and the tape/VTL medium. As a result, Altus assures that the fundamental requirement of compliance is met.