



# Multi-Cloud Active Archive

---

January 2018



## Table of Contents

<b>Benefits Summary</b>	<b>3</b>
<b>  3-Tier Architecture and the Storage Model</b>	<b>3</b>
<b>3-Tier Architecture</b>	<b>3</b>
Agents	3
Appliances	3
Web Monitor	4
<b>Storage Model</b>	<b>4</b>
<b>  A walk-through time...</b>	<b>4</b>
Day 1 - The big bang	4
Day 2 - The next snapshot	5
Day 3 – And something changes...	5
Day 4 – A couple more files change...	6
Day 5 – No changes	6
<b>  I don't need it anymore</b>	<b>7</b>
<b>Active Data Pruning</b>	<b>7</b>
Day 6 – File 10 changes	7
Day 7 – Nothing changes	8
Day 8 – Are you on vacation?	9
Day 9 – Our final example	10
In closing...	11
<b>Active Archive Summary</b>	<b>11</b>

Aparavi helps organizations master out of control unstructured data growth. We slow secondary storage growth by 75% with guaranteed availability regardless of how long data is retained. Our SaaS-based active archive delivers true storage independence with on-premises and multi-cloud mobility. This along with our open data format removes vendor-lock in forever. Aparavi pays for itself in reduced secondary storage spend and eliminates up-front investments with a pay-as-you-go model based on usage.

This paper explores how the system works and how you can create the most effective implementation.

## Benefits Summary

- Only sends changes to the appliance
- Keeps data only for the period you need it
- Removes data after it has expired
- Works across local and cloud storage destinations

## I. 3-Tier Architecture and the Storage Model

The Aparavi active archive model uses a 3-tier architecture of agent, software appliance, and web monitor to manage your storage resources.

### 3-Tier Architecture

#### *Agents*

Agents are the systems with data on them to be retained: the data sources. For maximum data security, all data is encrypted and compressed before it leaves the agent.

#### *Software Appliances*

Appliances are usually local, on-site machines that store data coming from the agents and have the Aparavi appliance software loaded. These systems can store any number of copies of agent data, although usually only a few copies are kept. These copies are utilized to recover data quickly in case of agent failure or loss of data.

Software appliances are also responsible for sending data to the next tier, such as a centralized data center in case of a private cloud, or to a public cloud like Amazon Web Services (AWS) or Microsoft Azure.

Although appliances are usually on-site machines, alternately they can live in the cloud if preferred. The advantage of placing them on-site is that cloud transfers via the internet are typically much slower than a local area network so agents have to work longer to get data up into the cloud. This can have a definite impact on production systems.

### Web Monitor

The web monitor is the overall monitoring and management system. It is where you go to set systems up. It contains no data, is not involved in any data transfers, and is used simply as the centralized clearinghouse to manage systems.

### Storage Model

So how do all three of these systems work together to protect and retain your data? Let's walk through what happens in a typical scenario. We will assume for this example that you have 10 files (File 1 through File 10) on one agent that you need to archive, one software appliance installed on-site, a web monitor sitting in AWS, and an S3 storage account with AWS that will be used to keep all long-term data.

We will also assume that most of the default settings for the Aparavi policies are kept, and that you have already installed and activated all three component tiers. There are many settings that you may customize, but the default policies shipped with the web monitor implement a pretty solid scheme and will work for most systems without modification.



## II. A walk-through time...

### Day 1 - The big bang

The first step is for the agent to connect to your software appliance, which connects to the web server to grab the current archive policy. This happens the very first time you start the agent. Any updates to the policy will automatically be implemented when they are changed via the web monitor.

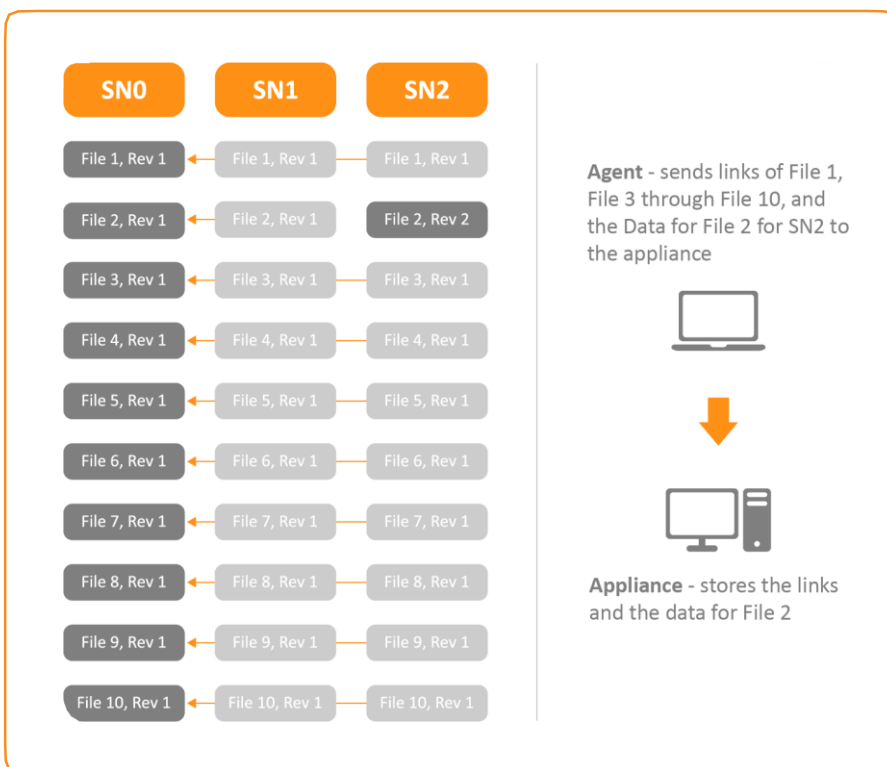
Once a valid policy is received, the agent starts processing the policy and will create its first snapshot which we will call SNO.

With this action, the appliance now has copies of File 1 through File 10. The files are all marked with revision number 1 (Rev 1) since it is the first revision of the file that is being protected.



### Day 2 - The next snapshot

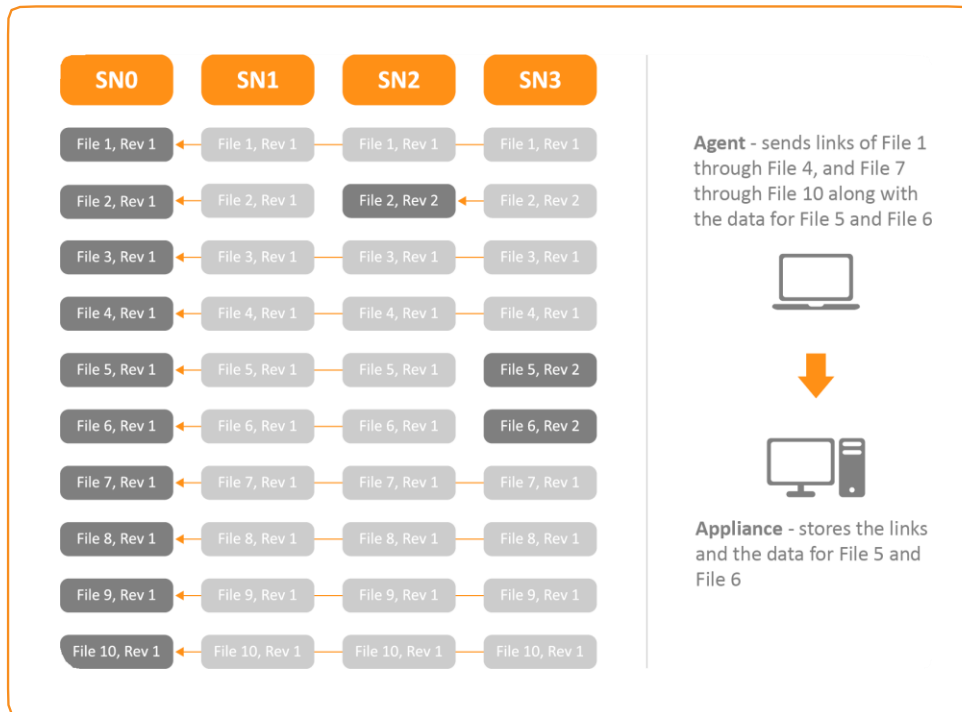
Let's assume none of the files change between the SN0 and the next snapshot SN1. This is not very likely, but let's explore what would happen. Instead of sending the files again, which would be completely redundant, Aparavi just sends links to the files contained in SN0.



### Day 3 – And something changes...

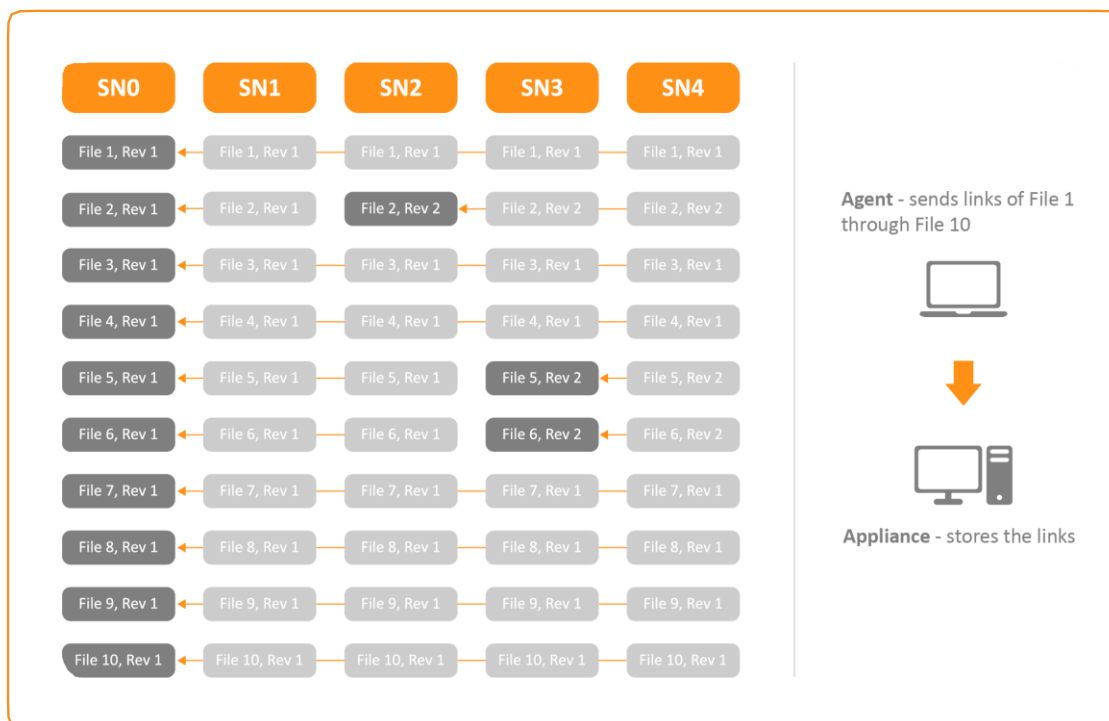
Now, let's say File 2 changes. What happens? Again, we don't need to resend the complete set of data from File 1 through File 10 as the software appliance already has that. The only thing we need to send are the links, of File 1 and File 3 through File 10, and the data for File 2. What does that look like?

Let's point out here that the appliance is now tracking 3 snapshots (SN0, SN1, SN2) and tracking exactly 11 files (File 1, Rev 1, File 2 Rev 1, File 2 Rev 2, File 3 Rev 1, etc.).



### Day 4 – A couple more files change...

Let's change File 5 and File 6. Pretty simple as it just follows the same pattern as from the day before. The only data that needs to be sent is the data for File 5 and File 6.



### Day 5 – No changes

The only thing that needs to be sent to the software appliance from the agent are links to existing data.

### III. I don't need it anymore

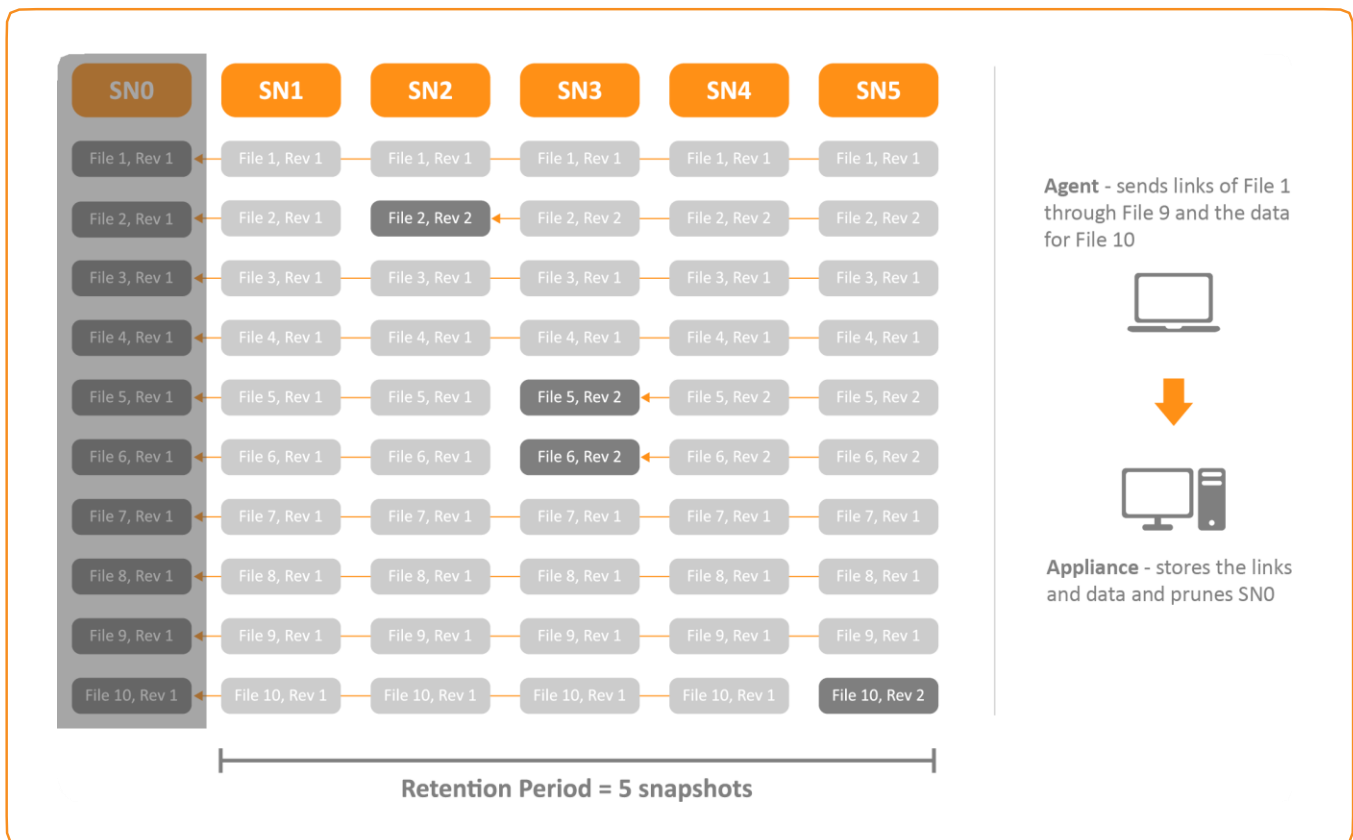
We now have 5 days of data on the software appliance, and if we keep going at this rate, our appliance is soon going to run out of storage. This growth is not sustainable over the long term. What Aparavi does to handle this is called Active Data Pruning.

#### Active Data Pruning

Active Data Pruning works as follows:

##### Day 6 – File 10 changes

Just like yesterday, the only thing sent from the agent to the software appliance is the links of File 1 through File 10 to the snapshots where the data actually resides, and the data from File 10.



However, the appliance performs an additional step. Since the agent's policy requires only five snapshots to be kept, and we now have six, we can now get rid of the first snapshot we made, or SN0. SN0 is not presented as a recoverable snapshot on the recover page of the web monitor, only the five that are recoverable. That is, SN1 through SN5 are shown.

However, even though it is not listed, some of the data contained in SN0 still needs to be kept. In the case of SN0, pretty much all the data needs to be kept, because one of our still recoverable snapshots, in particular SN1, still links to data in SN0. Therefore, nothing happens - yet.

### Day 7 – Nothing changes

Is it a holiday or something? With nothing changing, the agent only sends links again.

The Active Data Pruning process now determines that there is nothing in SN1 that is needed, so it completely removes SN1. However, all 10 files are still required from SN0, so it leaves them as they are. Even though SN0 does not show up on the recovery page, all its data is still there.

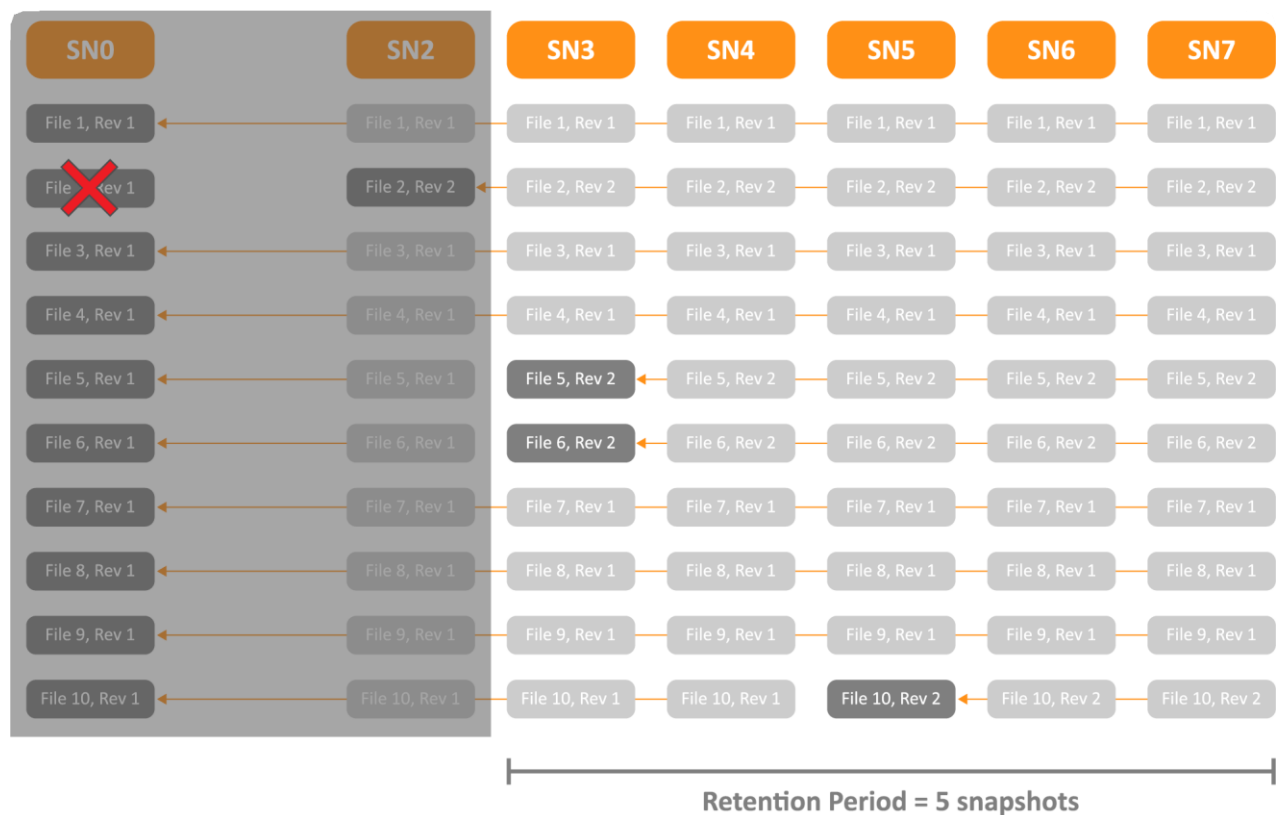




### Day 8 – Are you on vacation?


No changes for day 8, so, yet again, only links are sent from the agent to the software appliance. However, something different happens with the Active Data Pruning process.

Since File 2, Rev 1 is no longer referenced in any of our retained snapshots (SN3 through SN7), we can remove the data on the appliance occupied by File 2, Rev 1. Thus, the only recoverable revision of File 2 is now Rev 2.



  
**Agent** - sends links of  
 File 1 through File 10

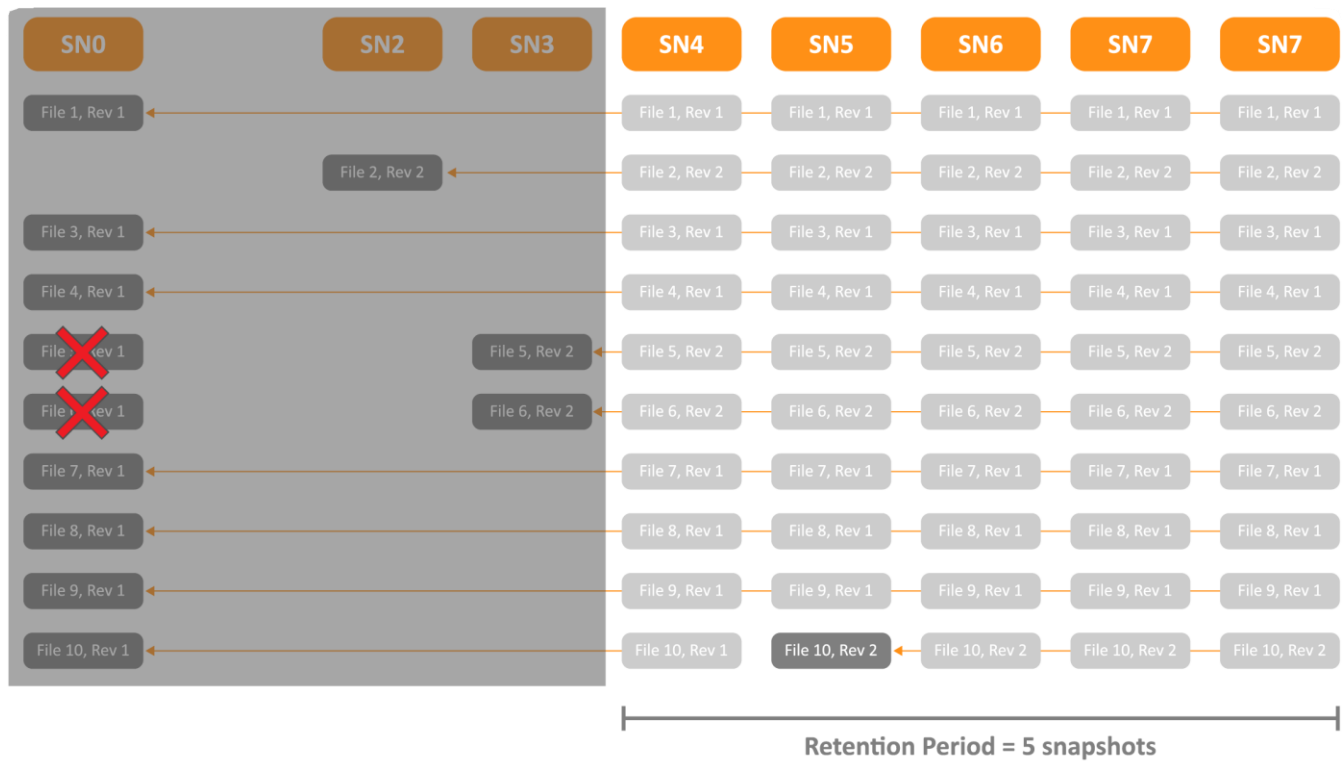


  
**Appliance** - stores the  
 links and prunes SN0, SN1

### Day 9 – Our final example

Once again, there are no changes so links are sent from the agent to the appliance for File 1 through File 10.

Active Data Pruning recognizes that we can now remove File 5 Rev 1 and File 6 Rev 1 as they are no longer recoverable from any of the retained snapshots.



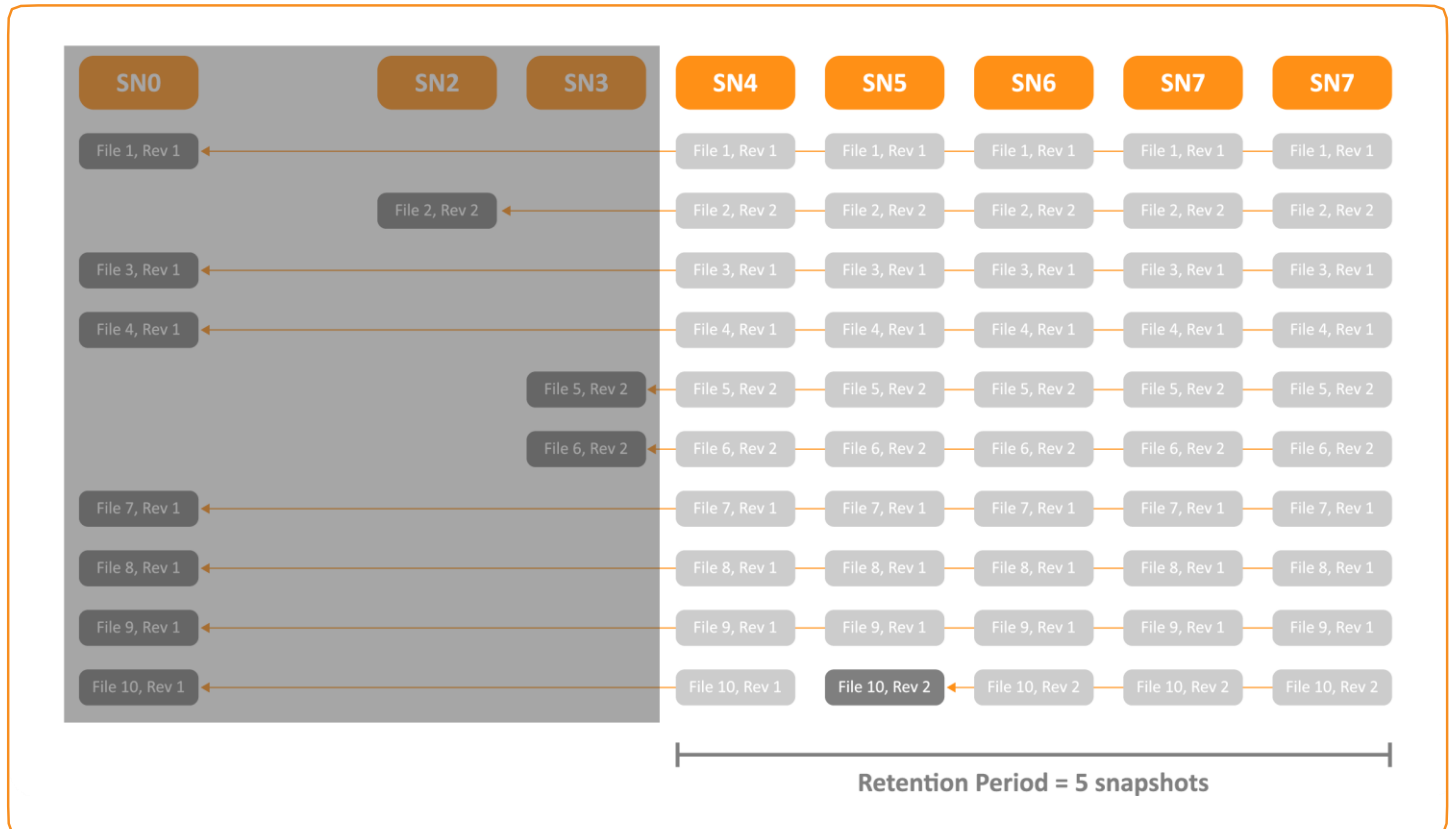
  
**Agent** - sends links of  
 File 1 through File 10



  
**Appliance** - stores the links  
 and prunes SN0, SN2, SN3

*In closing...*

What does the storage look like on the software appliance after these nine snapshots are stored?



From this, you can see Aparavi only keeps data as long as it needs to keep it to maintain your data retention policy. Instead of just continually growing your storage over time, this prunes data as it 'falls off the retention cycle'.

## Active Archive Summary

Active Data Pruning is a critical component of enabling active archive management over time for more efficient, cost-beneficial management of storage. It slows secondary storage growth, while retaining data per retention policies. In summary, Active Data Pruning helps organizations master out-of-control unstructured data growth for greater efficiency, better management, and lower costs.