Lab 10.: Introduction to Backup and Recovery
IN719 Systems Administration

Introduction

Even though this is the last major topic area we're covering in this paper, backup and recovery is probably the most important area of responsibility for systems administrators. Even though we are inclined to think of our systems as the critical resource, in the organisations where we work it is the *data* that is really valuable. We can get new servers, but if we lose our data there is really no point. On the other hand, if you have all of your important data carefully backed up, then you always have the potential to recover from a variety of problems. If a user mistakenly deletes important data, backups will let you get it back. If your organisation is hit with a ransomware attack, you can shrug it off and keep going as long as you have backups. If velociraptors trash the server room, you can set up in a new one as long as you have your data.

Notice that our topic is backup *and recovery*. If we don't have procedures to recover the data from our backup media and put that data back into use, then there was no point in backing it up. What's more, we need to have confidence that our restore procedures will work, which generally means that we need to test them regularly by running recovery drills. Often when we need to execute a recovery we're under a large amount of pressure because important services are degraded or unavailable. We don't want to "test" our recovery process under these circumstances.

1 Reasons to back up

Before we can plan and execute backup procedures, we need to identify the reasons why we back up our data. Generally these fall into three main categories.

1. Protection from accidental loss: Someone may delete a file or its contents might become corrupted. In this case we want a system that will let us find and restore specific files. We will need to retain many copies of the data from various points in time, although primarily just for the short to medium term.

2. Archival purposes: We might have data that we are required by law or policy to retain for long periods of time. A backup system for this use should store its data on stable removable media. It also needs to produce some sort of catalogue so that we can find data that may have been stored untouched for a long time. On the other hand, the recovery process probably does not need to be quick.

3. Service availability: If a critical system fails for some reason, there will be data that is needed to get the service functioning again. In this case we would like the recovery process to be as quick as possible, but we're probably only interested in very recent versions of the data. Another important consideration is that the data involved in this case includes a lot of system configuration data, where the other two cases involve user-produced data for the most part.

Our reasons for backup up are important when planning our strategies. For example, with many VM platforms it is possible to take "snapshots" of virtual machines. This works well for purpose number 3 above but very poorly for the other two. On the other hand, systems that back up data to tape are very good for purpose number 2, pretty good for number 1, but not ideal for number 3.

While all three reasons are important, in this paper we are going to focus on the third - backing up for service availability. This is because it is the reason most relevant to our working scenario.

2 Planning

Today we will work on planning our backup and recovery procedures. A good way to approach this is with *threat modelling*. What are the threats that we are concerned about? Basically, they are scenarios where one or more servers stops functioning. This could be because of a systems fault, a security breach, or even a gross error on the part of a team member (`sudo rm -rf /`). In these situations, the most likely recovery is going to involve launching a new virtual machine and restoring data to it required to make it function as desired.

Start your planning process by going over each server in your set and determining what data is on that server, above and beyond the base installation of the operating system, that is required for it to function? How is that data stored? Is it in a plain file, a database, or something else? Note that, since you installed OwnCloud earlier this week, you have some new things to consider.

Next, for each server, start outlining a procedure for how you could restore the data, probably onto a new VM. Since the new VM may not have the same IP address as the one it replaces, what needs to be done to handle that?

You might not have all the answers to the questions raised by this analysis. That's ok. Flag those questions and be sure you get them sorted in the next week of so.