



Journaling file systems

Alp E. ÖZKUL



Before “what is journaling?”

- Purpose of filesystems to exist is storing data.
- A filesystem uses metadata to maintain the data.
- Metadata is the data about the data.



Problems

- A filesystem expects to find metadata in a non-corrupted state. Otherwise filesystem won't be able to access files.
- When the system crashes or the file system can't be unmounted correctly, the metadata might be damaged.



Fsck (file system check)

- To solve the problem we discussed (the non-corrupted state of metadata), fsck exists.
- When a linux system boots, fsck scans the filesystems for errors and corrects them so that they can mount correctly.



Why fsck is bad?

- Using fsck seems good to make filesystem consistent.
- What if we have very very big filesystem?
- We have to wait hours and hours for system to boot up. Because fsck scans entire filesystem.



What is journaling?

- A filesystem is made journaling filesystem by adding a kind of data structure called journal.
- It is now a journaling file system.
- Now that it is, we don't need to use fsck anymore.



How it works?

- Before any changes made to the metadata, the filesystem writes data to the journal about what it will do.
- Then, writes metadata.
- By doing so, it saves all changes in a log of recent metadata modifications.
- It is a chronological order log.



Why don't we need to use fsck?

- Fsck scans entire metadata to find errors.
- We don't need it because the journaled filesystem understands where the problem is just in seconds by looking the log which has recent changes.
- It is very good to use a journaling file system for its speed of recovery.

Some journaling filesystems

- Ext3, Reiserfs, JFS, XFS.
- Ext3 is a journaling file system. It is developed by adding a journal to ext2 file system. Without journal, ext3 is same as ext2.
- Reiserfs is excellent for small files.
- JFS is developed by IBM.
- XFS is developed by Silicon Graphics.



Which to choose?

- Like everything, choosing the right filesystem to use depends on what to do.
- If small files are rare, reiserfs should be chosen for its performance.
- If we have existing ext2 filesystem, we can make it ext3.



Journaling

- Thanks to journal, the filesystem comes in a consistent state in a few seconds.
- It is very good not to wait for system to boot up on power failures.
- I have been using journaling file systems (ext3 and reiserfs) for years.



References

- http://en.wikipedia.org/wiki/Journaling_file_system
- <http://www-128.ibm.com/developerworks/library/l-fs.html>