

## JobSub - Bug #20659

### Extracting bziped tarballs is very slow

08/21/2018 05:38 PM - Dave Dykstra

<b>Status:</b>	Closed	<b>Start date:</b>	08/21/2018
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Dennis Box	<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	v1.2.9	<b>Spent time:</b>	0.00 hour
<b>First Occurred:</b>		<b>Stakeholders:</b>	
<b>Occurs In:</b>			

#### Description

Jobsub's tar file feature now uses bzip2 encoding. I noticed that extracting the tar files was taking a lot of cpu time on the test cvmfs publishing, and heard that uncompressing bzip2 was about 5 times slower than gzip. Dennis helped me to do a test by creating an approximately 1GB gzipped tar file with exactly the same contents as a formerly used bzip2 tar file, and found that it was 8.5 times faster: it took only 16 seconds to extract instead of 137 seconds. So I suggest switching to use gzip.

Dennis had originally used bzip2 because with gzip the hash changes each time, but the gzip -n option avoids that.

The cvmfs publishing itself is not a very high priority since that is just a test at this point, but this should also save significant time in decompression of these tar files on worker nodes.

#### History

##### #1 - 11/15/2018 03:05 PM - Dennis Box

- Assignee set to Dennis Box
- Target version set to v1.2.9

##### #2 - 11/21/2018 02:44 PM - Dennis Box

- Status changed from New to Feedback
- Assignee changed from Dennis Box to Shreyas Bhat

##### #3 - 11/29/2018 02:27 PM - Shreyas Bhat

- Status changed from Feedback to Accepted

Looks good.

##### #4 - 11/29/2018 02:28 PM - Shreyas Bhat

- Assignee changed from Shreyas Bhat to Dennis Box

##### #5 - 11/30/2018 01:28 PM - Dennis Box

- Status changed from Accepted to Resolved

merged to branch branch\_1.2.9\_rc0

##### #6 - 02/28/2019 10:30 AM - Dennis Box

- Status changed from Resolved to Closed