

Arvados - Bug #4309

[SDK] arv-copy collection copy performance

10/24/2014 05:21 PM - Peter Amstutz

Status:	New	Start date:	
Priority:	Normal	Due date:	
Assigned To:		% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	Arvados Future Sprints		

Description

Originally <https://arvados.org/issues/3699#note-43>

I'm copying a pipeline with a 5990M collection. I noticed this code:

```
data = src_keep.get(word)
dst_locator = dst_keep.put(data)
```

See attached image, there's a very clear falloff between blocks -- doing this sequentially isn't optimal. Download and upload could proceed concurrently. Also, it's possible we could get better utilization if we transferred multiple blocks at a time (e.g. 2x down / 2x up) by talking to multiple Keep servers. Consider a producer-consumer pattern using [Python queues](#).

History

#1 - 10/24/2014 05:22 PM - Peter Amstutz

- File arv-copy-perf.png added
- Description updated

#2 - 10/24/2014 05:23 PM - Peter Amstutz

- Description updated

#3 - 10/24/2014 05:26 PM - Peter Amstutz

- Description updated

#4 - 10/24/2014 05:28 PM - Peter Amstutz

- Subject changed from [SDK] arv-copy performance to [SDK] arv-copy collection copy performance
- Description updated

#5 - 10/27/2014 01:40 PM - Radhika Chippada

- Target version set to Bug Triage

#6 - 10/29/2014 06:14 PM - Tom Clegg

- Target version changed from Bug Triage to Arvados Future Sprints

#7 - 10/29/2014 06:51 PM - Ward Vandewege

- Story points set to 2.0

Files

arv-copy-perf.png	25.8 KB	10/24/2014	Peter Amstutz
-------------------	---------	------------	---------------