

Arvados - Story #10387

Faster downloading using arv-get

10/27/2016 05:43 PM - Joshua Randall

Status:	New	Start date:	10/27/2016
Priority:	Normal	Due date:	
Assigned To:		% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	Arvados Future Sprints		
Description			
<p>As a user wanting to access data kept in keep from an external (non-Arvidos) system, I'd like my `arv keep get` requests to be able to fully utilise available network bandwidth for a download from keep. The limiting factor currently seems to be that blocks are downloaded serially starting at the beginning of a manifest and working forward to the end). My expectation would be that destination files could be preallocated and that blocks could then be downloaded concurrently to fill in the files as the blocks arrive. The rate of parallel downloading could be increased until overall throughput stops increasing (indicating the network may be saturated) at which point the client could back off a bit. The maximum number of concurrent connections to each keep server could be limited by configuration.</p> <p>A request for parallel downloading may be incompatible with options such as `--md5sum` to display the hash of files as they come in (because the hash needs to be computed serially), but as a user I wouldn't mind making the options mutually exclusive and being able to choose either `--parallel` (or whatever) or `--md5sum`.</p>			
Related issues:			
Related to Arvados - Story #7824: [SDKs] arv-get and arv-ls should use new Py...		Resolved	11/19/2015

History

#1 - 08/29/2017 02:04 PM - Tom Morris

- Target version set to Arvados Future Sprints