

Corpus	chunker	size	compared to plain files	compared to 256K	compared to zstd -1 --long	command	Description		filenames	
Archlinux iso 3x	256K	2'355'828 KB	99.64 %	100.00 %	101.51 %	add --raw-leaves	Three uncompressed iso images for archlinux	cat *.iso > test.bin; zstd -1 --long test.bin	plain files	archlinux-2021.05.01-x86_64.iso archlinux-2021.06.01-x86_64.iso archlinux-2021.07.01-x86_64.iso
	128K	2'355'572 KB	99.63 %	99.99 %	101.50 %	add --raw-leaves --chunker size-131072				
	8K	2'356'384 KB	99.66 %	100.02 %	101.54 %	add --raw-leaves --chunker size-8192				
	buzhash	1'545'208 KB	65.35 %	65.59 %	66.58 %	add --raw-leaves --chunker buzhash				
	rabin	1'554'984 KB	65.77 %	66.01 %	67.00 %	add --raw-leaves --chunker rabin				
	rabin-4k-8k-16k	1'632'936 KB	69.06 %	69.31 %	70.36 %	add --raw-leaves --chunker rabin-4096-8192-16384				
	rabin-16k-32k-64k	1'502'596 KB	63.55 %	63.78 %	64.75 %	add --raw-leaves --chunker rabin-16384-32768-65536				
	rabin-32k-64k-128k	1'507'472 KB	63.76 %	63.99 %	64.96 %	add --raw-leaves --chunker rabin-32768-65536-131072				
	rabin-8k-16k-64k	1'543'280 KB	65.27 %	65.51 %	66.50 %	add --raw-leaves --chunker rabin-8192-16384-65536				
	rabin-12k-24k-48k	1'528'428 KB	64.64 %	64.88 %	65.86 %	add --raw-leaves --chunker rabin-12288-24576-49152				
Linux kernel pkgs 4x	256K	428'736 KB	100.03 %	100.00 %	141.52 %	add --raw-leaves	Four archlinux packages (uncompressed) for the linux kernel	cat *.tar > test.bin; zstd -1 --long test.bin	plain files	linux-5.12.6.arch11-x86_64.pkg.tar linux-5.12.7.arch11-x86_64.pkg.tar linux-5.12.8.arch11-x86_64.pkg.tar linux-5.12.9.arch11-x86_64.pkg.tar
	128K	428'592 KB	100.00 %	99.97 %	141.47 %	add --raw-leaves --chunker size-131072				
	8K	428'352 KB	99.94 %	99.91 %	141.39 %	add --raw-leaves --chunker size-8192				
	buzhash	426'384 KB	99.48 %	99.45 %	140.74 %	add --raw-leaves --chunker buzhash				
	rabin	428'508 KB	99.98 %	99.95 %	141.44 %	add --raw-leaves --chunker rabin				
	rabin-4k-8k-16k	480'604 KB	112.14 %	112.10 %	158.64 %	add --raw-leaves --chunker rabin-4096-8192-16384				
	rabin-16k-32k-64k	439'256 KB	102.49 %	102.45 %	144.99 %	add --raw-leaves --chunker rabin-16384-32768-65536				
	rabin-32k-64k-128k	432'276 KB	100.86 %	100.83 %	142.69 %	add --raw-leaves --chunker rabin-32768-65536-131072				
	rabin-8k-16k-64k	454'320 KB	106.00 %	105.97 %	149.96 %	add --raw-leaves --chunker rabin-8192-16384-65536				
	rabin-12k-24k-48k	445'508 KB	103.95 %	103.91 %	147.06 %	add --raw-leaves --chunker rabin-12288-24576-49152				
snapshots of Home Assistant 4x	256K	9'036'320 KB	43.56 %	100.00 %	238.65 %	add --raw-leaves -r	Private Home Assistant installation backups (daily) which contain python files, sqlite databases, a large mariadb database with sensor data and several different log files	tar cf test.tar *; zstd -1 --long test.tar	plain files	snapshot_2021-07-08/... snapshot_2021-07-10/... snapshot_2021-07-11/... snapshot_2021-07-12/...
	128K	8'736'136 KB	42.11 %	96.68 %	230.73 %	add --raw-leaves --chunker size-131072 -r				
	16K	8'132'660 KB	39.20 %	90.00 %	214.79 %	add --raw-leaves --chunker size-16384 -r				
	8K	8'177'688 KB	39.42 %	90.50 %	215.98 %	add --raw-leaves --chunker size-8192 -r				
	buzhash	13'014'612 KB	62.73 %	144.03 %	343.72 %	add --raw-leaves --chunker buzhash -r				
	rabin	12'091'432 KB	58.28 %	133.81 %	319.34 %	add --raw-leaves --chunker rabin -r				
	rabin-4k-8k-16k	9'412'424 KB	45.37 %	104.16 %	248.59 %	add --raw-leaves --chunker rabin-4096-8192-16384 -r				
	rabin-16k-32k-64k	9'631'948 KB	46.43 %	106.59 %	254.38 %	add --raw-leaves --chunker rabin-16384-32768-65536 -r				
	rabin-32k-64k-128k	10'759'392 KB	51.86 %	119.07 %	284.16 %	add --raw-leaves --chunker rabin-32768-65536-131072 -r				
	rabin-8k-16k-64k	9'349'636 KB	45.07 %	103.47 %	246.93 %	add --raw-leaves --chunker rabin-8192-16384-65536 -r				
rabin-12k-24k-48k	9'275'820 KB	44.71 %	102.65 %	244.98 %	add --raw-leaves --chunker rabin-12288-24576-49152 -r					
Oad-data	256K	4'561'400 KB	68.66 %	100.00 %	157.56 %	add --raw-leaves -r	Three archlinux packages which contain media data for the game Oad	cat *.tar > test.bin; zstd -1 --long test.bin	plain files	Oad-data-a23.1-1-any.pkg.tar Oad-data-a23.1-2-any.pkg.tar Oad-data-a24.b-1-any.pkg.tar
	128K	4'552'316 KB	68.53 %	99.80 %	157.25 %	add --raw-leaves --chunker size-131072 -r				
	8K	4'509'152 KB	67.88 %	98.85 %	155.76 %	add --raw-leaves --chunker size-8192 -r				
	buzhash	4'288'872 KB	64.56 %	94.03 %	148.15 %	add --raw-leaves --chunker buzhash -r				
	rabin	4'317'376 KB	64.99 %	94.65 %	149.13 %	add --raw-leaves --chunker rabin -r				
	rabin-4k-8k-16k	4'867'096 KB	73.26 %	106.70 %	168.12 %	add --raw-leaves --chunker rabin-4096-8192-16384 -r				
	rabin-16k-32k-64k	4'315'408 KB	64.96 %	94.61 %	149.07 %	add --raw-leaves --chunker rabin-16384-32768-65536 -r				
	rabin-32k-64k-128k	4'294'972 KB	64.65 %	94.16 %	148.36 %	add --raw-leaves --chunker rabin-32768-65536-131072 -r				
	rabin-8k-16k-64k	4'525'780 KB	68.13 %	99.22 %	156.33 %	add --raw-leaves --chunker rabin-8192-16384-65536 -r				
	rabin-12k-24k-48k	4'416'356 KB	66.48 %	96.82 %	152.55 %	add --raw-leaves --chunker rabin-12288-24576-49152 -r				
debian vmware images	256K	6'002'048 KB	100.03 %	100.00 %	106.49 %	add --raw-leaves -r	Three versions of debian as vmware images	tar cf test.tar *; zstd -1 --long test.tar	plain files	Debian_10.1.0_VMware_Image_LinuxVMIImages-disk1.vmdk Debian_10.1.0_VMware_Image_LinuxVMIImages.mf Debian_10.1.0_VMware_Image_LinuxVMIImages.ovf Debian_10.6.0_VM_LinuxVMIImages-disk1.vmdk Debian_10.6.0_VM_LinuxVMIImages.mf Debian_10.6.0_VM_LinuxVMIImages.ovf Debian_10.7.0_VM_LinuxVMIImages-disk1.vmdk Debian_10.7.0_VM_LinuxVMIImages.mf Debian_10.7.0_VM_LinuxVMIImages.ovf
	128K	6'003'676 KB	100.05 %	100.03 %	106.52 %	add --raw-leaves --chunker size-131072 -r				
	8K	5'888'660 KB	98.14 %	98.11 %	104.48 %	add --raw-leaves --chunker size-8192 -r				
	buzhash	6'095'684 KB	101.59 %	101.56 %	108.15 %	add --raw-leaves --chunker buzhash -r				
	rabin	6'124'884 KB	102.07 %	102.05 %	108.67 %	add --raw-leaves --chunker rabin -r				
	rabin-4k-8k-16k	5'529'188 KB	92.15 %	92.12 %	98.10 %	add --raw-leaves --chunker rabin-4096-8192-16384 -r				
	rabin-16k-32k-64k	6'080'264 KB	101.33 %	101.30 %	107.88 %	add --raw-leaves --chunker rabin-16384-32768-65536 -r				
	rabin-32k-64k-128k	6'229'632 KB	103.82 %	103.79 %	110.53 %	add --raw-leaves --chunker rabin-32768-65536-131072 -r				
	rabin-8k-16k-64k	5'789'288 KB	96.48 %	96.46 %	102.72 %	add --raw-leaves --chunker rabin-8192-16384-65536 -r				
	rabin-12k-24k-48k	5'859'608 KB	97.65 %	97.63 %	103.96 %	add --raw-leaves --chunker rabin-12288-24576-49152 -r				
Linux kernel git	256K	3'631'120 KB	100.03 %	100.00 %	107.93 %	add --raw-leaves -r	Linux git as bare repo, according to ipns://docs.ipfs.io/how-to/host-git-style-repo/commit 40226a3d96ef8ab8980f032681c8bf4d6d63874e	tar cf test.tar *; zstd -1 --long test.tar	plain files	linux.git/...
	128K	3'632'100 KB	100.05 %	100.03 %	107.96 %	add --raw-leaves --chunker size-131072 -r				
	8K	3'661'624 KB	100.87 %	100.84 %	108.84 %	add --raw-leaves --chunker size-8192 -r				
	buzhash	3'658'756 KB	100.79 %	100.76 %	108.75 %	add --raw-leaves --chunker buzhash -r				
	rabin	3'649'816 KB	100.54 %	100.51 %	108.49 %	add --raw-leaves --chunker rabin -r				
	rabin-4k-8k-16k	4'235'160 KB	116.67 %	116.64 %	125.89 %	add --raw-leaves --chunker rabin-4096-8192-16384 -r				
	rabin-16k-32k-64k	3'773'340 KB	103.95 %	103.92 %	112.16 %	add --raw-leaves --chunker rabin-16384-32768-65536 -r				
	rabin-32k-64k-128k	3'701'816 KB	101.98 %	101.95 %	110.03 %	add --raw-leaves --chunker rabin-32768-65536-131072 -r				
	rabin-8k-16k-64k	3'943'588 KB	108.64 %	108.61 %	117.22 %	add --raw-leaves --chunker rabin-8192-16384-65536 -r				
	rabin-12k-24k-48k	3'889'632 KB	107.15 %	107.12 %	115.62 %	add --raw-leaves --chunker rabin-12288-24576-49152 -r				

Category	File Size	Original Size	Retention %	Success %	Space Saved	Command	Source	Options	Destination	Count	
Linux src code 4x	256K	2'129'532 KB	46.44 %	100.00 %	285.12 %	add --raw-leaves -r	Linux git tags for each version without .git folder	tar cf test.tar *	plain files	linux-src/linux-5.10/... linux-src/linux-5.11/... linux-src/linux-5.12/... linux-src/linux-5.13/...	
	128K	2'125'692 KB	46.36 %	99.82 %	284.61 %	add --raw-leaves --chunker size-131072 -r		zstd -1 --long test.tar			
	8K	2'122'304 KB	46.28 %	99.66 %	284.15 %	add --raw-leaves --chunker size-8192 -r					
	buzhash	2'126'716 KB	46.38 %	99.87 %	284.74 %	add --raw-leaves --chunker buzhash -r			746'892 KB		4'585'528 KB
	rabin	2'128'360 KB	46.37 %	99.85 %	284.69 %	add --raw-leaves --chunker rabin -r					
	rabin-4k-8k-16k	2'069'780 KB	45.14 %	97.19 %	277.12 %	add --raw-leaves --chunker rabin-4096-8192-16384 -r					
	rabin-16k-32k-64k	2'076'884 KB	45.29 %	97.53 %	278.07 %	add --raw-leaves --chunker rabin-16384-32768-65536 -r					
	rabin-32k-64k-128k	2'107'312 KB	45.96 %	98.96 %	282.14 %	add --raw-leaves --chunker rabin-32768-65536-131072 -r					
	rabin-8k-16k-64k	2'057'944 KB	44.88 %	96.64 %	275.53 %	add --raw-leaves --chunker rabin-8192-16384-65536 -r					
	rabin-12k-24k-48k	2'048'992 KB	44.68 %	96.22 %	274.34 %	add --raw-leaves --chunker rabin-12288-24576-49152 -r					
webpages	256K	756'956 KB	94.31 %	100.00 %	439.57 %	add --raw-leaves -r	Randomly crawled webpages including fonts, js, xml, json, css, rss	tar cf test.tar *	plain files	3220 js 2788 html 465 woff2 373 json 221 xml 201 woff 128 eot 118 php 83 svg 80 ttf 79 txt 78 css 55 rss 50 asc 46 mjs 27 patch 27 diff 24 atom 17 webmanifest 11 pdf 8 otf 7 cast 6 cvrf 5 md 5 m3u8 4 m4v 4 dtd 2 readme 2 rdf 2 cgi 1 xsd 1 tmp 1 spx 1 sh 1 metalink 1 meta4 1 key 1 in 1 flatpakrepo 1 editorconfig 1 djuv 1 csv 1 cer 1 apk 1 am 1 ac 1 11	
	128K	755'004 KB	94.07 %	99.74 %	438.44 %	add --raw-leaves --chunker size-131072 -r		zstd -1 --long test.tar			
	8K	761'092 KB	94.83 %	100.55 %	441.97 %	add --raw-leaves --chunker size-8192 -r			172'204 KB		802'592 KB
	buzhash	741'056 KB	92.33 %	97.90 %	430.34 %	add --raw-leaves --chunker buzhash -r					
	rabin	749'260 KB	93.36 %	98.98 %	435.10 %	add --raw-leaves --chunker rabin -r					
	rabin-4k-8k-16k	760'432 KB	94.75 %	100.46 %	441.59 %	add --raw-leaves --chunker rabin-4096-8192-16384 -r					
	rabin-16k-32k-64k	727'648 KB	90.66 %	96.13 %	422.55 %	add --raw-leaves --chunker rabin-16384-32768-65536 -r					
	rabin-32k-64k-128k	737'376 KB	91.87 %	97.41 %	428.20 %	add --raw-leaves --chunker rabin-32768-65536-131072 -r					
	rabin-8k-16k-64k	735'580 KB	91.65 %	97.18 %	427.16 %	add --raw-leaves --chunker rabin-8192-16384-65536 -r					
	rabin-12k-24k-48k	731'524 KB	91.15 %	96.64 %	424.80 %	add --raw-leaves --chunker rabin-12288-24576-49152 -r					
Percent based avg Sum of used space											
	256K		81.59 %		28'901'940 KB						
	128K		81.35 %		28'589'088 KB						
	8K		80.85 %		27'860'228 KB						
	buzhash		76.24 %		27'060'364 KB						
	rabin		79.48 %		31'965'800 KB						
	rabin-4k-8k-16k		82.68 %		31'666'628 KB						
	rabin-16k-32k-64k		77.20 %		28'327'820 KB						
	rabin-32k-64k-128k		77.42 %		28'642'804 KB						
	rabin-8k-16k-64k		79.11 %		29'809'172 KB						
	rabin-12k-24k-48k		77.60 %		28'269'684 KB						
Limited to cases where deduplication was successful (below 95%) - ignoring 'Linux kernel git' and 'Linux kernel pkgs 4x'											
Percent based avg Sum of used space											
	256K		75.44 %		24'842'084 KB						
	128K		75.13 %		24'528'396 KB						
	8K		74.33 %		23'770'252 KB						
	buzhash		68.27 %		22'975'224 KB						
	rabin		72.55 %		27'887'476 KB						
	rabin-4k-8k-16k		72.11 %		26'950'864 KB						
	rabin-16k-32k-64k		68.53 %		24'115'224 KB						
	rabin-32k-64k-128k		69.42 %		24'508'712 KB						
	rabin-8k-16k-64k		69.71 %		25'411'264 KB						
	rabin-12k-24k-48k		68.28 %		23'934'544 KB						