

A Repository Statistics

In the following, numbers from the repository of various different projects are summarized. It looks at the storage properties of the different objects; how much disk space they take unpacked and packed also accounting for the block based nature of the file system. All sizes are in bytes.

Tig (<http://jonas.nitro.dk/tig/tig.git>): Tig is a small and young project with only one developer and not a lot of activity. It has no subdirectories and changes mostly touch only one file which can be seen from the fact that the number of commit, tree, and blob objects are roughly the same.

ELinks (<http://elinks.cz/elinks.git>): ELinks is bigger compared to Tig but is still not a very active project. It has 4-5 developers that participate from time to time. The project tree is structured into many directories containing few and often small files.

Git (<git://git.kernel.org/pub/scm/git/git.git#next>): Git is a very active project with a lot of different developers. It has frequent releases and the maintainer of the repository makes heavy use of topic branches to separate his own work and patches from other developers. It has a lot of files that are distributed over a small handful of directories. The *next* branch of the repository is the entry point for new functionality and tends to have more experimental features.

Cogito (<git://git.kernel.org/pub/scm/cogito/cogito.git>): Cogito started out as a fork of Git and so they share some of the repository history. However today Cogito does not share anything with Git in terms of code and consequently its source tree is much smaller with only less files since many is autogenerated. It is not very active and development is done primarily by one person.

Linux (<git://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux-2.6.git>): The Linux kernel is also a very active project with lots of developers. Release cycles are longer with ideally many changes merged into the repository in the start and fewer and smaller changes at the end of the cycle. The kernels directory tree is very big spanning several directories and a variety of different file types.

B Object Store Statistics

		Tag	Commit	Tree	Blob	Total
Objects		36	10392	40007	30518	80953
Sizes	Min size	468	149	53	24	24
	Max size	1516	2338	6964	68599	68599
	Average size	668	292	1807	6243	3121
Unpacked total	Overhead %	83	92	65	32	49
Packed total	Overhead %	24	1	1	0	0
Compression	level %	0	2	83	89	87

Figure 10: Summary of the repositories. The numbers are averages of the repositories in Figure 11, 12, 13, 14, and 15 with no respect to how big the repositories are compared to each other.

		Tag	Commit	Tree	Blob	Total
Objects		6	250	250	306	812
Sizes	Min size	133	144	81	50	50
	Max size	293	426	445	23668	23668
	Average size	187	219	271	13131	5101
Unpacked total	Size	1123	54866	67894	4018335	4142218
	On-disk size	24576	1024000	1024000	4706304	6778880
	Overhead %	95	94	93	14	38
Packed total	Size	1130	53603	26572	232805	314014
	On-disk size	4096	57344	28672	233472	315392
	Overhead %	72	6	7	0	0
Compression	Saved	-7	1263	41322	3785530	3828204
	Level %	0	2	60	94	92

Figure 11: The Tig repository.

		Tag	Commit	Tree	Blob	Total
Objects		3	2105	7944	5460	15512
Sizes	Min size	1459	155	49	21	21
	Max size	2465	2413	1319	102476	102476
	Average size	2119	237	654	4883	2087
Unpacked total	Size	6358	500449	5202037	26666162	32375006
	On-disk size	12288	8622080	32538624	41246720	82419712
	Overhead %	48	94	84	35	60
Packed total	Size	6376	484667	857492	4099820	5448259
	On-disk size	8192	487424	860160	4100096	5451776
	Overhead %	22	0	0	0	0
Compression	Saved	-18	15782	4344545	22566342	26926747
	Level %	0	3	83	84	83

Figure 12: The ELinks repository.

		Tag	Commit	Tree	Blob	Total
Objects		94	6907	9228	11668	27897
Sizes	Min size	275	137	47	17	17
	Max size	3397	2573	7262	42431	42431
	Average size	352	336	4241	4155	3225
Unpacked total	Size	33102	2325702	39139521	48489993	89988318
	On-disk size	385024	28291072	61480960	76812288	166969344
	Overhead %	91	91	36	36	46
Packed total	Size	32736	2269741	2268268	4032317	8602966
	On-disk size	32768	2273280	2269184	4034560	8605696
	Overhead %	0	0	0	0	0
Compression	Saved	366	55961	36871253	44457676	81385352
	Level %	1	2	94	91	90

Figure 13: The Git repository.

		Tag	Commit	Tree	Blob	Total
Objects		25	3139	3608	5646	12418
Sizes	Min size	236	158	47	17	17
	Max size	1057	1354	5482	19386	19386
	Average size	391	294	2056	2187	1667
Unpacked total	Size	9795	924853	7418479	12347802	20700929
	On-disk size	102400	12857344	15724544	26644480	55328768
	Overhead %	90	92	52	53	62
Packed total	Size	9711	905476	693125	1452865	3061081
	On-disk size	12288	909312	696320	1454080	3063808
	Overhead %	20	0	0	0	0
Compression	Saved	84	19377	6725354	10894937	17639848
	Level %	0	2	90	88	85

Figure 14: The Cogito repository.

		Tag	Commit	Tree	Blob	Total
Objects		52	39559	179006	129512	348129
Sizes	Min size	240	151	45	15	15
	Max size	372	4928	20315	155034	155034
	Average size	292	374	1817	6860	3529
Unpacked total	Size	15200	14801850	325350244	888514445	1228681739
	On-disk size	212992	162037760	849793024	1180659712	2192703488
	Overhead %	92	90	61	24	43
Packed total	Size	14988	14505268	27322674	92305137	134147971
	On-disk size	16384	14508032	27324416	92307456	134148096
	Overhead %	8	0	0	0	0
Compression	Saved	212	296582	298027570	796209308	1094533768
	Level %	1	2	91	89	89

Figure 15: The Linux repository

B.1 Packing Statistics

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	2	22219	22455	22337	223	3488	3488	3488	34
250	1	53603	53603	53603	214	7064	7064	7064	28

Figure 16: Tig: commits objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	2	9258	12973	11115	111	3488	3488	3488	34
250	1	26572	26572	26572	106	7064	7064	7064	28

Figure 17: Tig: trees objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	3	66568	92681	76090	760	3488	3488	3488	34
250	1	168750	168750	168750	675	7088	7088	7088	28
500	1	232805	232805	232805	760	8408	8408	8408	27

Figure 18: Tig: blobs objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	20	21487	28600	23546	235	3488	3488	3488	34
250	8	53200	65376	58240	232	7088	7088	7088	28
500	4	109979	126334	115635	231	13088	13088	13088	26
1000	2	222330	240077	231203	231	25088	25088	25088	25
2500	1	484667	484667	484667	230	51584	51584	51584	24

Figure 19: ELinks: commits objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	78	11322	34921	19429	194	3488	3488	3488	34
250	31	26668	49925	38762	155	7088	7088	7088	28
500	15	53745	81759	67564	135	13088	13088	13088	26
1000	7	104432	132110	122780	122	25088	25088	25088	25
2500	3	274427	292527	282101	112	61088	61088	61088	24
5000	1	552262	552262	552262	110	121088	121088	121088	24
7500	1	809705	809705	809705	107	181088	181088	181088	24
10000	1	857492	857492	857492	107	191720	191720	191720	24

Figure 20: ELinks: trees objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	54	17696	861976	269622	2696	3488	3488	3488	34
250	21	129990	1138364	569471	2277	7088	7088	7088	28
500	10	637347	1509524	999860	1999	13088	13088	13088	26
1000	5	1244127	2380150	1686222	1686	25088	25088	25088	25
2500	2	2318470	3124134	2721302	1088	61088	61088	61088	24
5000	1	4020395	4020395	4020395	804	121088	121088	121088	24
7500	1	4099820	4099820	4099820	750	132104	132104	132104	24

Figure 21: ELinks: blobs objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	31	21874	35571	29127	291	3488	3488	3488	34
250	12	58791	83833	72324	289	7088	7088	7088	28
500	6	117461	163355	144473	288	13088	13088	13088	26
1000	3	242966	316590	288383	288	25088	25088	25088	25
2500	1	722792	722792	722792	289	61088	61088	61088	24
5000	1	905476	905476	905476	288	76400	76400	76400	24
7500	1	905476	905476	905476	288	76400	76400	76400	24
10000	1	905476	905476	905476	288	76400	76400	76400	24

Figure 22: Cogito: commits objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	35	12551	28887	19851	198	3488	3488	3488	34
250	14	34172	60444	46277	185	7088	7088	7088	28
500	7	70116	104942	89291	178	13088	13088	13088	26
1000	3	140200	205602	182673	182	25088	25088	25088	25
2500	1	465459	465459	465459	186	61088	61088	61088	24
5000	1	693125	693125	693125	192	87656	87656	87656	24
7500	1	693125	693125	693125	192	87656	87656	87656	24
10000	1	693125	693125	693125	192	87656	87656	87656	24

Figure 23: Cogito: trees objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	55	41058	183950	94713	947	3488	3488	3488	34
250	22	70848	328476	158665	634	7088	7088	7088	28
500	11	120656	447441	236691	473	13088	13088	13088	26
1000	5	278616	543657	396053	396	25088	25088	25088	25
2500	2	753280	843096	798188	319	61088	61088	61088	24
5000	1	1334397	1334397	1334397	266	121088	121088	121088	24
7500	1	1452865	1452865	1452865	257	136568	136568	136568	24
10000	1	1452865	1452865	1452865	257	136568	136568	136568	24

Figure 24: Cogito: blobs objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	68	19235	39021	33040	330	3488	3488	3488	34
250	27	75020	87740	82024	328	7088	7088	7088	28
500	13	155825	174299	164208	328	13088	13088	13088	26
1000	6	313521	338898	327021	327	25088	25088	25088	25
2500	2	815744	835220	825482	330	61088	61088	61088	24
5000	1	1651721	1651721	1651721	330	121088	121088	121088	24
7500	1	2269741	2269741	2269741	328	166832	166832	166832	24
10000	1	2269741	2269741	2269741	328	166832	166832	166832	24

Figure 25: Git: commits objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	91	13804	46791	29908	299	3488	3488	3488	34
250	36	40126	90801	65303	261	7088	7088	7088	28
500	18	81340	150696	125441	250	13088	13088	13088	26
1000	9	163450	311692	251298	251	25088	25088	25088	25
2500	3	634053	701867	663930	265	61088	61088	61088	24
5000	1	1349789	1349789	1349789	269	121088	121088	121088	24
7500	1	1987850	1987850	1987850	265	181088	181088	181088	24
10000	1	2268268	2268268	2268268	245	222536	222536	222536	24

Figure 26: Git: trees objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	115	22560	363899	186407	1864	3488	3488	3488	34
250	46	85520	610518	334384	1337	7088	7088	7088	28
500	23	178519	1056426	499930	999	13088	13088	13088	26
1000	11	414614	1236992	731070	731	25088	25088	25088	25
2500	4	1050814	1562238	1282960	513	61088	61088	61088	24
5000	2	1912371	2207070	2059720	411	121088	121088	121088	24
7500	1	2893875	2893875	2893875	385	181088	181088	181088	24
10000	1	3617966	3617966	3617966	361	241088	241088	241088	24

Figure 27: Git: blobs objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	391	23549	52391	36287	362	3488	3488	3488	34
250	157	71634	113087	90559	362	7088	7088	7088	28
500	78	149853	210706	181430	362	13088	13088	13088	26
1000	39	326777	399818	363791	363	25088	25088	25088	25
2500	15	841399	968026	912950	365	61088	61088	61088	24
5000	7	1719551	1909713	1827476	365	121088	121088	121088	24
7500	5	2605356	2829994	2749475	366	181088	181088	181088	24
10000	3	3503598	3777206	3654637	365	241088	241088	241088	24

Figure 28: Linux: commits objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	1772	11575	189614	64252	642	3488	3488	3488	34
250	713	34502	293777	120610	482	7088	7088	7088	28
500	357	62518	414010	194826	389	13088	13088	13088	26
1000	178	166070	607265	320720	320	25088	25088	25088	25
2500	71	428839	961217	613127	245	61088	61088	61088	24
5000	35	877941	1308072	1055048	211	121088	121088	121088	24
7500	23	1219082	1745041	1448135	193	181088	181088	181088	24
10000	17	1625123	2109295	1851864	185	241088	241088	241088	24

Figure 29: Linux: trees objects.

		Pack file sizes				Index file sizes			
Objects	Packs	Min	Max	Avg	Pr obj	Min	Max	Avg	Pr obj
100	1282	45800	1861636	536677	5366	3488	3488	3488	34
250	515	162037	3606412	1275512	5102	7088	7088	7088	28
500	258	368162	5989508	2458778	4917	13088	13088	13088	26
1000	129	754650	9506100	4667246	4667	25088	25088	25088	25
2500	51	2714589	18526695	10442205	4176	61088	61088	61088	24
5000	25	9562550	27554397	18044311	3608	121088	121088	121088	24
7500	17	17851561	30285835	23881460	3184	181088	181088	181088	24
10000	12	23699091	35751651	28709864	2870	241088	241088	241088	24

Figure 30: Linux: blobs objects.