

Developing the Raster Big Data Benchmark: A Comparison of Raster Analysis on Big Data Platforms

David Haynes ^{1,*}, Philip Mitchell ² and Eric Shook ³

¹ Institute for Health Informatics, University of Minnesota, Minneapolis, MN 55455, USA

² Ali I. Al-Naimi Petroleum Engineering Research Center, King Abdullah University of Science and Technology, Thuwal 23955, Kingdom of Saudi Arabia; mitc0415@umn.edu

³ Geography Environment and Society; University of Minnesota, Minneapolis, MN 55455, USA, eshook@umn.edu

* Correspondence: dahaynes@umn.edu

Table S1. Pixel Count Performance Times on all Platforms all Datasets.

GLC					MERIS				NLCD			
Partition size	PostGIS	GeoTrellis	GeoTrellis (Cached)	SCIDB	PostGIS	GeoTrellis s	GeoTrellis (Cached)	SCID B	PostGIS	GeoTrellis s	GeoTrellis (Cached)	SCIDB
25		1.498	0.080			10.548	0.207			97.047	10.917	
50	2.611	1.231	0.053		16.164	10.010	0.129		1184.822	84.956	6.112	
100	2.009	1.183	0.042		13.680	10.317	0.104		1039.693	86.418	5.101	
200	1.893	1.316	0.034		12.834	11.508	0.103		1000.756	100.159	5.305	
300	1.876	1.492	0.048		12.526	12.885	0.098		994.960	119.041	5.767	
400	1.823	1.750	0.045		13.392	14.722	0.105		990.351	137.887	6.281	
500	1.844	1.965	0.037	0.325	12.562	16.704	0.094	0.581	991.586	167.465	6.942	17.603
600	2.025	2.215	0.038		13.356	18.864	0.103		996.428	196.552	7.454	
700	1.852	2.603	0.041		13.545	21.478	0.100		991.928	231.496	7.961	
800	2.083	2.965	0.045		13.665	24.712	0.116		999.223	258.017	8.461	
900	2.021	3.532	0.052		12.784	27.174	0.102		1005.009	309.726	9.123	
1000	2.205	3.714	0.054	0.200	13.922	30.229	0.106	0.564	1005.128	355.632	9.849	14.363
1500	1.848			0.189	13.121			0.428	1002.517			21.935
2000	2.435			0.192	17.032			0.489	1032.112			19.495
2500				0.244				0.528				19.856
3000				0.211				0.481				20.794
3500				0.239				0.503				21.867
4000				0.236								20.955

Table S2. Focal Performance Times on all Platforms all Datasets

GLC					
Partition size	PostGIS	GeoTrellis	GeoTrellis (Cached)	SCIDB	SciDB_overlap
25		1.563	0.135		
50	118.883	1.272	0.124		
100	123.314	1.309	0.114		
200	126.583	1.393	0.109		
300	128.053	1.581	0.138		
400	127.472	1.821	0.146		
500	128.040	2.086	0.156	8.380	4.420
600	128.084	2.355	0.179		
700	129.083	2.689	0.188		
800	128.354	3.263	0.208		
900	128.390	3.524	0.272		
1000	128.422	3.828	0.275	14.148	8.243
1500				19.708	7.223
MERIS					
Partition size	PostGIS	GeoTrellis	GeoTrellis (Cached)	SCIDB	SciDB_overlap
25		11.058	0.608		
50	1231.641	10.670	0.570		
100	1280.320	10.763	0.520		
200	1311.697	11.984	0.547		
300	1317.308	13.446	0.592		
400	1322.023	15.375	0.745		
500	1327.230	17.291	0.721	37.672	8.501
600	1327.812	19.693	0.738		
700	1330.271	22.085	0.681		
800	1327.911	25.583	0.780		
900	1333.428	27.942	0.774		
1000	1336.509	31.150	0.785	39.628	37.444
1500				42.295	---
NLCD					
Partition size	PostGIS	GeoTrellis	GeoTrellis (Cached)	SCIDB	SciDB_overlap
25		137.310	52.299		

50	122.124	44.903		
100	124.050	43.732		
200	140.795	47.232		
300	162.284	50.222		
400	183.028	53.012		
500	214.100	55.736	---	---
600	244.912	57.261		
700	280.607	58.864		
800	308.179	59.607		
900	359.523	61.130		
1000	407.264	62.609	---	---
1500				

Table S3. Polygonal Summaries on all Platforms for all Datasets

GLC		PostGIS			SciDB Join Only				SciDB All Operations				GeoTrellis		
Partition size	regions	states	counties	tracts	regions	states	counties	tracts	regions	states	counties	tracts	states	counties	tracts
25													257.76	1881.55	40685.1
50	4.85	5.69	13.51	106.60									69.21	619.06	12982.8
100	2.74	3.89	9.84	78.88									21.17	242.25	4897.61
200	2.24	3.85	8.85	73.48									7.77	100.15	2205.59
300	2.48	4.33	9.21	86.45									5.49	78.06	1747.58
400	2.41	4.70	10.20	106.00									4.60	70.39	1579.98
500	2.61	4.82	11.58	134.25									4.31	68.79	1486.75
600	2.74	4.99	13.29	166.94									4.33	75.33	1571.24
700	3.21	5.00	15.36	210.62									4.39	80.38	1647.18
800	3.75	5.07	17.76	260.27									4.57	82.52	1726.48
900	3.95	5.19	20.84	321.47	4.81	97.42	1881.31								
1000	3.75	5.43	23.43	372.19	1.73	1.60	1.67	1.55	16.06	16.81	16.65	22.56	4.96	94.26	1909.86
1500					1.81	1.95	1.93	1.86	16.15	16.15	15.32	23.28			
2000					2.29	2.37	2.41	2.35	16.12	16.66	17.42	30.88			
2500					3.45	3.40	3.40	3.34	18.79	18.40	18.30	34.80			
3000					4.35	4.29	4.26	4.37	21.84	23.53	21.83	27.54			
3500					4.991	5.033	4.936	4.948	22.491	24.544	23.306	29.860			
4000					5.532	5.550	5.584	5.467	23.274	26.491	24.858	43.031			
MERIS		PostGIS			SciDB Join Only				SciDB All Operations				GeoTrellis		
Partition size	regions	states	counties	tracts	regions	states	counties	tracts	regions	states	counties	tracts	states	counties	tracts
25													17025.81	2605.76	380758.74
50	29.19	17.95	35.58	205.84									4610.52	678.45	101375.91
100	12.59	13.11	37.72	163.39									1379.44	185.95	30047.95
200	8.45	12.61	42.87	149.54									522.58	60.97	11309.04
300	8.02	14.91	45.67	160.30									359.90	37.52	7602.61

400	8.25	15.98	48.58	182.26	9.21	9.33	9.11	9.24	109.63	131.42	133.56	170.66	312.80	32.18	6524.13								
500	9.15	20.96	50.94	211.98									265.54	29.25	5542.39								
600	9.79	21.89	53.82	253.80									233.64	29.18	4957.02								
700	10.22	25.62	57.12	305.09									227.49	30.39	4837.95								
800	10.26	26.06	59.86	358.73									244.42	33.91	5201.31								
900	10.96	28.57	64.03	425.06	9.76	9.72	9.71	9.76	126.60	112.14	120.60	190.55	228.50	35.90	4699.39								
1000	13.30	30.90	68.28	498.24									244.88	37.89	5268.79								
1500													11.20	11.46	10.85	11.10	121.77	130.77	115.43	157.33			
2000													11.67	11.61	11.61	11.43	121.92	110.03	118.53	222.65			
2500													12.86	12.80	12.97	12.83	142.16	132.85	118.29	203.43			
3000					14.30	14.11	14.38	14.22	170.52	193.08	122.11	255.34											
3500					13.29	13.43	13.36	13.25	145.26	132.03	136.39	141.79											
4000					16.82	17.51	16.53	16.73	141.18	142.98	135.49	299.22											
NLCD		PostGIS			SciDB Join Only				SciDB All Operations				GeoTrellis										
Partition size	regions	states	counties	tracts	regions	states	counties	tracts	regions	states	counties	tracts	states	counties	tracts								
25													---	---	---								
50	2090	873	520	1303									---	---	---								
100	800	533	514	1144									---	---	---								
200	468	432	629	1390									---	---	---								
300	404	413	783	1711									---	---	---								
400	385	414	954	2013	39.5	39.3	39.5	42.2	13365.3	11549.8	16045.2	20641.8	---	---	---								
500	386	422	1130	2361									---	---	---								
600	377	423	1311	2574									---	---	---								
700	380	439	1489	2791									---	---	---								
800	377	453	1641	2973									---	---	---								
900	379	461	1807	3170	60.0	60.4	59.7	62.3	15302.1	17836.7	17536.2	19319.1	---	---	---								
1000	378	481	1944	3367									---	---	---								
1500													74.5	74.8	74.4	75.7	16612.9	16732.7	16802.4	19457.6			
2000													90.4	90.4	91.1	91.1	16450.0	16855.3	15461.1	18509.7			

2500	121.9	121.9	122.1	123.8	16538.2	15568.9	15960.0	18384.0
3000	144.3	144.1	146.2	144.9	17623.4	18052.5	17653.1	18703.4
3500	169.1	169.4	171.3	171.0	15337.7	17863.2	16750.4	22298.1
4000	175.2	175.1	175.7	177.0	16477.3	17026.3	18416.2	20002.9

Table S4. SciDB Zonal Operation Components by Percentage of Time

Raster	Vector	Rasterize	Load	Redimension	Join
GLC	states	7.9%	31.7%	35.7%	24.7%
	counties	7.6%	32.0%	36.9%	23.4%
	tracts	31.8%	31.9%	21.5%	14.9%
MERIS	states	31.6%	19.2%	36.3%	12.8%
	counties	36.6%	17.5%	34.8%	11.1%
	tracts	46.6%	18.0%	28.4%	7.0%
NLCD	states	20.0%	9.3%	65.7%	5.1%
	counties	22.5%	8.8%	63.7%	5.0%
	tracts	44.4%	4.6%	46.7%	4.2%