

Spatial Partitioning Techniques in SpatialHadoop

Ahmed Eldawy Louai Alarabi Mohamed F. Mokbel Department of Computer Science and Engineering University of Minnesota

Big Spatial Data in HDFS

Spatial Applications

Brain simulation, Climate changes Event detection and analysis, GIS application Volunteer GIS applications



Spatial Partitioning Techniques













Experimental Setup

Summary of techniques

Technique Partitioning Bounda								
Grid	Space	Replication						
Quad-tr	e Space	Replication						
STR	Data	Distribution						
STR+	Data	Replication						
K-d tree	Data	Replication						
Z-Curve	SFC	Distribution						
Hilbert	SFC	Distribution						

Datasets

Dataset	Size	Record
All Object	s 88GB	250M
Buildings	25GB	109M
Roads	23GB	70M
Lakes	8.6GB	9M
Cities	1.4GB	170K

Quality Measures

 $Q1 = \Sigma(P.width \times P.height)$

Measures the amount of dead area

$Q2 = \Sigma$ (Partitions overlap)

Measures the overlap between partitions

$Q3 = \Sigma(P.width + P.height)$

Measures the squareness of partitions

$Q4 = \Sigma(P.size) / \Sigma(Block size)$

Disk utilization

Q5=Std Dev(P.count)

Skewness of partitions (Load balance)

Range Query

Throughput, running time **Spatial Join** Running time

Quality Measures





Partition techniques are relatively stable across datasets





Range Query





Bitonic behavior due to the tradeoff between quality and load balance



Single machine is preferable over MapReduce for small queries

The block size with peak performance depends on the query size



The speedup of range query is capped by number of simultaneous queries the system runs



Tradeoff between Q1 and Q4 with HDFS block size



Spatial Join	
High correlation	
between Q1 and	
the performance	
of spatial join	

		Lakes (8.6GB)							Buildings (25GB)					
		Q1	0.75	0.89	0.89	1.47	1.47	2.71	0.50	0.71	0.78	0.71	1.07	1.81
	Q1		Quad	STR+	K-d	STR	Hilbert	Zcurve	Quad	STR+	K-d	STR	Hilbert	Zcurve
3)	0.52	Quad	276	440	463	675	749	1446	628	919	888	895	1054	1633
Ð	0.66	STR+	419	381	439	577	613	1048	801	740	834	763	960	1346
23	0.71	K-d	465	404	416	560	664	1058	787	836	735	782	935	1321
ls (0.71	STR	409	401	437	570	599	1031	834	765	795	733	928	1321
ad	1.03	Hilbert	512	464	491	612	630	1100	1248	1085	1026	1030	1184	1646
Ro	1.88	Zcurve	740	589	608	837	780	1246	1979	1587	1519	1519	1644	2223

