

SparkRDF: Elastic Discreted RDF Graph Processing Engine With Distributed Memory

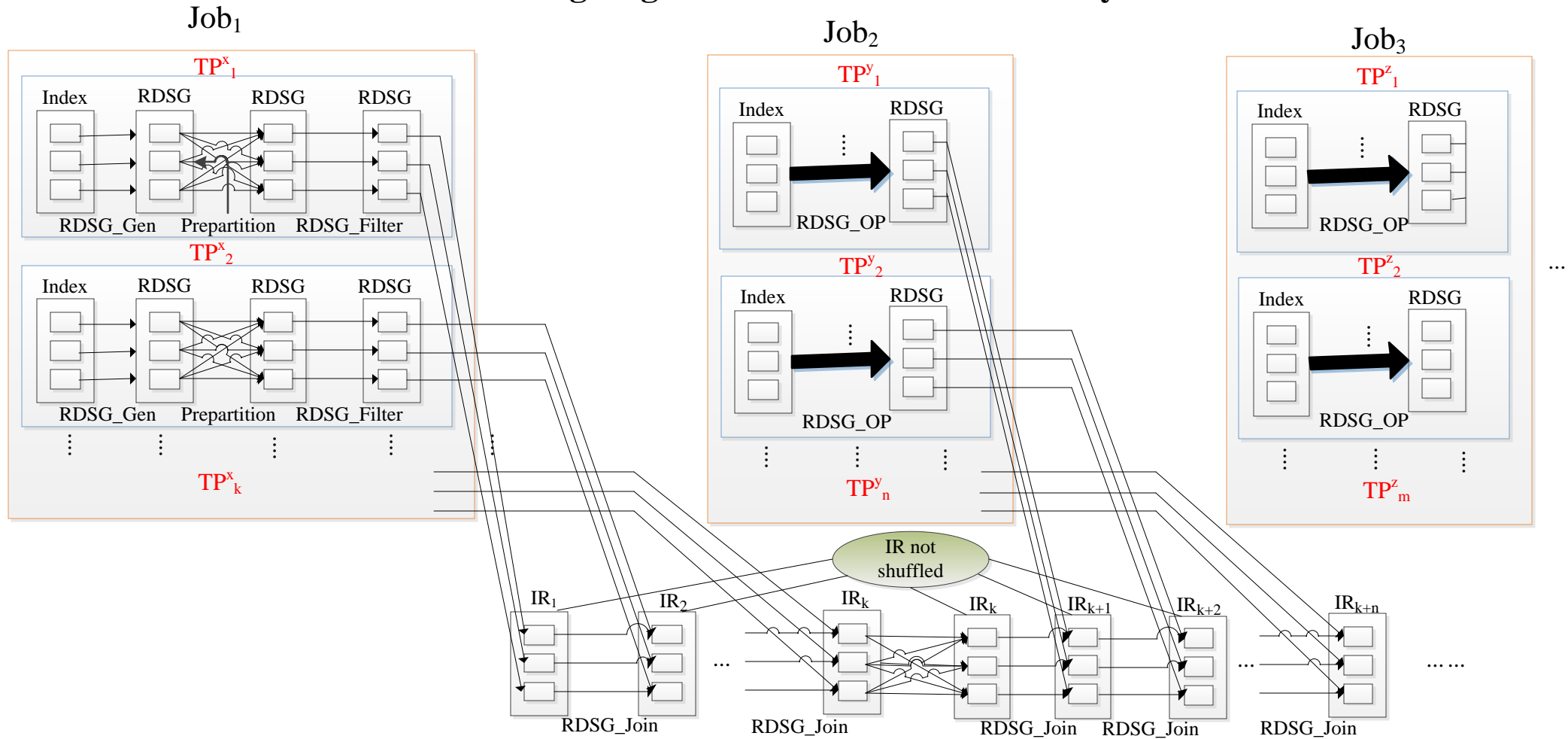


Fig.1. Iterative Query Model of SparkRDF

	LUBM-10000			LUBM-20000			LUBM-30000		
	cluster systems		centralized system	cluster systems		centralized system	cluster systems		centralized system
	SRDF	HRDF	RDF3X	SRDF	HRDF	RDF3X	SRDF	HRDF	RDF3X
Q1	478.5	8475.4	2131.4	1123.2	>3h	4380.3	1435.4	>3.5h	failed
Q2	11.9	3425.2	13.8	25.8	>2h	28.9	40.3	>2.5h	43.5
Q3	1.4	6869.7	24.6	1.4	>2.5h	90.7	1.4	>3h	failed
Q4	14.4	11940.3	0.7	23.8	>4h	0.8	32.5	>8h	0.8
Q5	6.8	2587.5	0.7	10.9	>1h	0.7	13.0	>3h	0.7
Q6	10.3	7210.5	0.6	16.4	>2.5h	0.7	20.0	>3h	0.7
Q7	54.6	1911.2	101.5	112.5	>0.7h	198.5	201.3	>1h	853.0

Fig.2. Performance comparison for SparkRDF(SRDF), HadoopRDF(HRDF) and RDF3X