

# Machine Learning techniques to Model Data Intensive Application Performance

A. Battistello      P. Ferretti

## Contents

<b>1</b>	<b>Fixed Datasize, all the features</b>	<b>4</b>
1.1	Query R1 . . . . .	4
1.1.1	Query R1 — Datasize 250GB . . . . .	4
1.1.2	Query R1 — Datasize 500GB . . . . .	5
1.1.3	Query R1 — Datasize 750GB . . . . .	6
1.1.4	Query R1 — Datasize 1000GB . . . . .	7
1.2	Query R2 . . . . .	8
1.2.1	Query R2 — Datasize 250GB . . . . .	8
1.2.2	Query R2 — Datasize 500GB . . . . .	9
1.2.3	Query R2 — Datasize 750GB . . . . .	10
1.2.4	Query R2 — Datasize 1000GB . . . . .	11
1.3	Query R3 . . . . .	12
1.3.1	Query R3 — Datasize 250GB . . . . .	12
1.3.2	Query R3 — Datasize 500GB . . . . .	13
1.3.3	Query R3 — Datasize 750GB . . . . .	14
1.3.4	Query R3 — Datasize 1000GB . . . . .	15
1.4	Query R4 . . . . .	16
1.4.1	Query R4 — Datasize 250GB . . . . .	16
1.4.2	Query R4 — Datasize 500GB . . . . .	17
1.4.3	Query R4 — Datasize 750GB . . . . .	18
1.4.4	Query R4 — Datasize 1000GB . . . . .	19
1.5	Query R5 . . . . .	20
1.5.1	Query R5 — Datasize 250GB . . . . .	20
1.5.2	Query R5 — Datasize 500GB . . . . .	21
1.5.3	Query R5 — Datasize 750GB . . . . .	22
1.5.4	Query R5 — Datasize 1000GB . . . . .	23
<b>2</b>	<b>Fixed Datasize, <math>ncores^{-1}</math>, all the features</b>	<b>24</b>
2.1	Query R1 . . . . .	24
2.1.1	Query R1 — Datasize 250GB . . . . .	24
2.1.2	Query R1 — Datasize 500GB . . . . .	25
2.1.3	Query R1 — Datasize 750GB . . . . .	26

2.1.4	Query R1 — Datasize 1000GB . . . . .	27
2.2	Query R2 . . . . .	28
2.2.1	Query R2 — Datasize 250GB . . . . .	28
2.2.2	Query R2 — Datasize 500GB . . . . .	29
2.2.3	Query R2 — Datasize 750GB . . . . .	30
2.2.4	Query R2 — Datasize 1000GB . . . . .	31
2.3	Query R3 . . . . .	32
2.3.1	Query R3 — Datasize 250GB . . . . .	32
2.3.2	Query R3 — Datasize 500GB . . . . .	33
2.3.3	Query R3 — Datasize 750GB . . . . .	34
2.3.4	Query R3 — Datasize 1000GB . . . . .	35
2.4	Query R4 . . . . .	36
2.4.1	Query R4 — Datasize 250GB . . . . .	36
2.4.2	Query R4 — Datasize 500GB . . . . .	37
2.4.3	Query R4 — Datasize 750GB . . . . .	38
2.4.4	Query R4 — Datasize 1000GB . . . . .	39
2.5	Query R5 . . . . .	40
2.5.1	Query R5 — Datasize 250GB . . . . .	40
2.5.2	Query R5 — Datasize 500GB . . . . .	41
2.5.3	Query R5 — Datasize 750GB . . . . .	42
2.5.4	Query R5 — Datasize 1000GB . . . . .	43
<b>3</b>	<b>Query comparison</b>	<b>44</b>
3.1	R5 → R2 (Fast - Fast) . . . . .	44
3.1.1	R5 → R2 — Datasize 250GB . . . . .	44
3.1.2	R5 → R2 — Datasize 500GB . . . . .	45
3.1.3	R5 → R2 — Datasize 750GB . . . . .	46
3.1.4	R5 → R2 — Datasize 1000GB . . . . .	47
3.2	R2 → R5 (Fast - Fast) . . . . .	48
3.2.1	R2 → R5 — Datasize 250GB . . . . .	48
3.2.2	R2 → R5 — Datasize 500GB . . . . .	49
3.2.3	R2 → R5 — Datasize 750GB . . . . .	50
3.2.4	R2 → R5 — Datasize 1000GB . . . . .	51
3.3	R3 → R4 (Slow - Slow) . . . . .	52
3.3.1	R3 → R4 — Datasize 250GB . . . . .	52
3.3.2	R3 → R4 — Datasize 500GB . . . . .	53
3.3.3	R3 → R4 — Datasize 750GB . . . . .	54
3.3.4	R3 → R4 — Datasize 1000GB . . . . .	55
3.4	R4 → R3 (Slow - Slow) . . . . .	56
3.4.1	R4 → R3 — Datasize 250GB . . . . .	56
3.4.2	R4 → R3 — Datasize 500GB . . . . .	57
3.4.3	R4 → R3 — Datasize 750GB . . . . .	58
3.4.4	R4 → R3 — Datasize 1000GB . . . . .	59
3.5	R1, R2, R4 → R3 (Fast and Slow - Average) . . . . .	60

3.5.1	R1,R2,R4 $\rightarrow$ R3 — Datasize 250GB . . . . .	60
3.5.2	R1,R2,R4 $\rightarrow$ R3 — Datasize 500GB . . . . .	61
3.5.3	R1,R2,R4 $\rightarrow$ R3 — Datasize 750GB . . . . .	62
3.5.4	R1,R2,R4 $\rightarrow$ R3 — Datasize 1000GB . . . . .	63
3.6	R2, R5 $\rightarrow$ R3 (Fast - Slow) . . . . .	64
3.6.1	R2,R5 $\rightarrow$ R3 — Datasize 250GB . . . . .	64
3.6.2	R2,R5 $\rightarrow$ R3 — Datasize 500GB . . . . .	65
3.6.3	R2,R5 $\rightarrow$ R3 — Datasize 750GB . . . . .	66
3.6.4	R2,R5 $\rightarrow$ R3 — Datasize 1000GB . . . . .	67
3.7	R3, R4 $\rightarrow$ R2 (Fast - Slow) . . . . .	68
3.7.1	R3,R4 $\rightarrow$ R2 — Datasize 250GB . . . . .	68
3.7.2	R3,R4 $\rightarrow$ R2 — Datasize 500GB . . . . .	69
3.7.3	R3,R4 $\rightarrow$ R2 — Datasize 750GB . . . . .	70
3.7.4	R3,R4 $\rightarrow$ R2 — Datasize 1000GB . . . . .	71
<b>4</b>	<b>Fixed Datasize, all the features,</b>	<b>73</b>
4.1	Query Q2 . . . . .	73
4.1.1	Query Q2 — Datasize 40GB . . . . .	73
4.1.2	Query Q2 — Datasize 50GB . . . . .	74
4.2	Query Q3 . . . . .	75
4.2.1	Query Q3 — Datasize 50GB . . . . .	75
4.3	Query Q4 . . . . .	76
4.3.1	Query Q4 — Datasize 50GB . . . . .	76
<b>5</b>	<b>Fixed Datasize, all features with 1/nCores,</b>	<b>77</b>
5.1	Query Q2 . . . . .	77
5.1.1	Query Q2 — Datasize 40GB . . . . .	77
5.1.2	Query Q2 — Datasize 50GB . . . . .	78
5.2	Query Q3 . . . . .	79
5.2.1	Query Q3 — Datasize 50GB . . . . .	79
5.3	Query Q4 . . . . .	80
5.3.1	Query Q4 — Datasize 50GB . . . . .	80

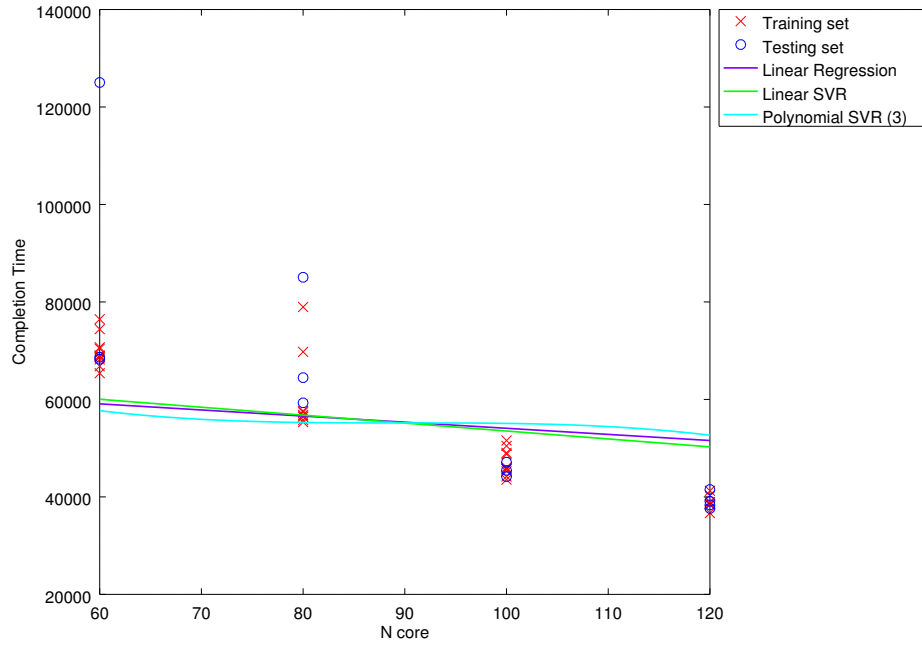
# 1 Fixed Datasize, all the features

## 1.1 Query R1

### 1.1.1 Query R1 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.3041	0.9591	2480	0.0329
Linear SVR	0.3418	0.9830	2820	0.0375
Polynomial SVR (2)	0.7566	0.7509	10643	0.2065
Polynomial SVR (3)	4.2307	0.7818	23208	0.2408
Polynomial SVR (4)	1.1959	0.4622	11066	0.1907
Polynomial SVR (6)	5.3772	0.4964	37535	0.4284
Gaussian SVR	1.2070	0.4548	7808	0.0864

**Table 1:** Results for R1-250

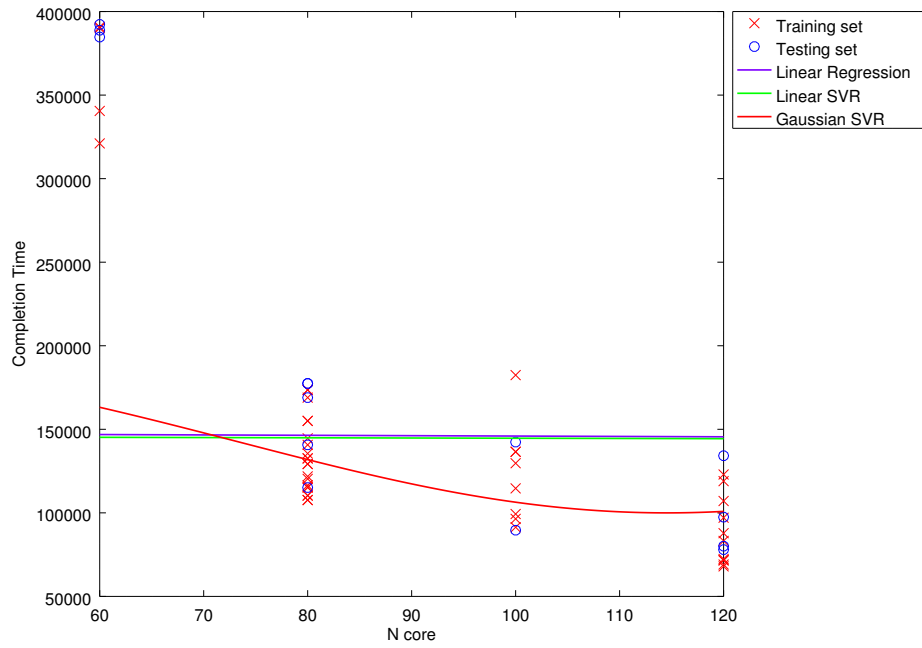


**Figure 1:** Completion time vs ncores for query R1 with datasize 250GB

### 1.1.2 Query R1 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0905	0.9954	5714	0.0293
Linear SVR	0.0994	0.9979	6740	0.0356
Polynomial SVR (2)	0.4862	0.9173	30727	0.1741
Polynomial SVR (3)	0.2537	0.9662	13302	0.0643
Polynomial SVR (4)	0.6772	0.8746	42865	0.2285
Polynomial SVR (6)	1.3799	0.6979	67886	0.3128
Gaussian SVR	0.3571	0.9779	17589	0.0756

**Table 2:** Results for R1-500

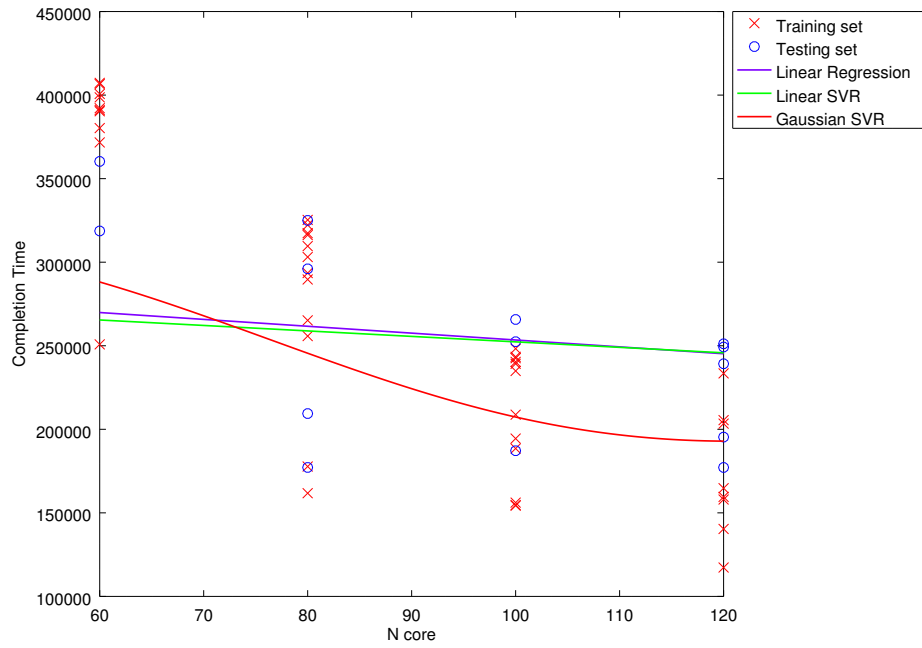


**Figure 2:** Completion time vs ncores for query R1 with datasize 500GB

### 1.1.3 Query R1 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1813	0.9216	12361	0.0474
Linear SVR	0.1727	0.9390	11060	0.0413
Polynomial SVR (2)	0.6893	0.1868	47136	0.1980
Polynomial SVR (3)	0.4048	0.6131	27687	0.1097
Polynomial SVR (4)	0.5559	0.2879	39023	0.1624
Polynomial SVR (6)	0.5809	0.2267	41711	0.1717
Gaussian SVR	0.1855	0.9592	12583	0.0531

**Table 3:** Results for R1-750

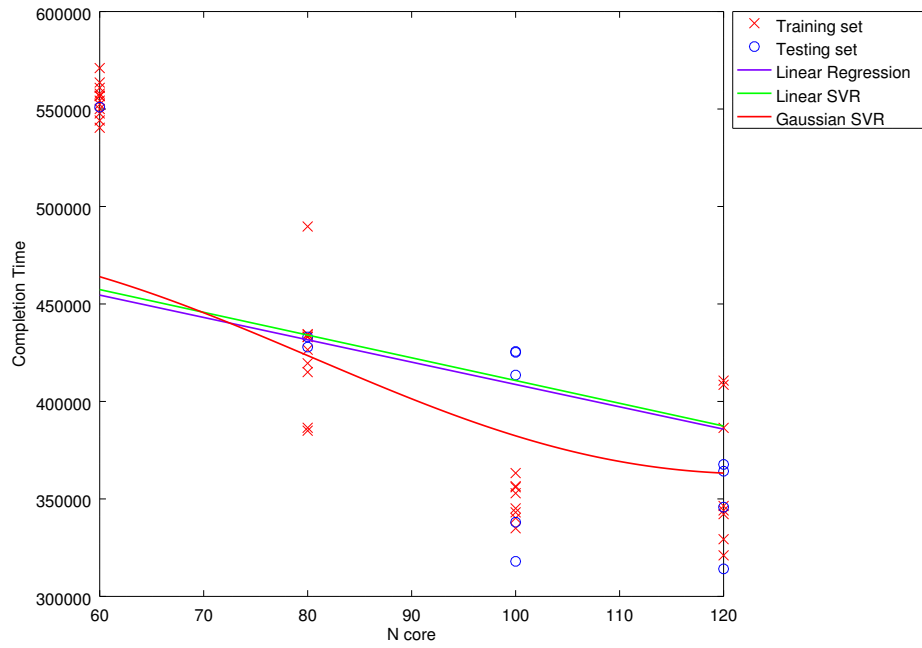


**Figure 3:** Completion time vs ncores for query R1 with datasize 750GB

#### 1.1.4 Query R1 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1784	0.9416	12742	0.0332
Linear SVR	0.1914	0.9451	13208	0.0343
Polynomial SVR (2)	0.6728	0.4426	51407	0.1411
Polynomial SVR (3)	0.4131	0.7854	31112	0.0836
Polynomial SVR (4)	1.0215	0.0334	55894	0.1605
Polynomial SVR (6)	1.1153	0.0093	60608	0.1735
Gaussian SVR	0.3595	0.8131	15070	0.0441

**Table 4:** Results for R1-1000



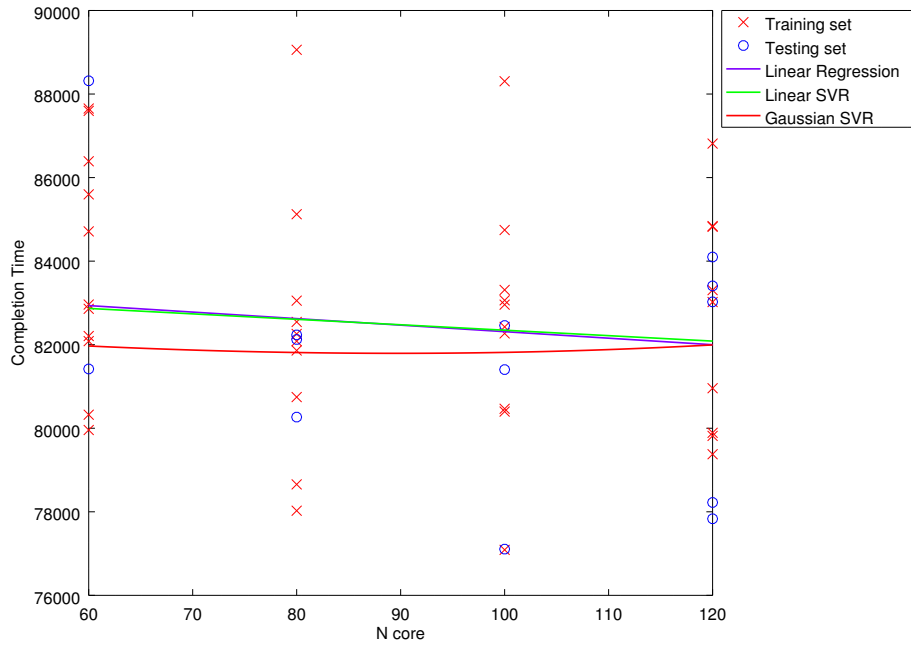
**Figure 4:** Completion time vs ncores for query R1 with datasize 1000GB

## 1.2 Query R2

### 1.2.1 Query R2 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2914	0.9080	609	0.0073
Linear SVR	0.2882	0.9215	622	0.0075
Polynomial SVR (2)	1.0548	0.3165	2442	0.0303
Polynomial SVR (3)	0.5157	0.8003	1189	0.0149
Polynomial SVR (4)	0.9650	0.1946	2218	0.0276
Polynomial SVR (6)	1.0183	0.1074	2319	0.0289
Gaussian SVR	0.4247	0.8933	1021	0.0126

**Table 5:** Results for R2-250



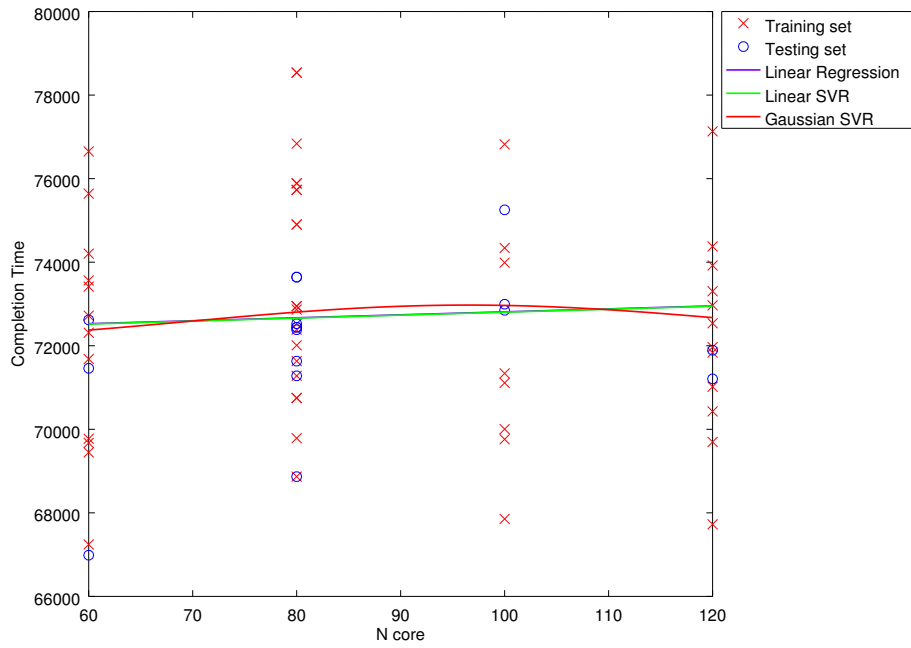
**Figure 5:** Completion time vs ncores for query R2 with datasize 250GB



### 1.2.2 Query R2 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1427	0.9539	278	0.0039
Linear SVR	0.1346	0.9662	279	0.0039
Polynomial SVR (2)	0.6942	0.2101	1256	0.0178
Polynomial SVR (3)	0.3123	0.8255	643	0.0089
Polynomial SVR (4)	0.6433	0.5642	1228	0.0174
Polynomial SVR (6)	0.7042	0.4659	1277	0.0181
Gaussian SVR	0.2514	0.8786	574	0.0079

**Table 6:** Results for R2-500

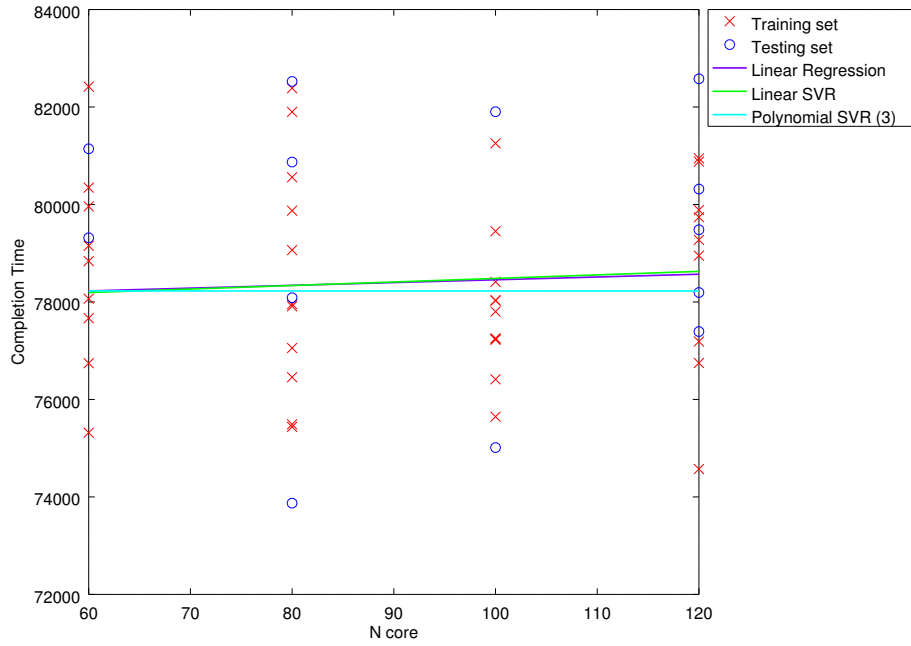


**Figure 6:** Completion time vs ncores for query R2 with datasize 500GB

### 1.2.3 Query R2 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2345	0.9596	469	0.0059
Linear SVR	0.2177	0.9748	445	0.0056
Polynomial SVR (2)	1.2193	0.0000	2315	0.0292
Polynomial SVR (3)	0.4706	0.8904	896	0.0114
Polynomial SVR (4)	1.2193	0.0000	2315	0.0292
Polynomial SVR (6)	1.2193	0.0000	2315	0.0292
Gaussian SVR	0.5978	0.8044	1099	0.0138

**Table 7:** Results for R2-750

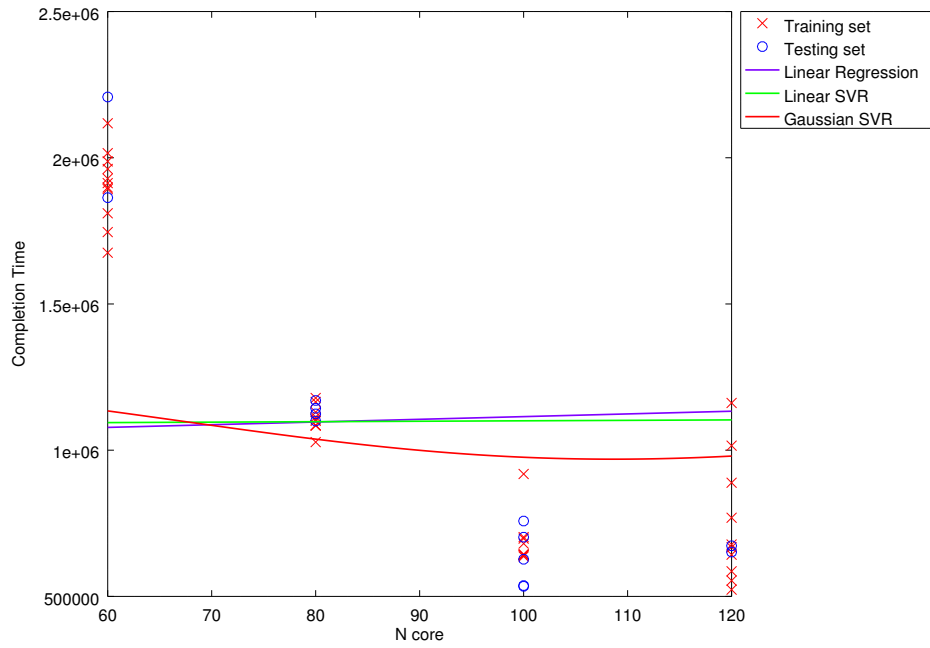


**Figure 7:** Completion time vs ncores for query R2 with datasize 750GB

### 1.2.4 Query R2 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0529	0.9971	21336	0.0227
Linear SVR	0.0872	0.9939	39397	0.0449
Polynomial SVR (2)	0.7767	0.4473	363910	0.4851
Polynomial SVR (3)	0.5370	0.8054	210031	0.2266
Polynomial SVR (4)	0.9436	0.5587	387439	0.4181
Polynomial SVR (6)	2.3977	0.5408	663955	0.6566
Gaussian SVR	0.2309	0.9596	94043	0.1177

**Table 8:** Results for R2-1000



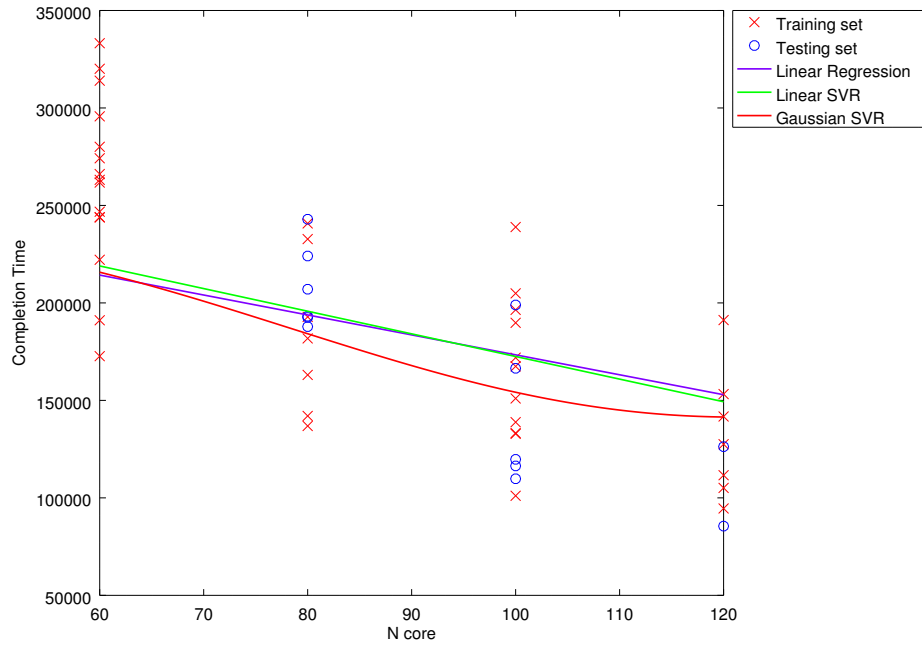
**Figure 8:** Completion time vs ncores for query R2 with datasize 1000GB

### 1.3 Query R3

#### 1.3.1 Query R3 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1109	0.9786	5118	0.0293
Linear SVR	0.1106	0.9889	5059	0.0292
Polynomial SVR (2)	0.7172	0.1678	38067	0.2733
Polynomial SVR (3)	0.2449	0.8985	13157	0.0911
Polynomial SVR (4)	0.6752	0.2430	35012	0.2465
Polynomial SVR (6)	0.6198	0.3670	35378	0.2375
Gaussian SVR	0.1868	0.9628	8716	0.0663

**Table 9:** Results for R3-250

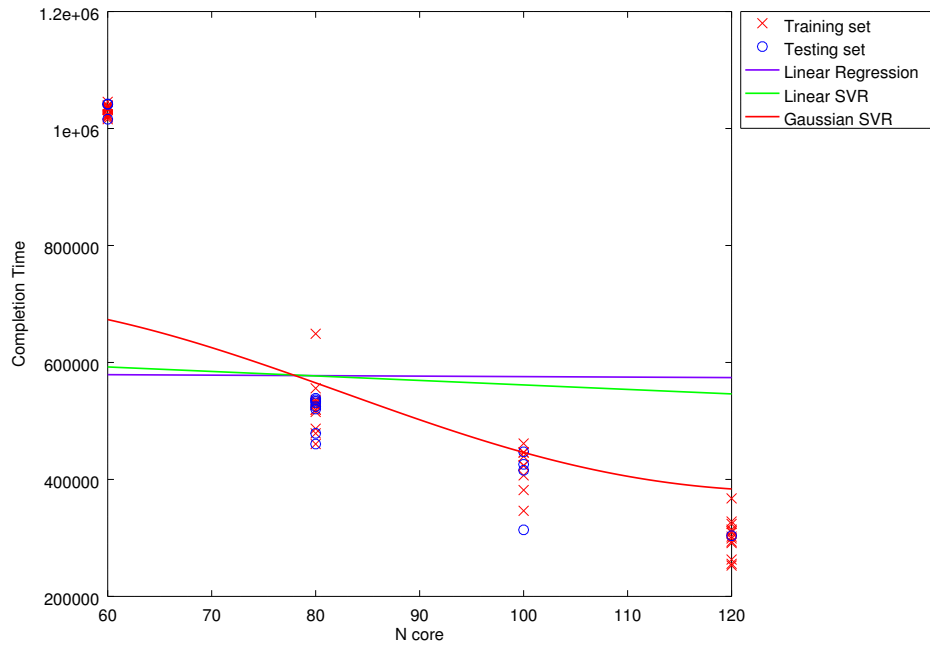


**Figure 9:** Completion time vs ncores for query R3 with datasize 250GB

### 1.3.2 Query R3 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0535	0.9963	11098	0.0224
Linear SVR	0.0606	0.9955	13859	0.0280
Polynomial SVR (2)	0.2930	0.9013	59492	0.1343
Polynomial SVR (3)	0.1548	0.9691	36584	0.0771
Polynomial SVR (4)	0.1646	0.9658	37111	0.0703
Polynomial SVR (6)	0.3303	0.8652	60515	0.1288
Gaussian SVR	0.0698	0.9937	14553	0.0288

**Table 10:** Results for R3-500

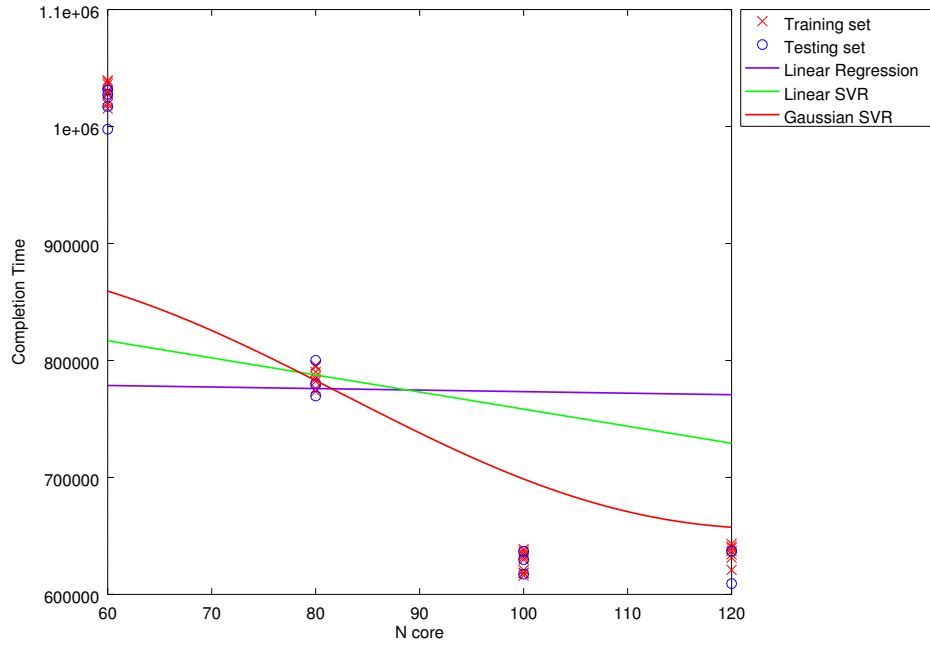


**Figure 10:** Completion time vs ncores for query R3 with datasize 500GB

### 1.3.3 Query R3 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0220	0.9996	2722	0.0034
Linear SVR	0.0888	0.9948	13574	0.0177
Polynomial SVR (2)	1.0647	0.4290	145696	0.1758
Polynomial SVR (3)	0.5284	0.8347	73215	0.0972
Polynomial SVR (4)	0.8089	0.6248	113146	0.1515
Polynomial SVR (6)	0.8551	0.4785	114570	0.1545
Gaussian SVR	0.2530	0.9796	34529	0.0434

**Table 11:** Results for R3-750

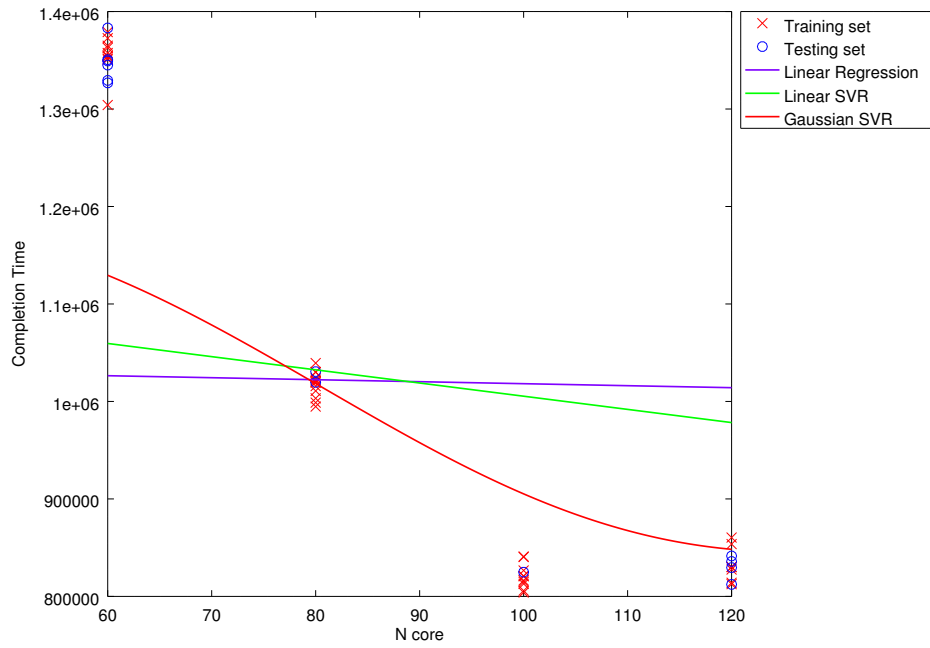


**Figure 11:** Completion time vs ncores for query R3 with datasize 750GB

### 1.3.4 Query R3 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0734	0.9955	9574	0.0085
Linear SVR	0.1067	0.9922	18970	0.0187
Polynomial SVR (2)	0.7430	0.6271	156805	0.1511
Polynomial SVR (3)	0.3859	0.9314	63411	0.0560
Polynomial SVR (4)	0.9817	0.2514	194521	0.1816
Polynomial SVR (6)	1.1743	0.0219	217951	0.2044
Gaussian SVR	0.3360	0.9471	54863	0.0472

**Table 12:** Results for R3-1000



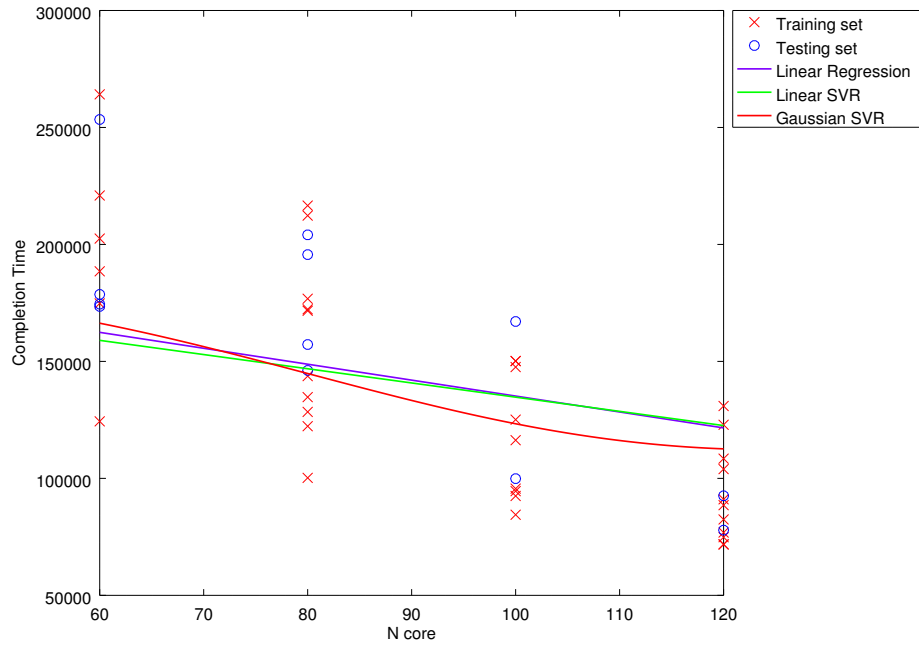
**Figure 12:** Completion time vs ncores for query R3 with datasize 1000GB

## 1.4 Query R4

### 1.4.1 Query R4 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1719	0.9691	5155	0.0313
Linear SVR	0.1665	0.9721	5476	0.0331
Polynomial SVR (2)	0.8901	0.2806	37089	0.2708
Polynomial SVR (3)	0.6349	0.8391	23943	0.1357
Polynomial SVR (4)	0.8922	0.4584	37940	0.2624
Polynomial SVR (6)	0.8339	0.5929	35512	0.2419
Gaussian SVR	0.2832	0.9483	9304	0.0558

**Table 13:** Results for R4-250



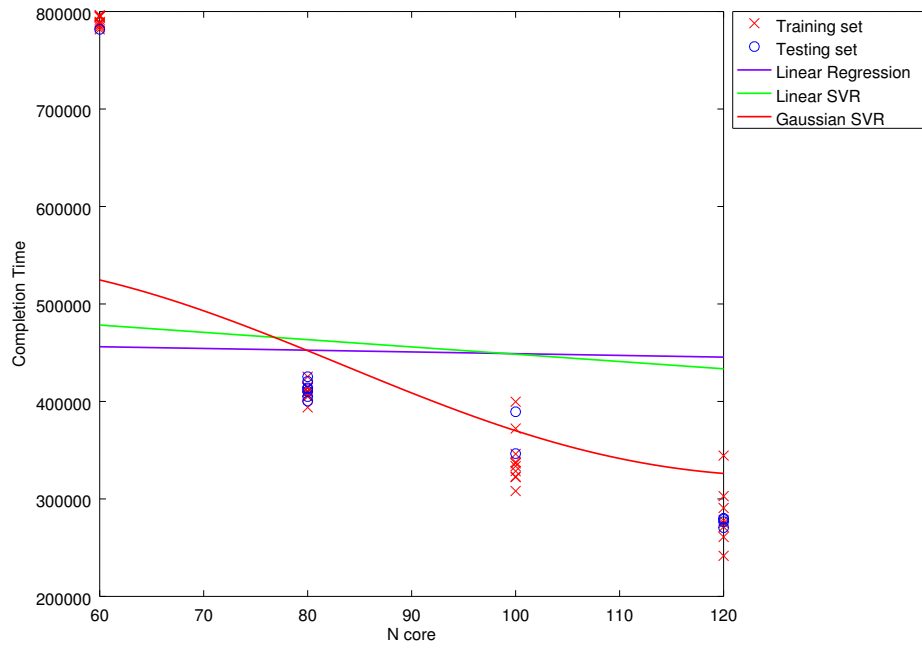
**Figure 13:** Completion time vs ncores for query R4 with datasize 250GB



#### 1.4.2 Query R4 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0523	0.9941	5530	0.0157
Linear SVR	0.0815	0.9947	13073	0.0334
Polynomial SVR (2)	0.3104	0.8825	40014	0.1194
Polynomial SVR (3)	0.1446	0.9635	19321	0.0472
Polynomial SVR (4)	0.2598	0.9353	35711	0.0904
Polynomial SVR (6)	0.2765	0.8948	36156	0.0950
Gaussian SVR	0.0816	0.9881	10981	0.0270

**Table 14:** Results for R4-500

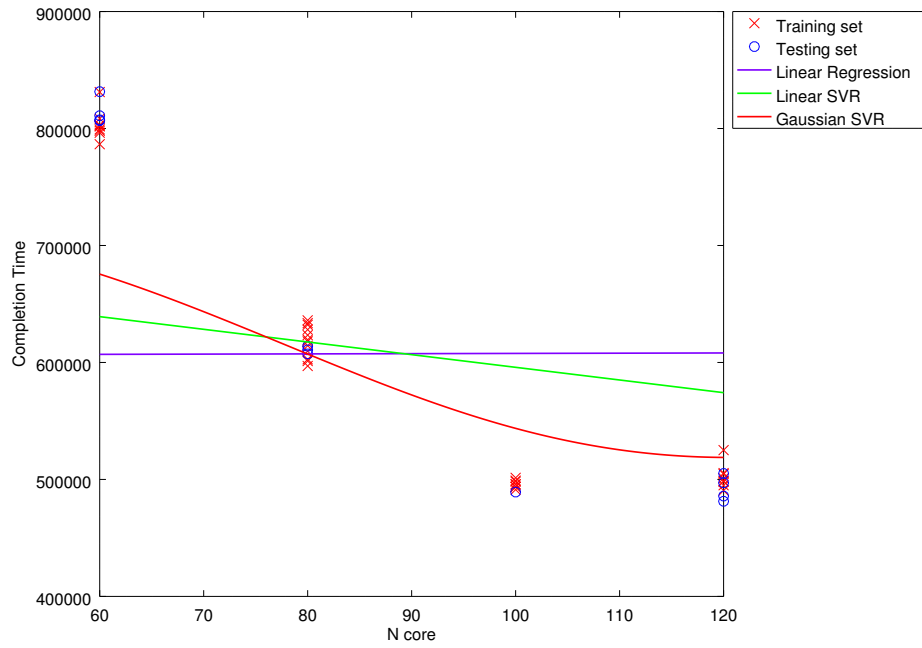


**Figure 14:** Completion time vs ncores for query R4 with datasize 500GB

### 1.4.3 Query R4 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0215	0.9996	2345	0.0038
Linear SVR	0.0808	0.9957	9071	0.0162
Polynomial SVR (2)	0.6955	0.7218	85402	0.1413
Polynomial SVR (3)	0.3624	0.9413	39380	0.0615
Polynomial SVR (4)	0.8872	0.5977	94544	0.1441
Polynomial SVR (6)	1.0199	0.3856	106583	0.1602
Gaussian SVR	0.1759	0.9783	14459	0.0248

**Table 15:** Results for R4-750

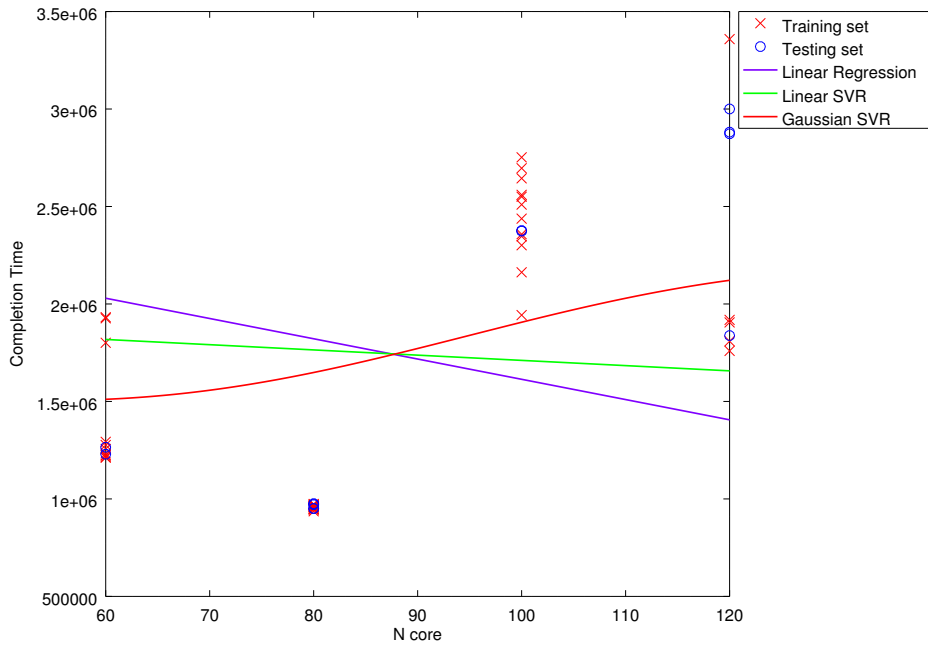


**Figure 15:** Completion time vs ncores for query R4 with datasize 750GB

#### 1.4.4 Query R4 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0944	0.9924	56636	0.0333
Linear SVR	0.1024	0.9936	60674	0.0388
Polynomial SVR (2)	0.8395	0.4422	580227	0.3698
Polynomial SVR (3)	0.4507	0.9060	296096	0.2340
Polynomial SVR (4)	0.8560	0.6428	607746	0.4239
Polynomial SVR (6)	0.9475	0.5411	673570	0.4938
Gaussian SVR	0.2280	0.9589	129214	0.0695

**Table 16:** Results for R4-1000



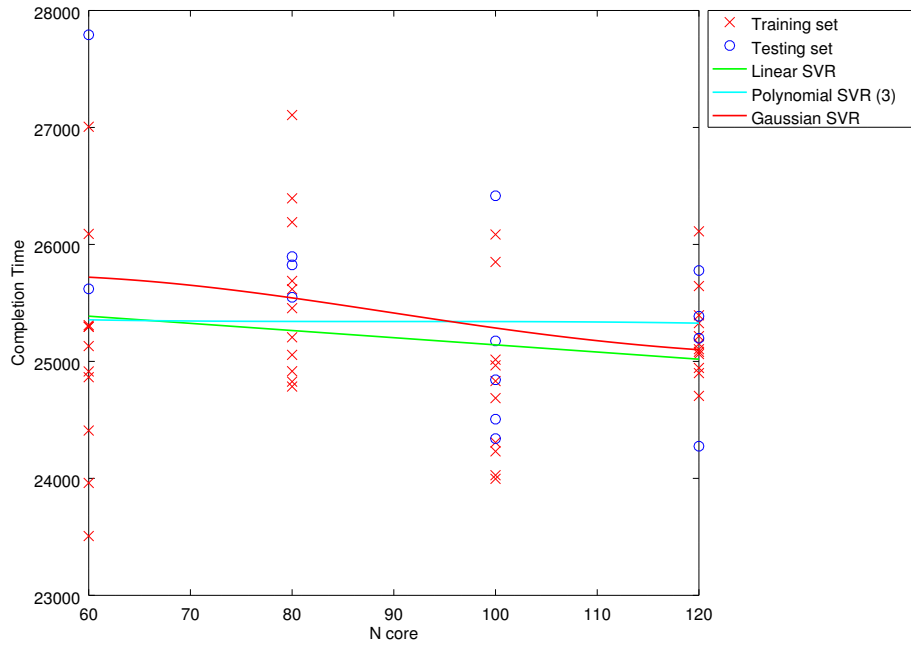
**Figure 16:** Completion time vs ncores for query R4 with datasize 1000GB

## 1.5 Query R5

### 1.5.1 Query R5 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	1.1021	-0.0551	713	0.0275
Linear SVR	1.0368	0.3190	639	0.0246
Polynomial SVR (2)	1.1427	0.0277	769	0.0299
Polynomial SVR (3)	1.0482	0.1294	688	0.0266
Polynomial SVR (4)	1.1212	0.0921	728	0.0282
Polynomial SVR (6)	1.1117	0.0337	696	0.0270
Gaussian SVR	0.9581	0.4034	595	0.0229

**Table 17:** Results for R5-250

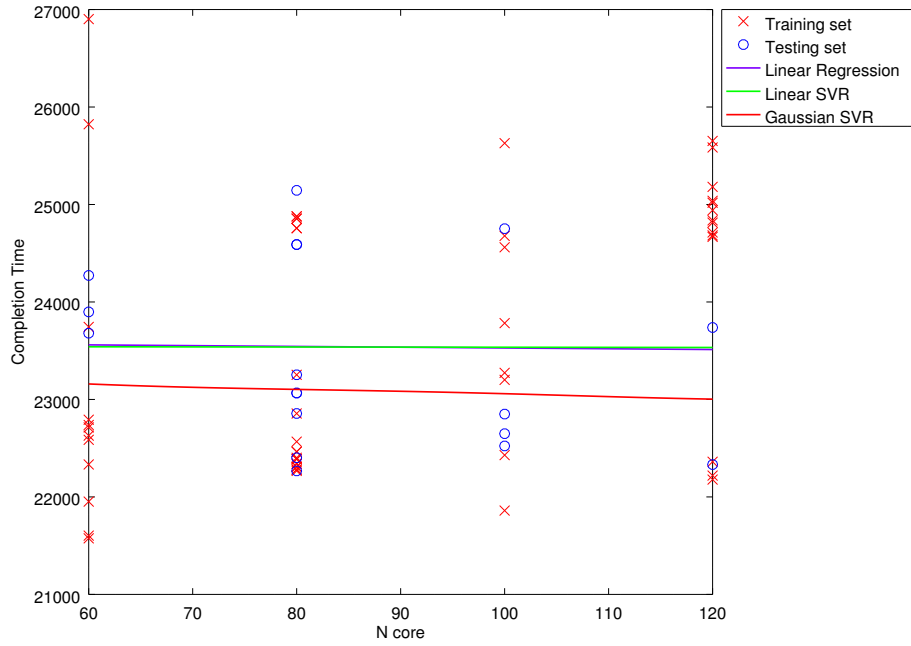


**Figure 17:** Completion time vs ncores for query R5 with datasize 250GB

### 1.5.2 Query R5 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.5027	0.5247	398	0.0169
Linear SVR	0.4869	0.6065	386	0.0163
Polynomial SVR (2)	0.7132	0.2590	753	0.0318
Polynomial SVR (3)	0.6277	0.4057	524	0.0221
Polynomial SVR (4)	0.8572	0.0574	770	0.0325
Polynomial SVR (6)	0.7629	0.0794	741	0.0314
Gaussian SVR	0.5268	0.5836	433	0.0182

**Table 18:** Results for R5-500

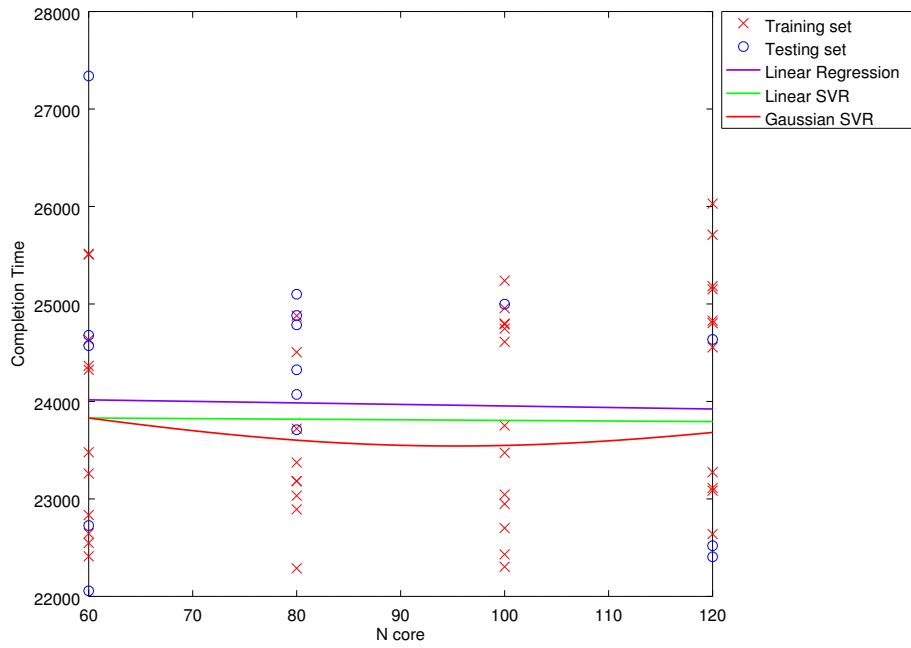


**Figure 18:** Completion time vs ncores for query R5 with datasize 500GB

### 1.5.3 Query R5 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.5755	0.6749	440	0.0176
Linear SVR	0.6121	0.7036	465	0.0186
Polynomial SVR (2)	1.4732	0.1597	1181	0.0486
Polynomial SVR (3)	0.8230	0.5766	841	0.0343
Polynomial SVR (4)	1.4430	0.0262	1312	0.0534
Polynomial SVR (6)	1.2012	0.0051	1189	0.0484
Gaussian SVR	0.6737	0.5912	521	0.0212

**Table 19:** Results for R5-750

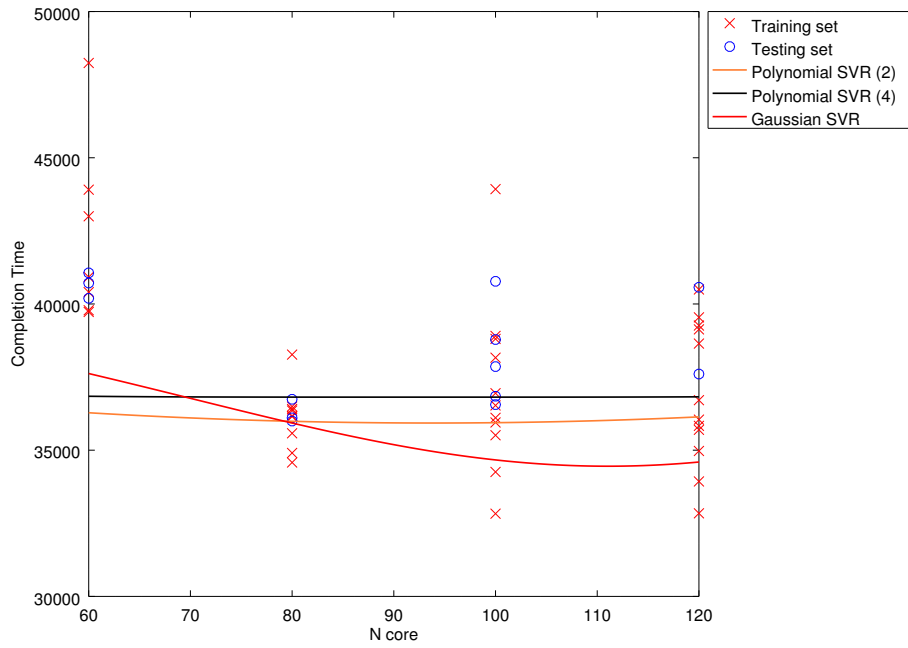


**Figure 19:** Completion time vs ncores for query R5 with datasize 750GB

### 1.5.4 Query R5 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.5015	0.1498	1322	0.0338
Linear SVR	0.4946	0.3382	1208	0.0309
Polynomial SVR (2)	0.4199	0.5664	947	0.0237
Polynomial SVR (3)	0.5480	0.2594	1525	0.0388
Polynomial SVR (4)	0.4707	0.3918	1138	0.0287
Polynomial SVR (6)	0.5688	0.3014	1456	0.0366
Gaussian SVR	0.3646	0.5791	820	0.0207

**Table 20:** Results for R5-1000



**Figure 20:** Completion time vs ncores for query R5 with datasize 1000GB

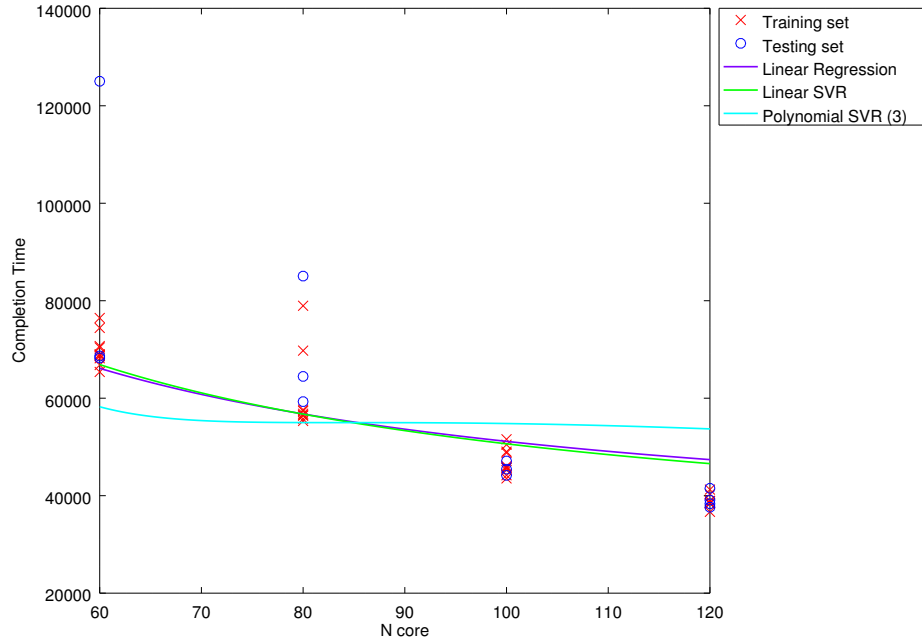
## 2 Fixed Datasize, $ncores^{-1}$ , all the features

### 2.1 Query R1

#### 2.1.1 Query R1 — Datasize 250GB

Model	RMSE	$R^2$	Mean absolute error	Mean relative error
Linear regression	0.3569	0.9437	2586	0.0332
Linear SVR	0.3897	0.9727	2836	0.0362
Polynomial SVR (2)	0.7267	0.7759	10787	0.2045
Polynomial SVR (3)	3.6861	0.7898	20915	0.2254
Polynomial SVR (4)	2.1868	0.2291	18693	0.2484
Polynomial SVR (6)	3.1348	0.3153	25360	0.3269
Gaussian SVR	1.2154	0.4402	7712	0.0841

**Table 21:** Results for R1-250GB with the nonlinear  $1/ncores$  feature



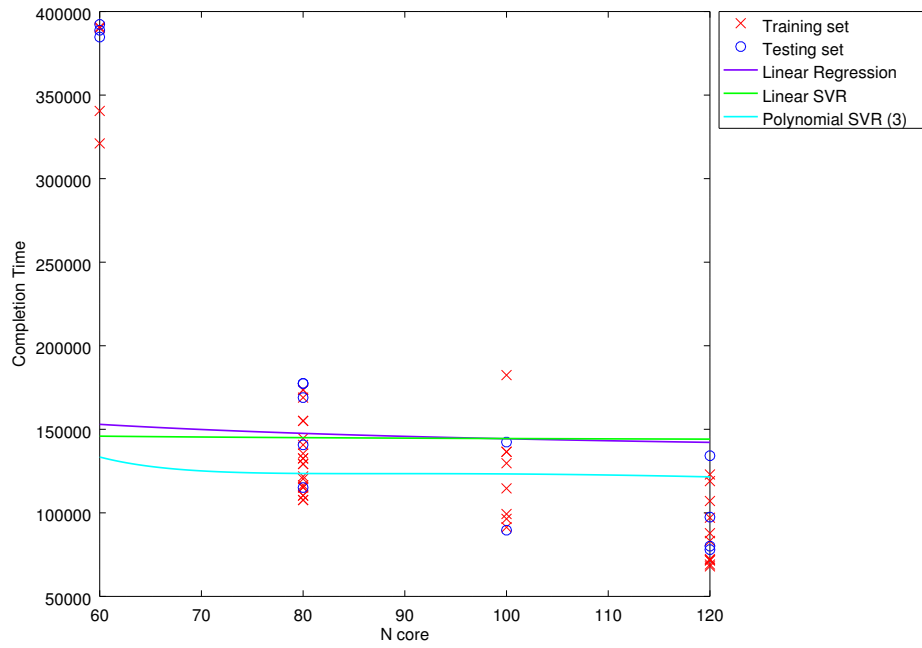
**Figure 21:** Completion time vs ncores for query R1 with datasize 250GB



### 2.1.2 Query R1 — Datasize 500GB

Model	RMSE	$R^2$	Mean absolute error	Mean relative error
Linear regression	0.0878	0.9957	5544	0.0283
Linear SVR	0.0994	0.9979	6748	0.0357
Polynomial SVR (2)	0.5341	0.9077	32780	0.1738
Polynomial SVR (3)	0.1464	0.9900	8855	0.0540
Polynomial SVR (4)	0.6090	0.8974	38612	0.2160
Polynomial SVR (6)	1.0016	0.8146	55698	0.2804
Gaussian SVR	0.3567	0.9755	18202	0.0825

**Table 22:** Results for R1-500GB with the nonlinear 1/ncores feature

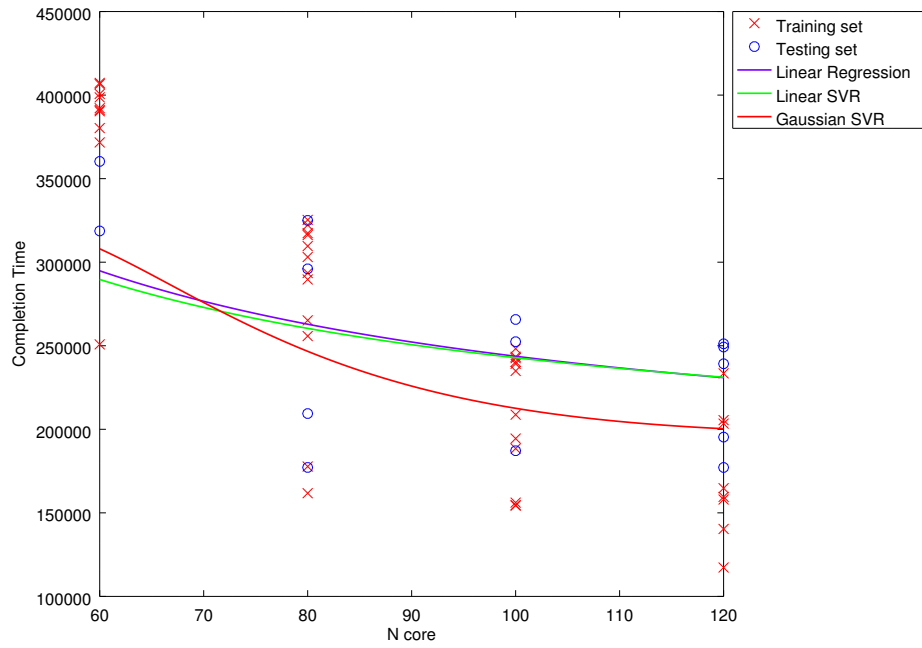


**Figure 22:** Completion time vs ncores for query R1 with datasize 500GB

### 2.1.3 Query R1 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1409	0.9526	10604	0.0437
Linear SVR	0.1435	0.9626	10499	0.0427
Polynomial SVR (2)	0.6469	0.2361	42774	0.1805
Polynomial SVR (3)	0.4315	0.5604	27541	0.1053
Polynomial SVR (4)	0.8832	0.0817	53947	0.2200
Polynomial SVR (6)	0.5945	0.2311	42316	0.1704
Gaussian SVR	0.1752	0.9606	11544	0.0496

**Table 23:** Results for R1-750GB with the nonlinear 1/ncores feature

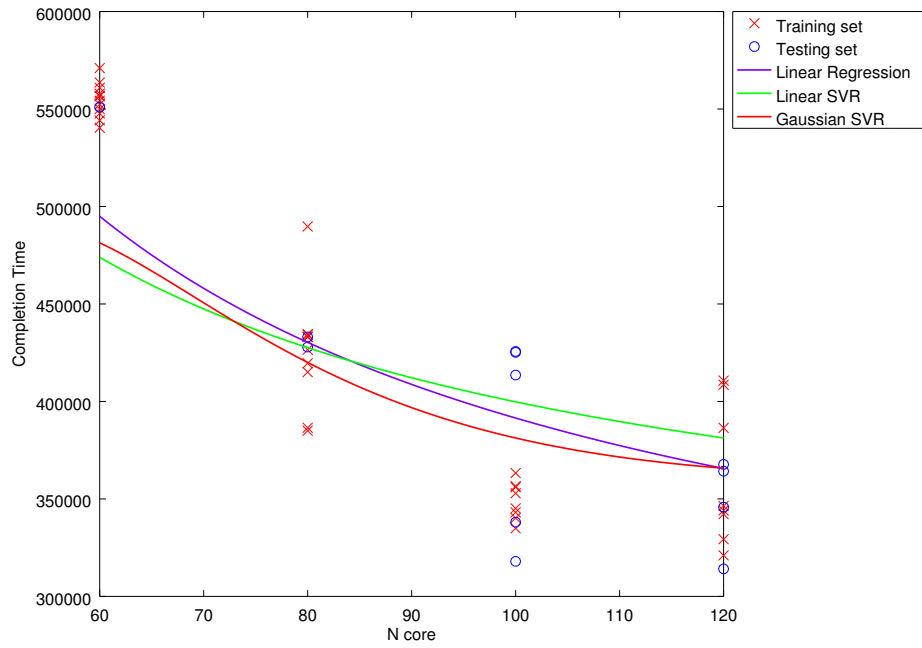


**Figure 23:** Completion time vs ncores for query R1 with datasize 750GB

### 2.1.4 Query R1 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1575	0.9544	10766	0.0283
Linear SVR	0.1449	0.9662	10234	0.0269
Polynomial SVR (2)	0.6723	0.3832	49056	0.1358
Polynomial SVR (3)	0.3719	0.7768	28996	0.0774
Polynomial SVR (4)	0.9733	0.0448	54529	0.1555
Polynomial SVR (6)	0.9608	0.0447	55540	0.1577
Gaussian SVR	0.3636	0.8120	15303	0.0446

**Table 24:** Results for R1-1000GB with the nonlinear 1/ncore feature



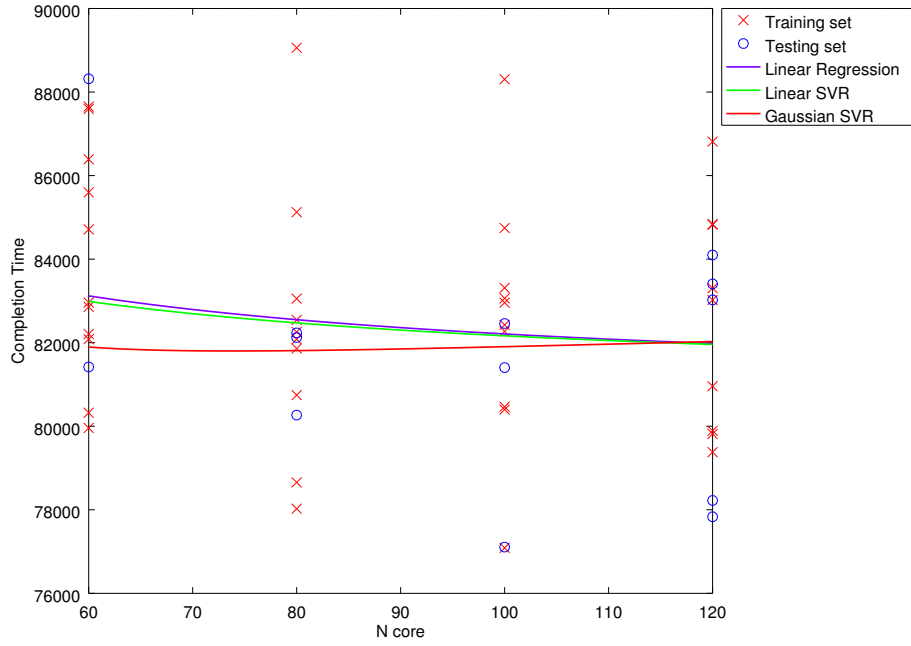
**Figure 24:** Completion time vs ncores for query R1 with datasize 1000GB

## 2.2 Query R2

### 2.2.1 Query R2 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.3116	0.8948	656	0.0079
Linear SVR	0.2890	0.9187	608	0.0073
Polynomial SVR (2)	1.0484	0.3489	2432	0.0302
Polynomial SVR (3)	0.5203	0.7803	1223	0.0153
Polynomial SVR (4)	0.9899	0.1627	2283	0.0284
Polynomial SVR (6)	1.0160	0.1256	2307	0.0287
Gaussian SVR	0.4203	0.8942	1005	0.0124

**Table 25:** Results for R2-250GB with the nonlinear 1/ncores feature

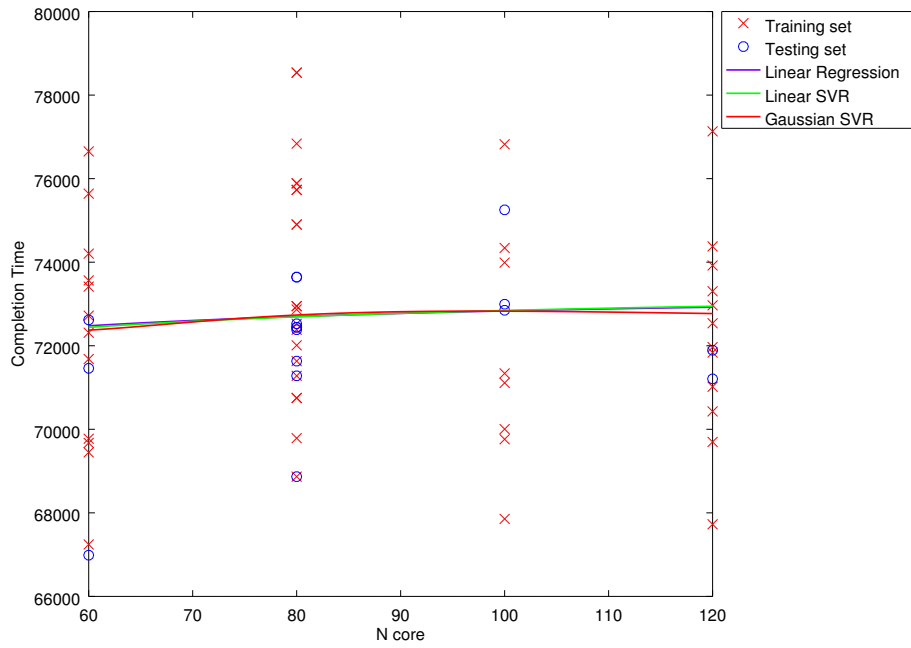


**Figure 25:** Completion time vs ncores for query R2 with datasize 250GB

### 2.2.2 Query R2 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1459	0.9517	284	0.0040
Linear SVR	0.1396	0.9640	293	0.0041
Polynomial SVR (2)	0.6872	0.3168	1249	0.0177
Polynomial SVR (3)	0.3714	0.7477	824	0.0116
Polynomial SVR (4)	0.6409	0.5494	1218	0.0172
Polynomial SVR (6)	0.5920	0.5919	1350	0.0190
Gaussian SVR	0.2339	0.8938	536	0.0074

**Table 26:** Results for R2-500GB with the nonlinear 1/ncore feature

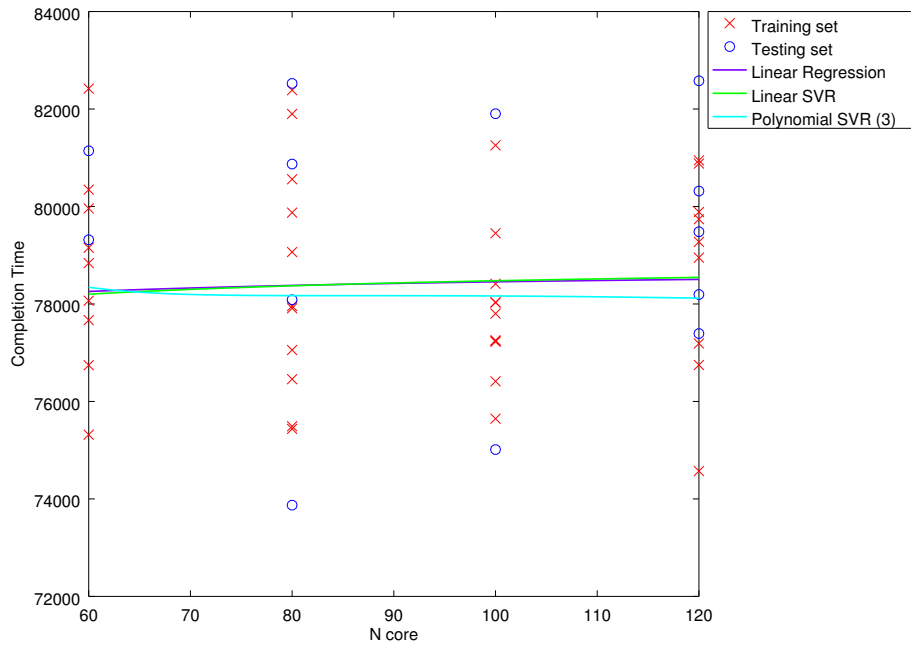


**Figure 26:** Completion time vs ncores for query R2 with datasize 500GB

### 2.2.3 Query R2 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2407	0.9575	472	0.0060
Linear SVR	0.2404	0.9726	476	0.0060
Polynomial SVR (2)	1.2193	0.0000	2315	0.0292
Polynomial SVR (3)	0.6488	0.8600	1046	0.0133
Polynomial SVR (4)	1.2193	0.0000	2315	0.0292
Polynomial SVR (6)	1.2193	0.0000	2315	0.0292
Gaussian SVR	0.5945	0.8089	1108	0.0139

**Table 27:** Results for R2-750GB with the nonlinear 1/ncore feature

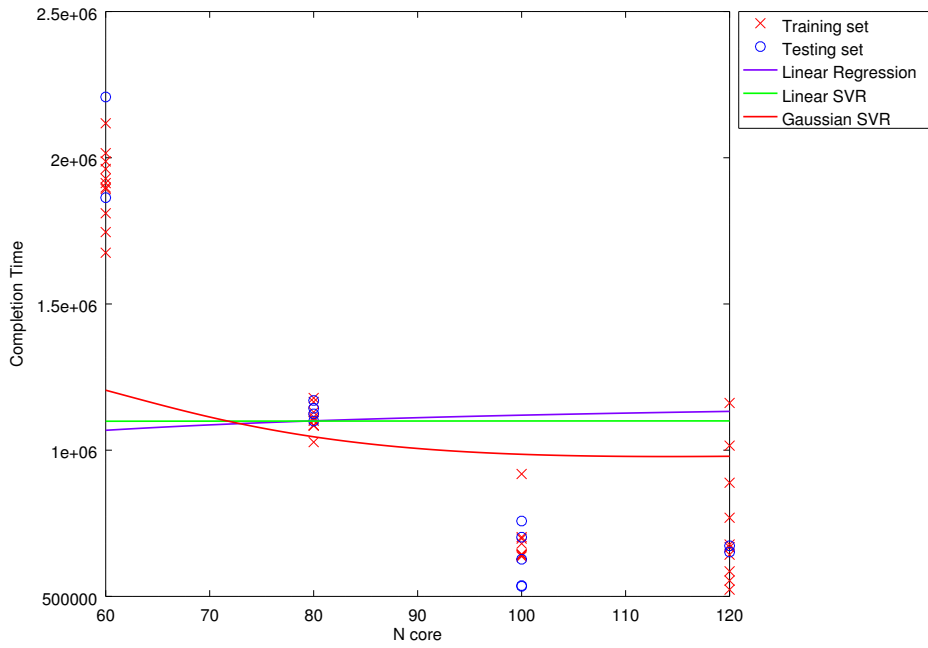


**Figure 27:** Completion time vs ncores for query R2 with datasize 750GB

### 2.2.4 Query R2 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0534	0.9971	21205	0.0221
Linear SVR	0.0843	0.9942	37707	0.0421
Polynomial SVR (2)	0.7009	0.5037	336756	0.4033
Polynomial SVR (3)	0.5103	0.8111	206774	0.2177
Polynomial SVR (4)	0.8812	0.5707	374286	0.4127
Polynomial SVR (6)	0.8290	0.5363	351110	0.4688
Gaussian SVR	0.2298	0.9590	93580	0.1148

**Table 28:** Results for R2-1000GB with the nonlinear 1/ncore feature



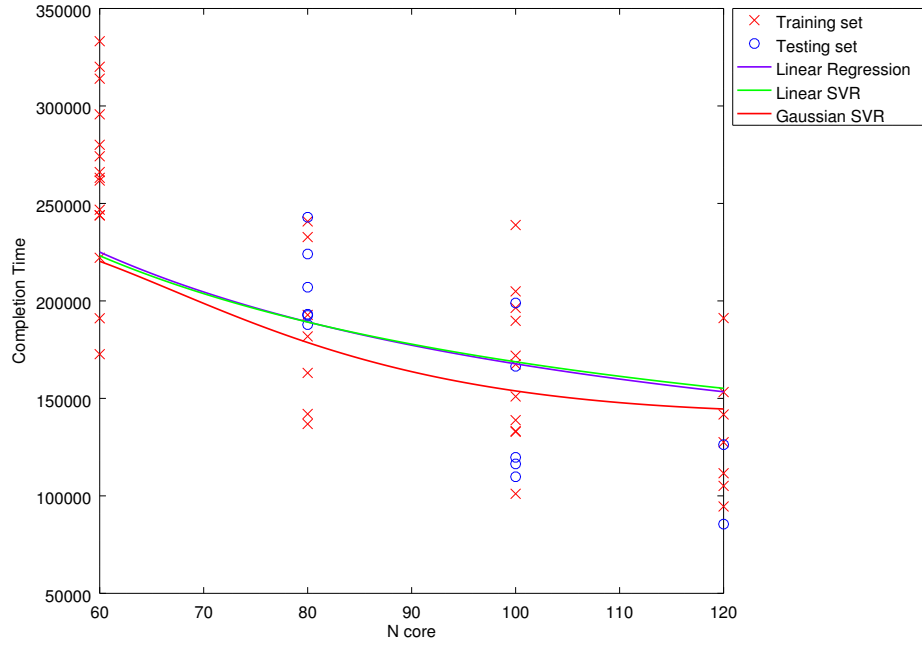
**Figure 28:** Completion time vs ncores for query R2 with datasize 1000GB

## 2.3 Query R3

### 2.3.1 Query R3 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0780	0.9894	3915	0.0252
Linear SVR	0.1057	0.9816	5093	0.0299
Polynomial SVR (2)	0.7201	0.1170	41915	0.2811
Polynomial SVR (3)	0.2708	0.9232	15247	0.0945
Polynomial SVR (4)	0.6645	0.2415	36141	0.2452
Polynomial SVR (6)	0.6164	0.4095	34311	0.2194
Gaussian SVR	0.1788	0.9665	8642	0.0656

**Table 29:** Results for R3-250GB with the nonlinear  $1/\text{ncores}$  feature



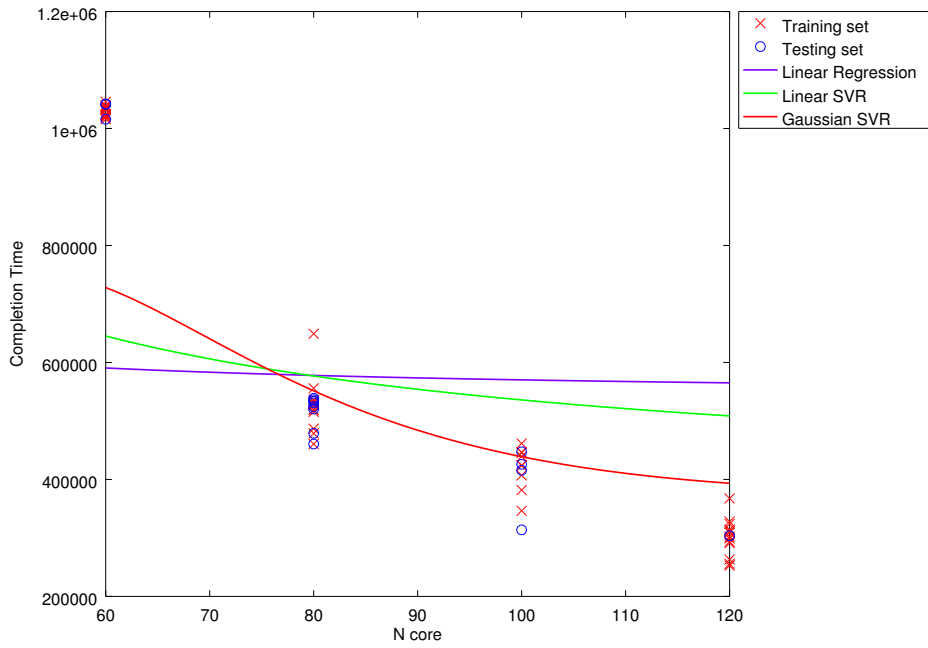
**Figure 29:** Completion time vs ncores for query R3 with datasize 250GB



### 2.3.2 Query R3 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0539	0.9962	10845	0.0220
Linear SVR	0.0626	0.9949	12632	0.0269
Polynomial SVR (2)	0.2821	0.8979	59067	0.1231
Polynomial SVR (3)	0.1297	0.9782	32168	0.0654
Polynomial SVR (4)	0.1922	0.9545	40999	0.0889
Polynomial SVR (6)	0.2877	0.8937	54259	0.1205
Gaussian SVR	0.0703	0.9939	15349	0.0321

**Table 30:** Results for R3-500GB with the nonlinear 1/ncores feature

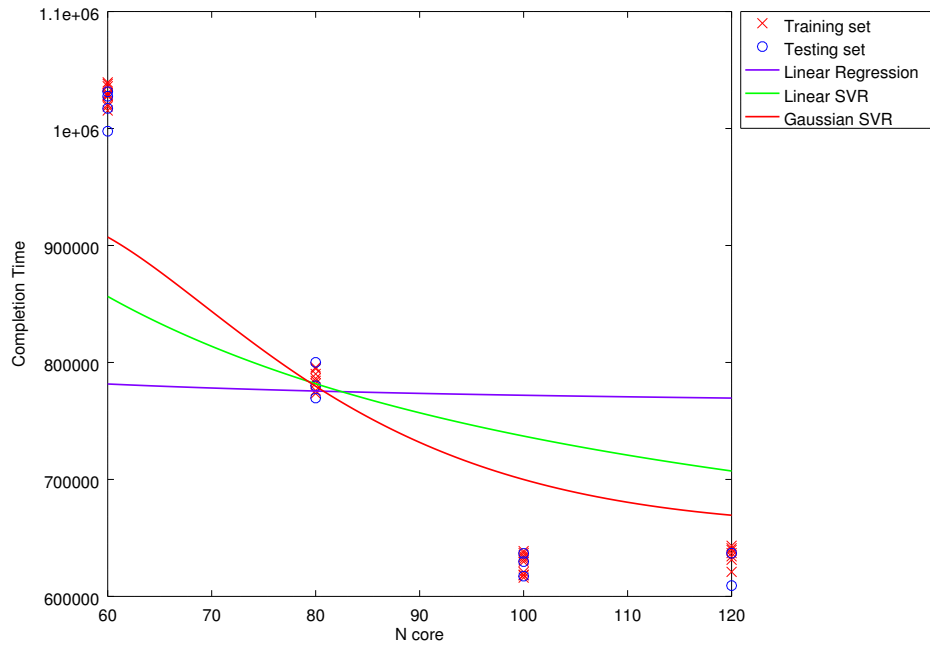


**Figure 30:** Completion time vs ncores for query R3 with datasize 500GB

### 2.3.3 Query R3 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0209	0.9996	2658	0.0033
Linear SVR	0.0806	0.9961	11228	0.0149
Polynomial SVR (2)	1.0051	0.5158	140581	0.1721
Polynomial SVR (3)	0.4684	0.9239	65851	0.0849
Polynomial SVR (4)	0.7927	0.6494	111418	0.1502
Polynomial SVR (6)	0.6810	0.6845	95464	0.1247
Gaussian SVR	0.2061	0.9777	26875	0.0359

**Table 31:** Results for R3-750GB with the nonlinear 1/ncores feature

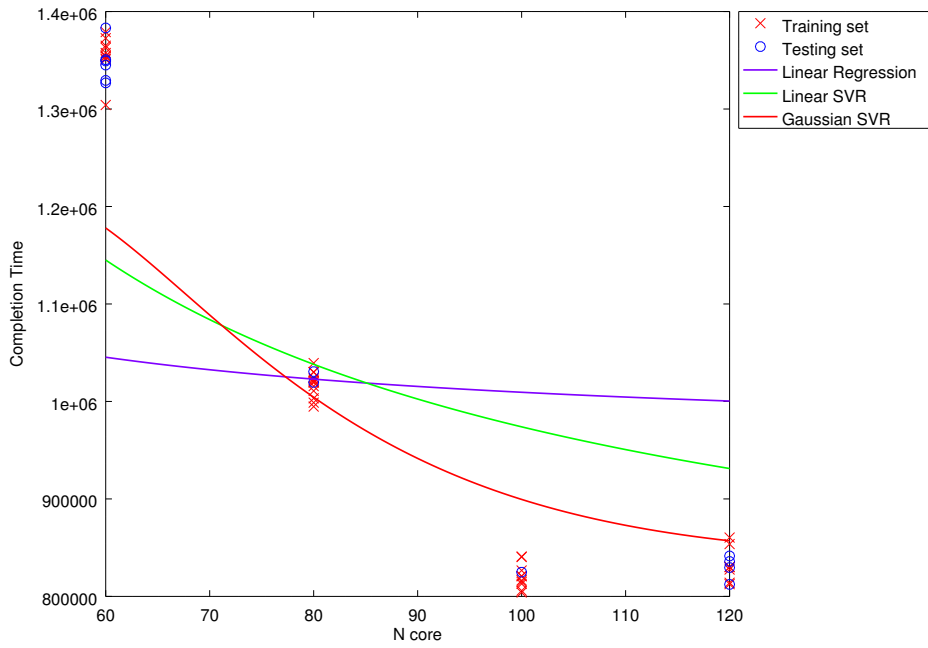


**Figure 31:** Completion time vs ncores for query R3 with datasize 750GB

### 2.3.4 Query R3 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0723	0.9956	10267	0.0094
Linear SVR	0.0959	0.9948	20283	0.0199
Polynomial SVR (2)	0.6954	0.6772	148312	0.1435
Polynomial SVR (3)	0.3570	0.9367	58398	0.0507
Polynomial SVR (4)	0.8566	0.5419	174113	0.1625
Polynomial SVR (6)	0.8709	0.4828	171350	0.1591
Gaussian SVR	0.2990	0.9557	49677	0.0438

**Table 32:** Results for R3-1000GB with the nonlinear 1/ncore feature



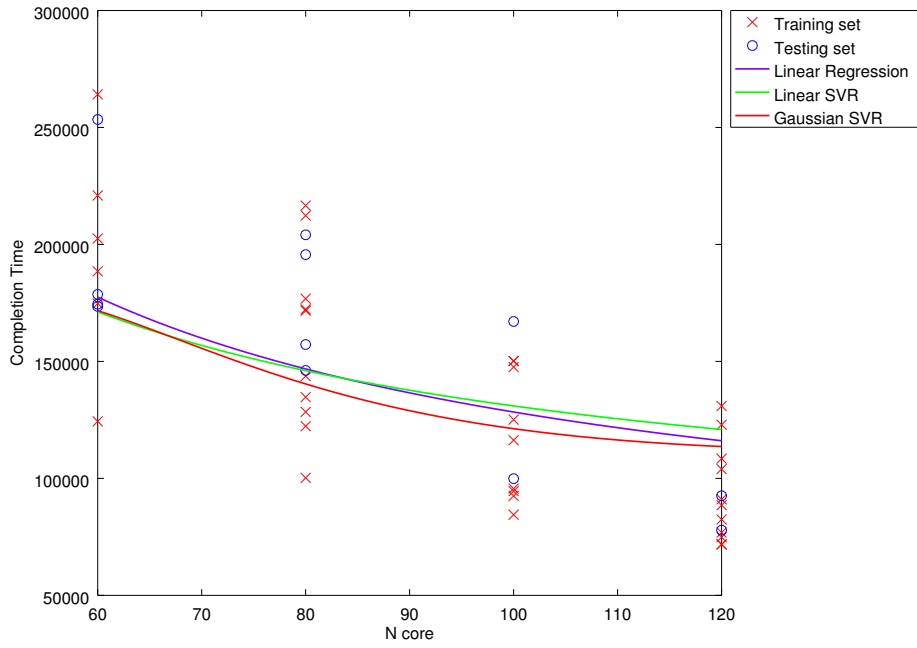
**Figure 32:** Completion time vs ncores for query R3 with datasize 1000GB

## 2.4 Query R4

### 2.4.1 Query R4 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1407	0.9793	4859	0.0282
Linear SVR	0.1048	0.9893	4109	0.0239
Polynomial SVR (2)	0.8143	0.4103	33794	0.2429
Polynomial SVR (3)	0.4882	0.8326	18105	0.1119
Polynomial SVR (4)	0.8747	0.5418	37483	0.2594
Polynomial SVR (6)	0.8999	0.5869	38677	0.2634
Gaussian SVR	0.2798	0.9470	9282	0.0572

**Table 33:** Results for R4-250GB with the nonlinear 1/ncore feature

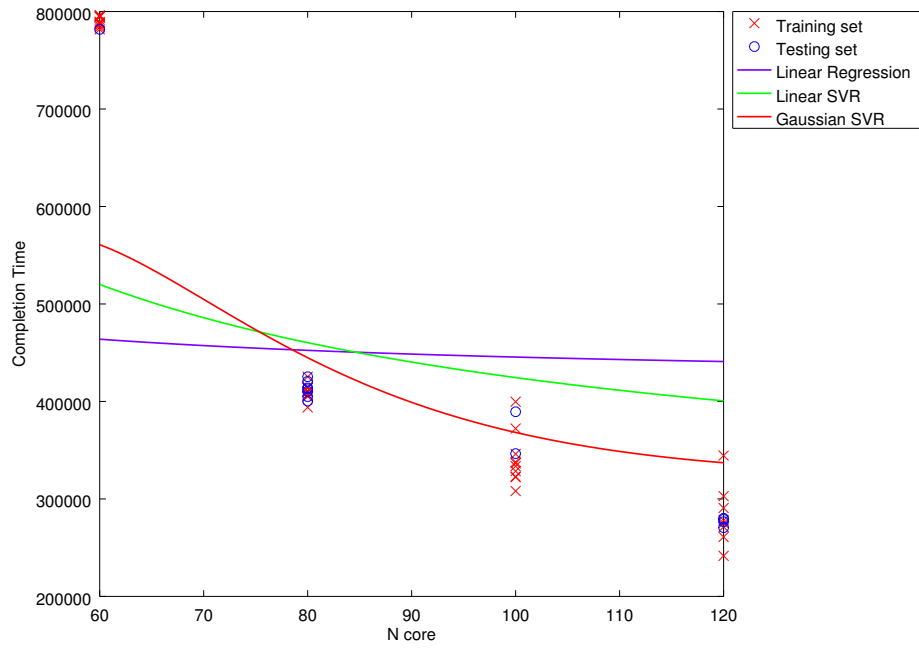


**Figure 33:** Completion time vs ncores for query R4 with datasize 250GB

### 2.4.2 Query R4 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0501	0.9946	5178	0.0145
Linear SVR	0.0850	0.9872	14922	0.0397
Polynomial SVR (2)	0.2128	0.9400	29712	0.0914
Polynomial SVR (3)	0.1287	0.9766	19858	0.0507
Polynomial SVR (4)	0.2801	0.9035	42883	0.1152
Polynomial SVR (6)	0.2924	0.8767	40499	0.1112
Gaussian SVR	0.0791	0.9876	11411	0.0287

**Table 34:** Results for R4-500GB with the nonlinear 1/ncores feature

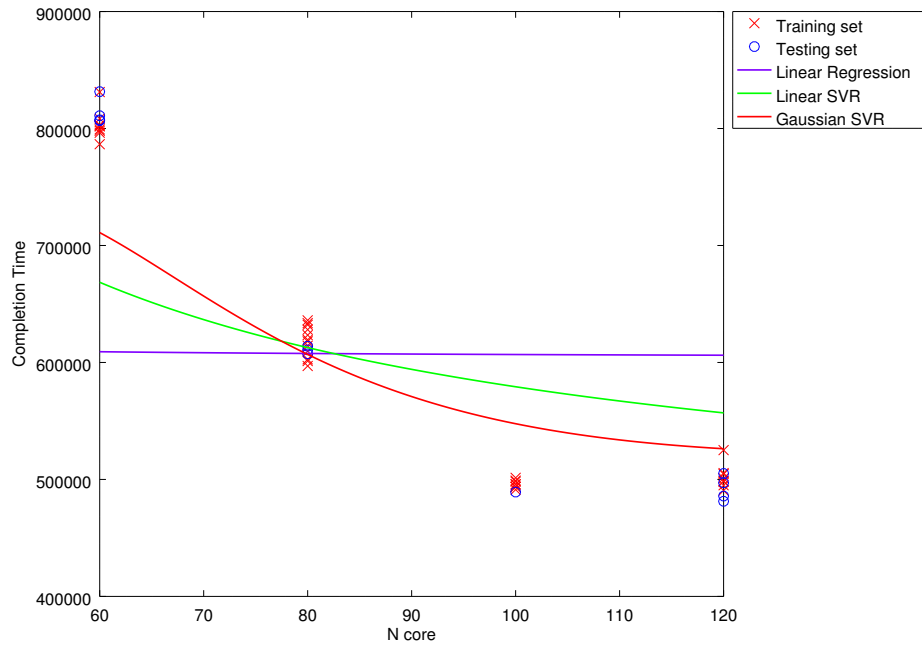


**Figure 34:** Completion time vs ncores for query R4 with datasize 500GB

### 2.4.3 Query R4 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0216	0.9996	2333	0.0038
Linear SVR	0.0773	0.9968	8179	0.0146
Polynomial SVR (2)	0.6766	0.7235	82143	0.1361
Polynomial SVR (3)	0.3497	0.9401	37764	0.0591
Polynomial SVR (4)	0.7932	0.7572	91479	0.1430
Polynomial SVR (6)	0.9234	0.6666	101239	0.1550
Gaussian SVR	0.1553	0.9830	13013	0.0221

**Table 35:** Results for R4-750GB with the nonlinear 1/ncores feature

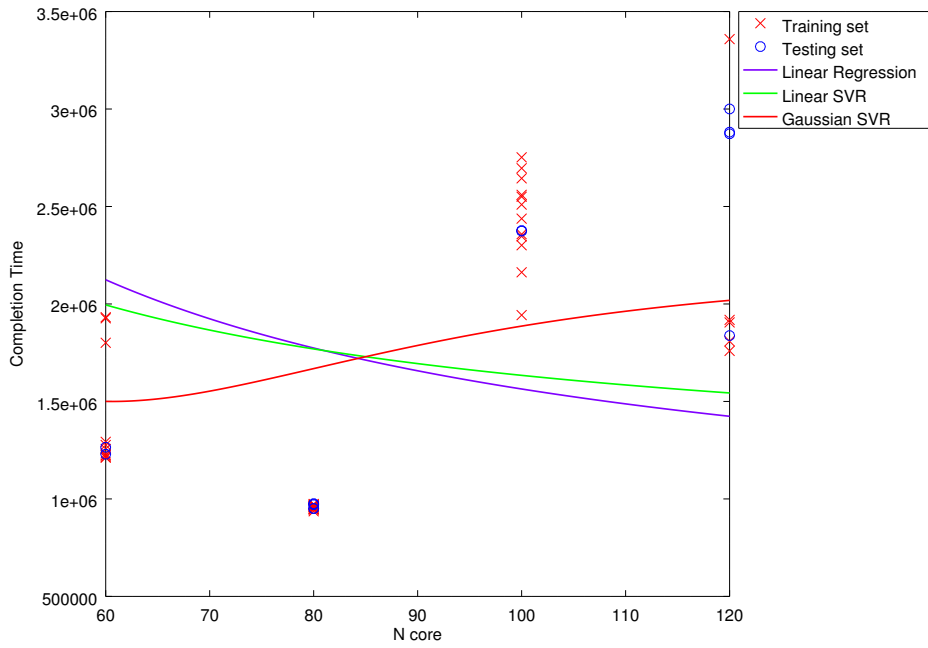


**Figure 35:** Completion time vs ncores for query R4 with datasize 750GB

#### 2.4.4 Query R4 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0812	0.9944	45917	0.0244
Linear SVR	0.0785	0.9966	46627	0.0315
Polynomial SVR (2)	0.7046	0.6535	446048	0.2688
Polynomial SVR (3)	0.3197	0.9286	196869	0.1466
Polynomial SVR (4)	0.8131	0.7058	578071	0.4145
Polynomial SVR (6)	0.9574	0.5389	680038	0.4961
Gaussian SVR	0.2282	0.9582	129449	0.0703

**Table 36:** Results for R4-1000GB with the nonlinear 1/ncore feature



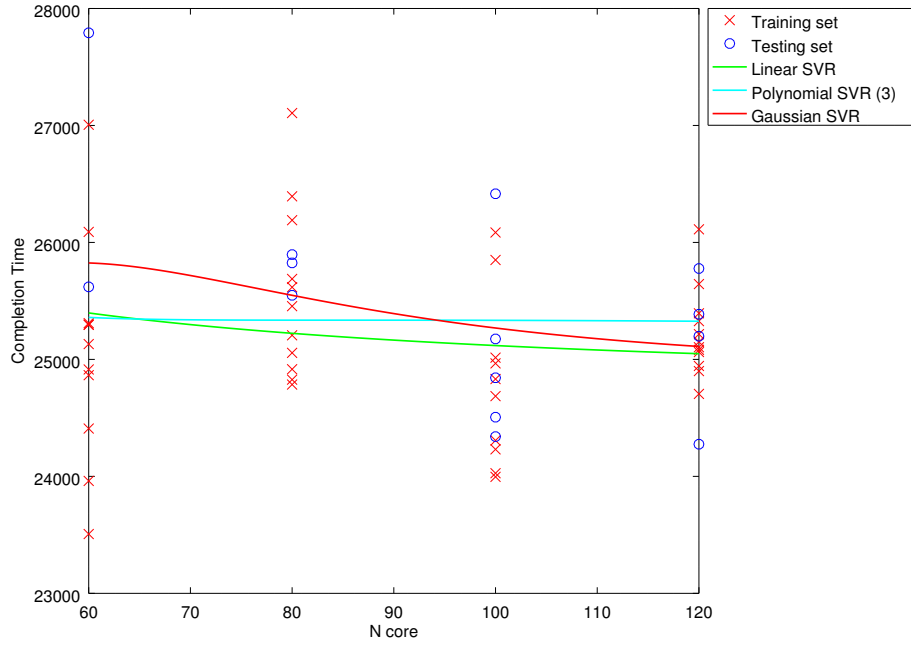
**Figure 36:** Completion time vs ncores for query R4 with datasize 1000GB

## 2.5 Query R5

### 2.5.1 Query R5 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	1.1046	-0.0599	720	0.0277
Linear SVR	1.0254	0.3409	642	0.0247
Polynomial SVR (2)	1.1448	0.0293	773	0.0301
Polynomial SVR (3)	1.0585	0.1048	706	0.0273
Polynomial SVR (4)	1.0900	0.0650	729	0.0283
Polynomial SVR (6)	1.0613	0.0496	679	0.0264
Gaussian SVR	0.9708	0.3643	606	0.0234

**Table 37:** Results for R5-250GB with the nonlinear  $1/\text{ncores}$  feature



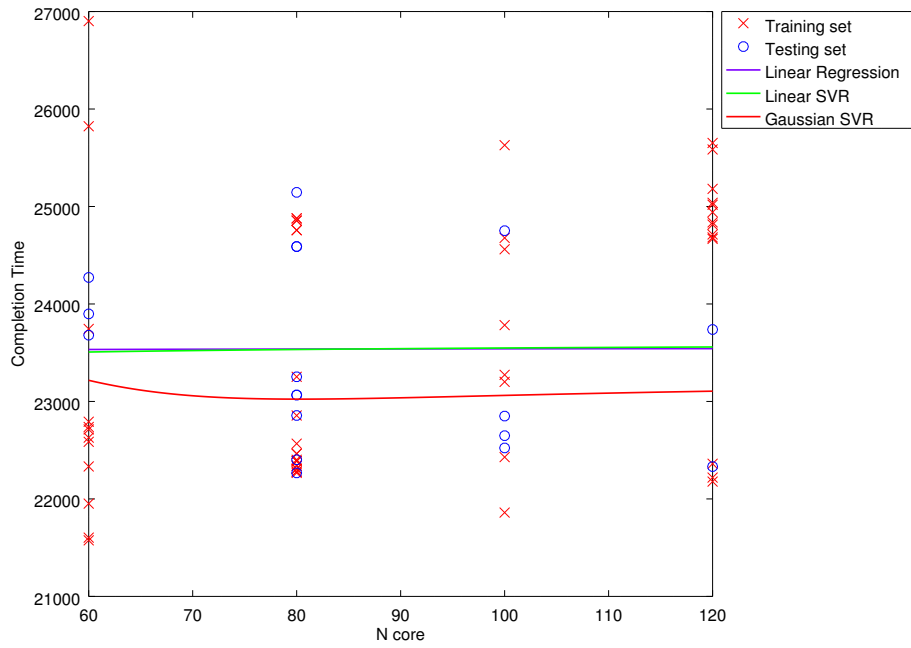
**Figure 37:** Completion time vs ncores for query R5 with datasize 250GB



### 2.5.2 Query R5 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.4987	0.5323	389	0.0164
Linear SVR	0.4932	0.5986	381	0.0161
Polynomial SVR (2)	0.7350	0.2975	705	0.0297
Polynomial SVR (3)	0.6196	0.4052	521	0.0220
Polynomial SVR (4)	0.7239	0.2502	636	0.0265
Polynomial SVR (6)	0.6725	0.1906	683	0.0289
Gaussian SVR	0.5501	0.5643	416	0.0175

**Table 38:** Results for R5-500GB with the nonlinear 1/ncore feature

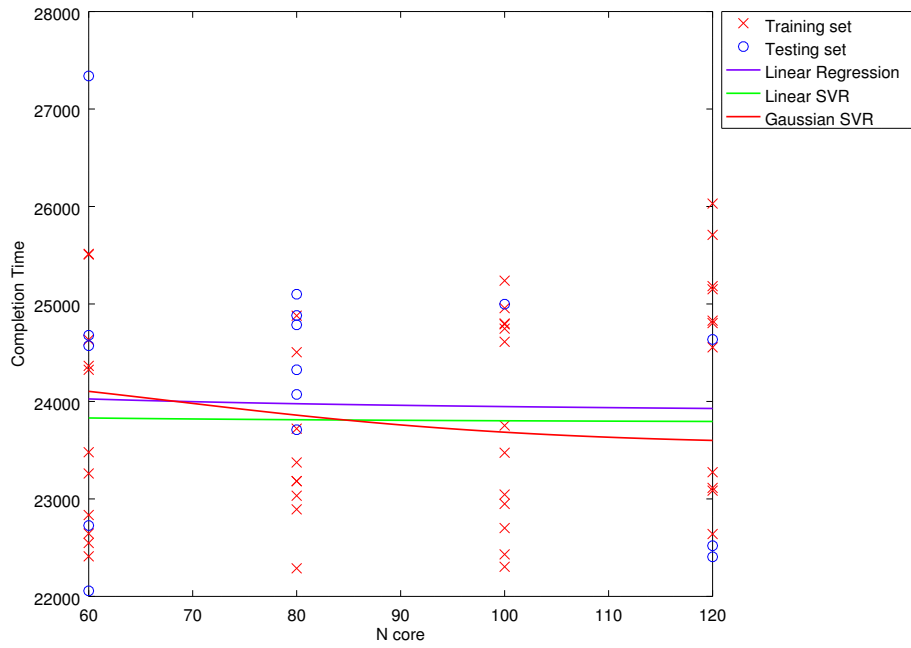


**Figure 38:** Completion time vs ncores for query R5 with datasize 500GB

### 2.5.3 Query R5 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.5758	0.6745	442	0.0177
Linear SVR	0.6121	0.7041	466	0.0186
Polynomial SVR (2)	2.5714	0.0675	1530	0.0618
Polynomial SVR (3)	0.8141	0.5165	830	0.0340
Polynomial SVR (4)	1.9622	0.0219	1418	0.0570
Polynomial SVR (6)	2.1235	0.0003	1574	0.0630
Gaussian SVR	0.6321	0.6870	486	0.0195

**Table 39:** Results for R5-750GB with the nonlinear 1/ncores feature

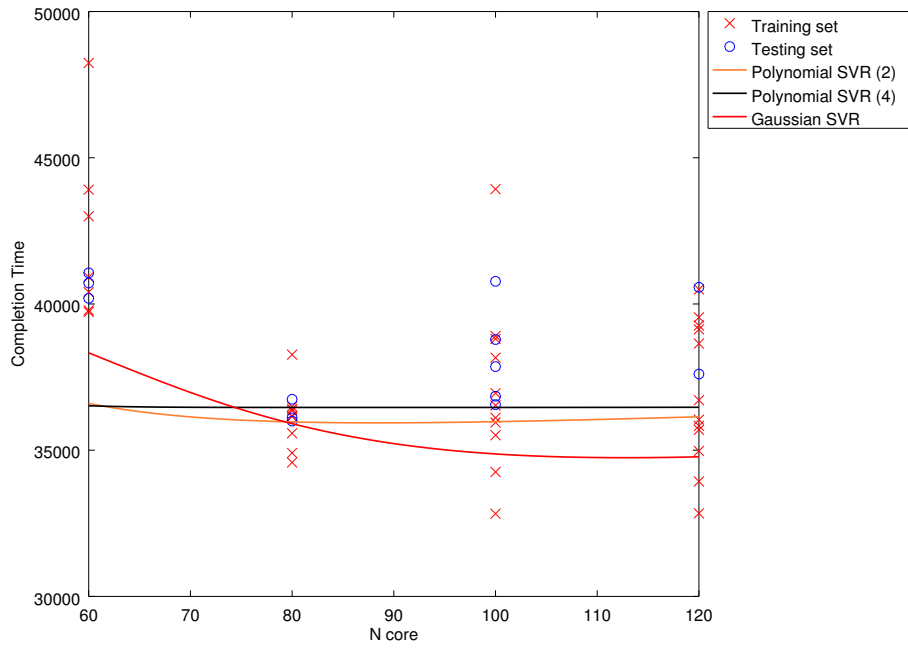


**Figure 39:** Completion time vs ncores for query R5 with datasize 750GB

### 2.5.4 Query R5 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.5047	0.1390	1333	0.0341
Linear SVR	0.4982	0.3487	1237	0.0316
Polynomial SVR (2)	0.4124	0.5666	873	0.0219
Polynomial SVR (3)	0.6250	0.2996	1597	0.0404
Polynomial SVR (4)	0.5453	0.3911	1321	0.0332
Polynomial SVR (6)	0.5542	0.2966	1421	0.0358
Gaussian SVR	0.3568	0.5856	851	0.0217

**Table 40:** Results for R5-1000GB with the nonlinear 1/ncore feature



**Figure 40:** Completion time vs ncores for query R5 with datasize 1000GB

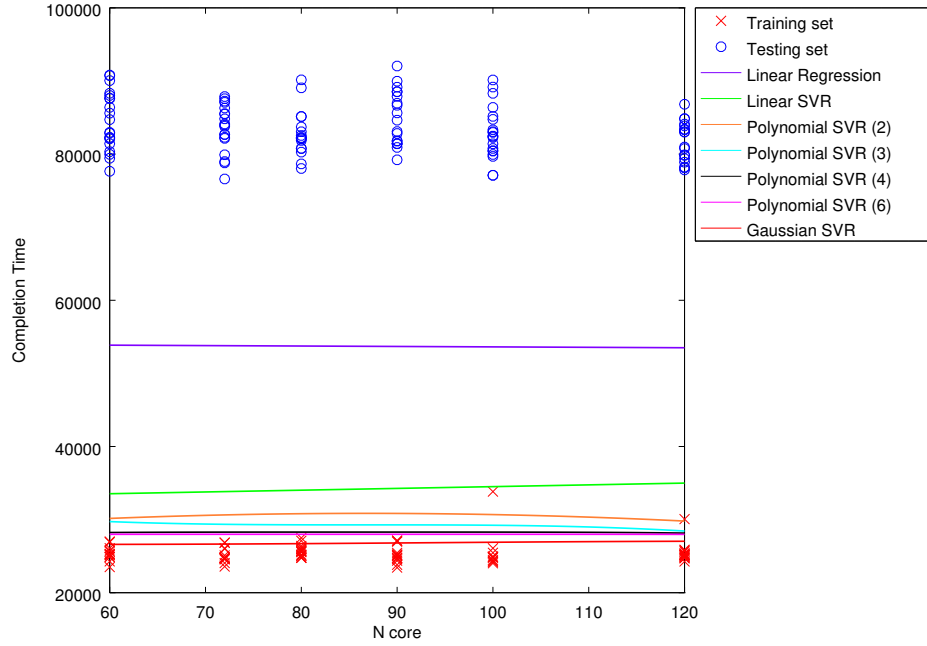
### 3 Query comparison

#### 3.1 R5 $\rightarrow$ R2 (Fast - Fast)

##### 3.1.1 R5 $\rightarrow$ R2 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2664	-3.8012	7131	0.0868
Linear SVR	1.4192	0.4500	41150	0.4928
Polynomial SVR (2)	2.0017	0.3318	57963	0.6936
Polynomial SVR (3)	1.7351	0.3201	50336	0.6032
Polynomial SVR (4)	2.0052	0.2591	58051	0.6946
Polynomial SVR (6)	2.0177	0.3148	58314	0.6974
Gaussian SVR	1.9531	0.0953	56646	0.6783

**Table 41:** Results for R5  $\rightarrow$  R2-250

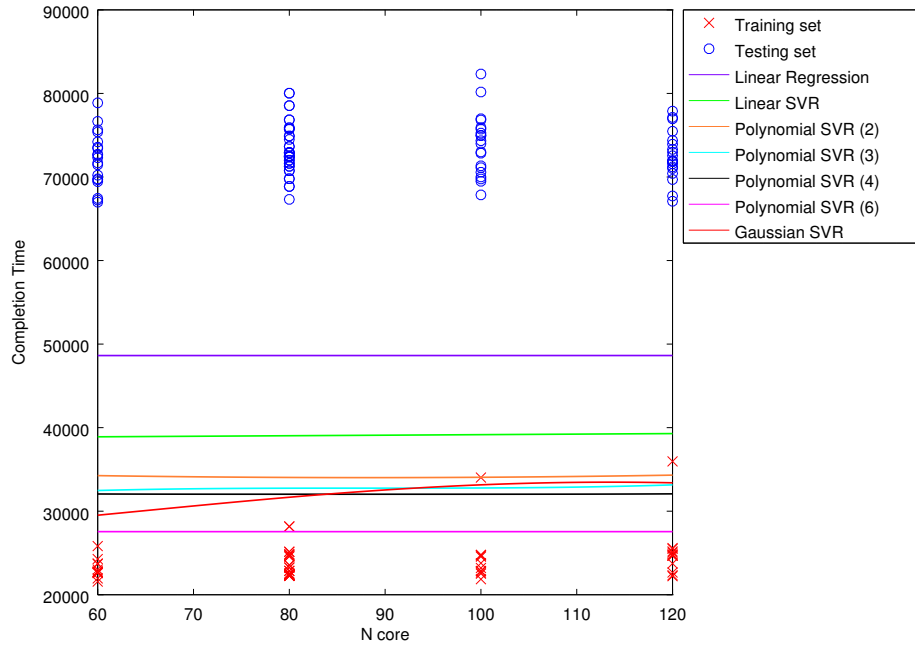


**Figure 41:** Completion time vs ncores for: R5  $\rightarrow$  R2 with datasize 250GB

### 3.1.2 R5 $\rightarrow$ R2 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0927	0.4825	1627	0.0219
Linear SVR	0.8177	0.8804	20156	0.2760
Polynomial SVR (2)	1.9565	0.8888	47958	0.6544
Polynomial SVR (3)	1.3110	0.8506	32346	0.4437
Polynomial SVR (4)	1.9813	0.8130	48399	0.6598
Polynomial SVR (6)	1.9723	0.7576	48394	0.6606
Gaussian SVR	1.8458	0.4395	45448	0.6213

**Table 42:** Results for R5  $\rightarrow$  R2-500

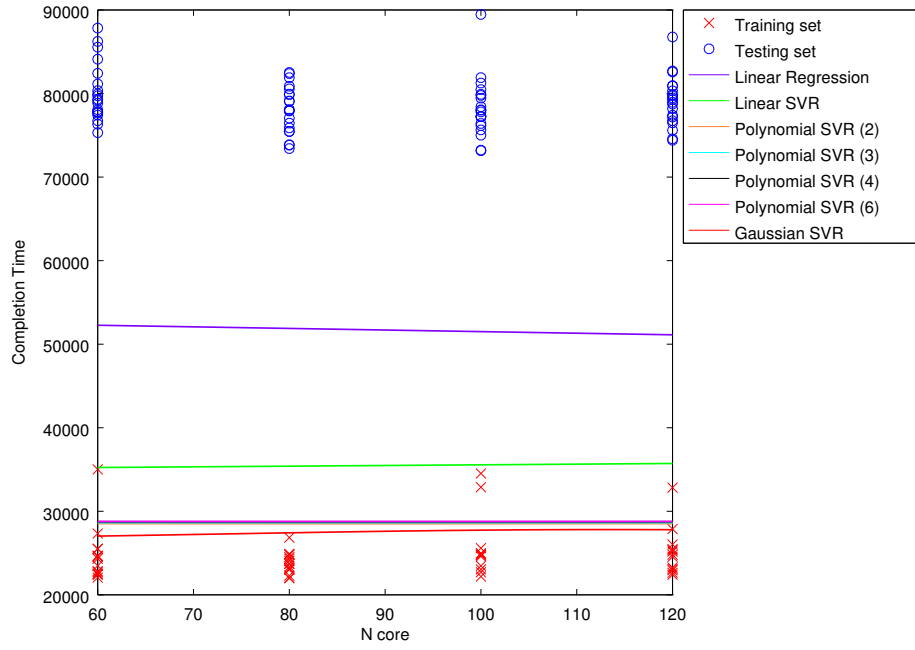


**Figure 42:** Completion time vs ncores for: R5  $\rightarrow$  R2 with datasize 500GB

### 3.1.3 R5 $\rightarrow$ R2 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.7284	-	19886	0.2523
Linear SVR	1.2200	0.7629	33371	0.4226
Polynomial SVR (2)	1.9610	0.7081	53638	0.6792
Polynomial SVR (3)	1.7155	0.7155	46976	0.5953
Polynomial SVR (4)	1.9544	0.6962	53405	0.6760
Polynomial SVR (6)	1.9545	0.6360	53347	0.6749
Gaussian SVR	1.8982	0.4132	51944	0.6579

**Table 43:** Results for R5  $\rightarrow$  R2-750

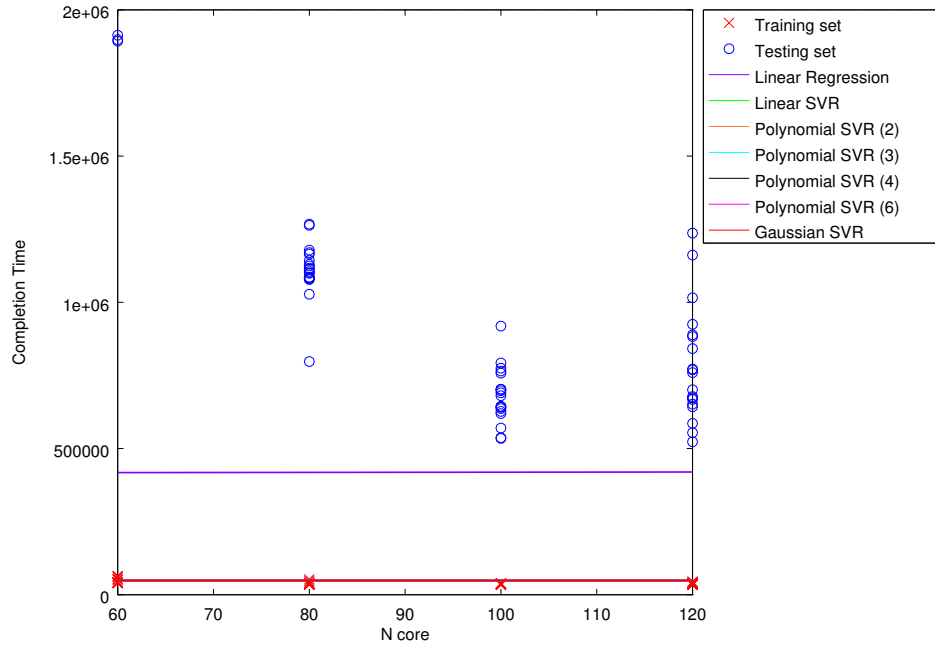


**Figure 43:** Completion time vs ncores for: R5  $\rightarrow$  R2 with datasize 750GB

### 3.1.4 R5 → R2 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2791	0.8182	124323	0.1468
Linear SVR	1.9131	-0.0000	865805	0.9413
Polynomial SVR (2)	1.9131	-0.0000	865805	0.9413
Polynomial SVR (3)	1.9131	-0.0000	865805	0.9413
Polynomial SVR (4)	1.9131	-0.0000	865805	0.9413
Polynomial SVR (6)	1.9131	-0.0000	865805	0.9413
Gaussian SVR	1.9131	-0.0000	865805	0.9413

**Table 44:** Results for R5 → R2-1000



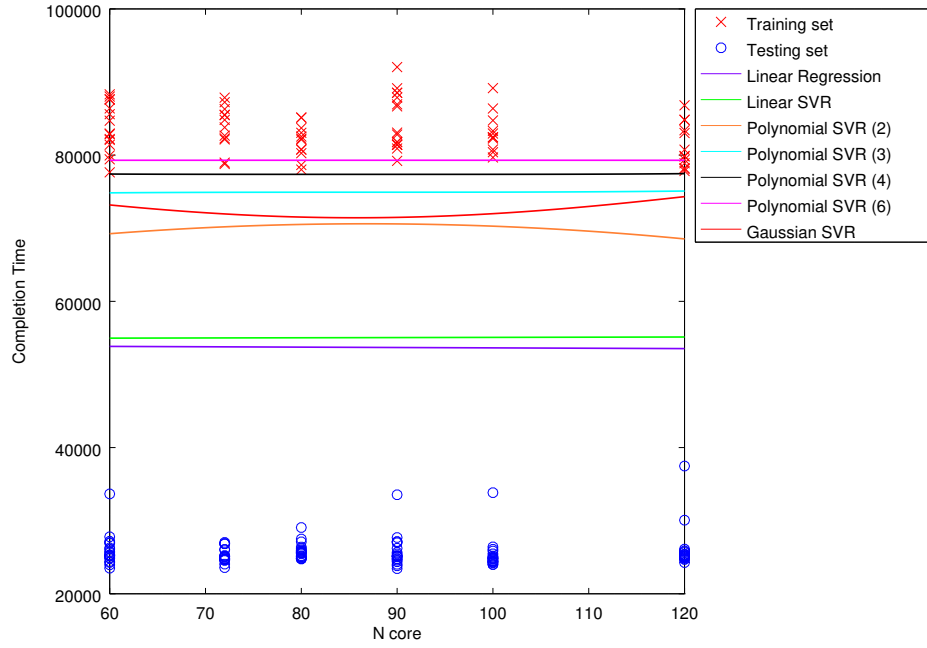
**Figure 44:** Completion time vs ncores for: R5 → R2 with datasize 1000GB

### 3.2 R2 $\rightarrow$ R5 (Fast - Fast)

#### 3.2.1 R2 $\rightarrow$ R5 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.5076	- 54.2010	14618	0.5690
Linear SVR	0.0742	0.5354	1906	0.0748
Polynomial SVR (2)	1.9582	0.5005	56793	2.2284
Polynomial SVR (3)	1.4334	0.4575	41627	1.6299
Polynomial SVR (4)	1.9578	0.4334	56818	2.2282
Polynomial SVR (6)	1.9529	0.3733	56692	2.2226
Gaussian SVR	1.8646	0.4311	54134	2.1221

**Table 45:** Results for R2  $\rightarrow$  R5-250



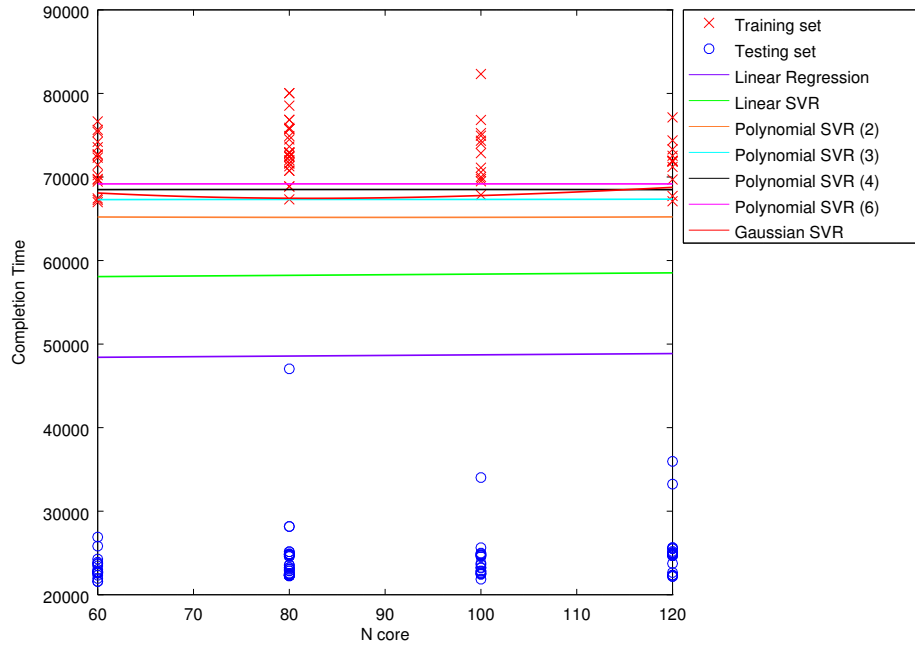
**Figure 45:** Completion time vs ncores for: R2  $\rightarrow$  R5 with datasize 250GB



### 3.2.2 R2 → R5 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2910	-3.7899	6672	0.2718
Linear SVR	0.7805	0.3478	19144	0.8083
Polynomial SVR (2)	1.9764	0.3328	48648	2.0492
Polynomial SVR (3)	1.5362	0.3112	37827	1.5916
Polynomial SVR (4)	1.9546	0.2841	48117	2.0265
Polynomial SVR (6)	1.9390	0.2654	47733	2.0103
Gaussian SVR	1.9761	0.1983	48681	2.0483

**Table 46:** Results for R2 → R5-500

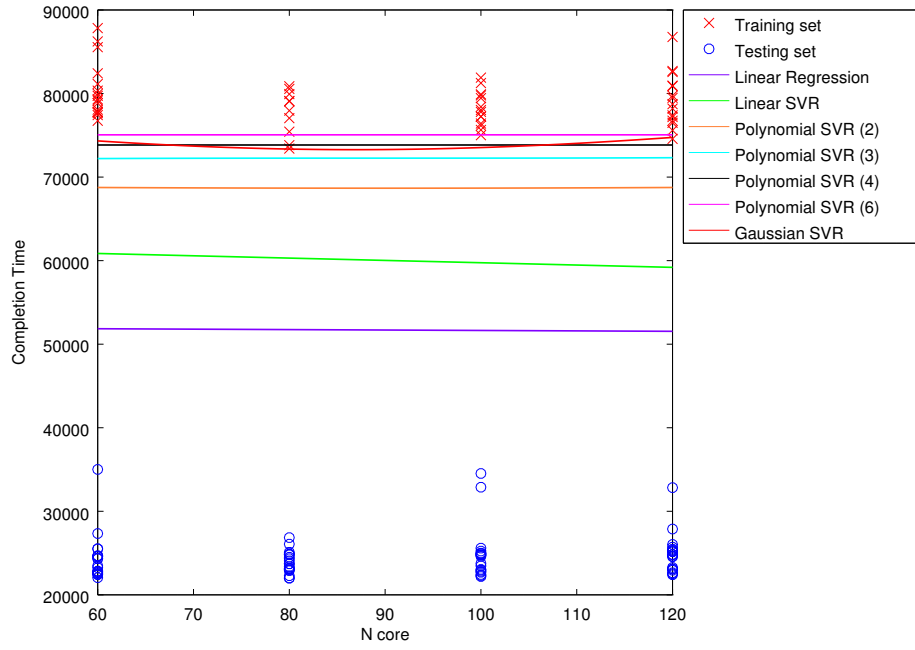


**Figure 46:** Completion time vs ncores for: R2 → R5 with datasize 500GB

### 3.2.3 R2 $\rightarrow$ R5 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.7050	-	19273	0.7906
Linear SVR	0.5814	0.1630	15781	0.6549
Polynomial SVR (2)	1.9918	0.3443	54498	2.2536
Polynomial SVR (3)	1.5014	0.3534	41121	1.6974
Polynomial SVR (4)	1.9754	0.2597	54059	2.2349
Polynomial SVR (6)	1.9710	0.2735	53943	2.2300
Gaussian SVR	2.0285	0.4331	55559	2.2948

**Table 47:** Results for R2  $\rightarrow$  R5-750

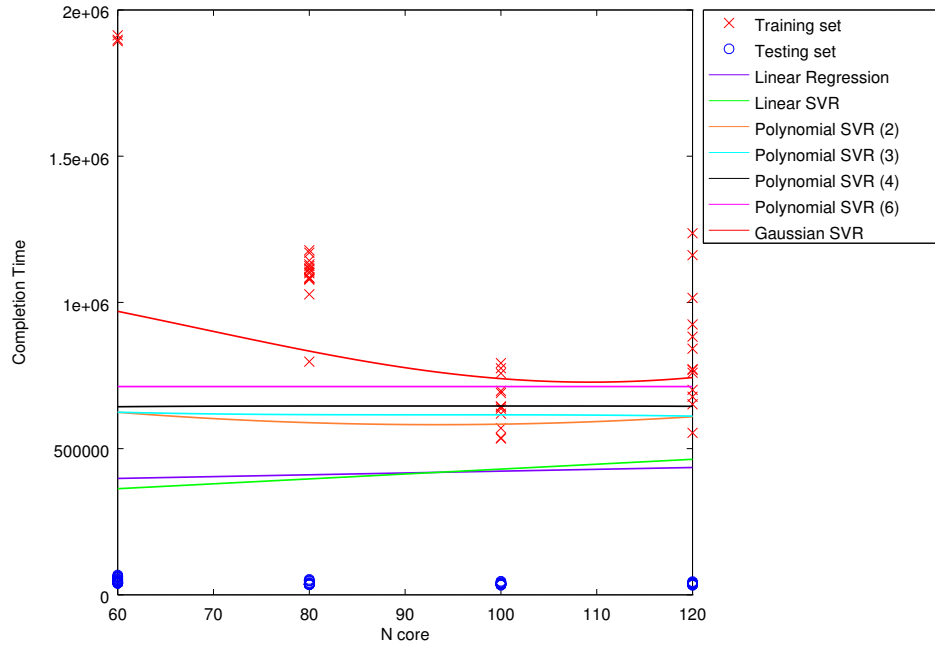


**Figure 47:** Completion time vs ncores for: R2  $\rightarrow$  R5 with datasize 750GB

### 3.2.4 R2 → R5 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2457	-	117597	2.9684
Linear SVR	0.0746	0.0791	31327	0.7786
Polynomial SVR (2)	1.5254	0.2338	734175	18.4731
Polynomial SVR (3)	0.8871	0.1001	426009	10.6683
Polynomial SVR (4)	1.4660	0.0268	705234	17.7261
Polynomial SVR (6)	1.4552	0.0225	700727	17.5862
Gaussian SVR	2.1204	0.0853	1020795	25.5869

**Table 48:** Results for R2 → R5-1000



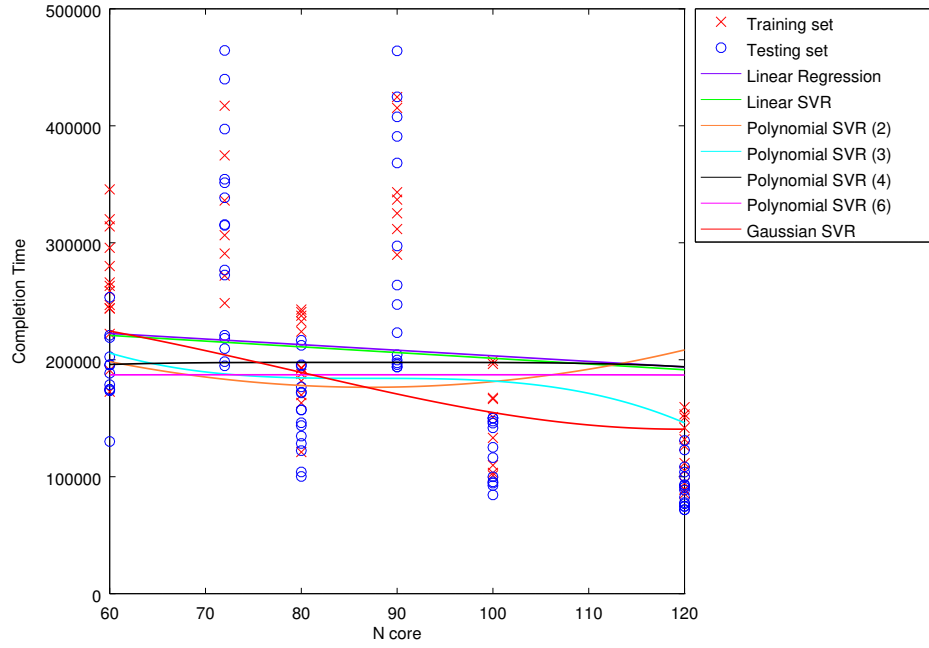
**Figure 48:** Completion time vs ncores for: R2 → R5 with datasize 1000GB

### 3.3 R3 → R4 (Slow - Slow)

#### 3.3.1 R3 → R4 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0999	0.9905	7743	0.0556
Linear SVR	0.1017	0.9936	7894	0.0559
Polynomial SVR (2)	1.2864	0.0000	109952	0.8321
Polynomial SVR (3)	0.5276	0.8127	40426	0.2156
Polynomial SVR (4)	0.9298	0.2785	77573	0.5374
Polynomial SVR (6)	0.8746	0.2982	69520	0.4722
Gaussian SVR	0.5125	0.8738	39988	0.2815

**Table 49:** Results for R3 → R4-250

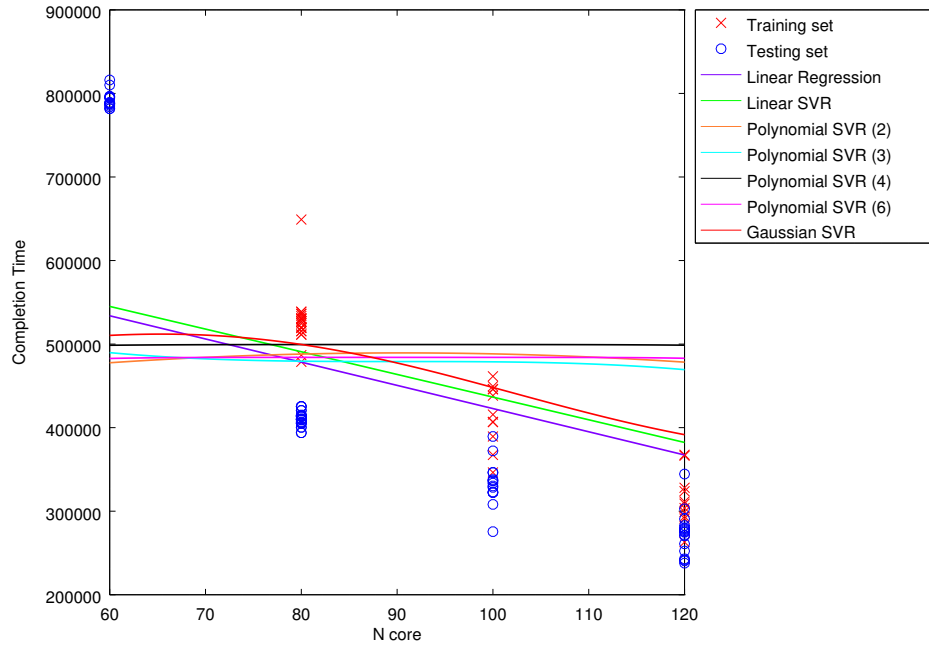


**Figure 49:** Completion time vs ncores for: R3 → R4 with datasize 250GB

### 3.3.2 R3 → R4 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.5575	0.7917	79305	0.1806
Linear SVR	0.6248	0.8895	87857	0.1990
Polynomial SVR (2)	1.4064	0.3856	175966	0.4039
Polynomial SVR (3)	0.9288	0.7713	129311	0.3215
Polynomial SVR (4)	1.3185	0.1435	177714	0.4267
Polynomial SVR (6)	1.4713	0.1476	177441	0.4040
Gaussian SVR	1.0797	0.3126	134221	0.2887

**Table 50:** Results for R3 → R4-500

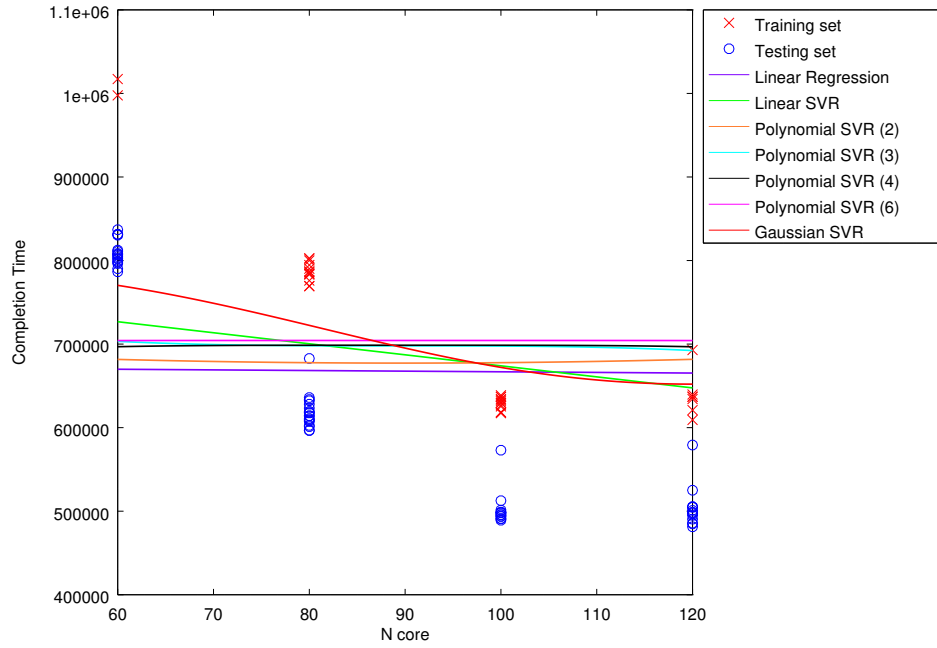


**Figure 50:** Completion time vs ncores for: R3 → R4 with datasize 500GB

### 3.3.3 R3 → R4 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0933	0.9899	12026	0.0195
Linear SVR	0.2629	0.9912	33564	0.0561
Polynomial SVR (2)	1.8337	0.7689	217984	0.3828
Polynomial SVR (3)	0.7352	0.7039	87799	0.1548
Polynomial SVR (4)	1.8539	0.4641	210264	0.3621
Polynomial SVR (6)	1.1873	0.4401	148303	0.2645
Gaussian SVR	1.0774	0.8421	126509	0.2349

**Table 51:** Results for R3 → R4-750

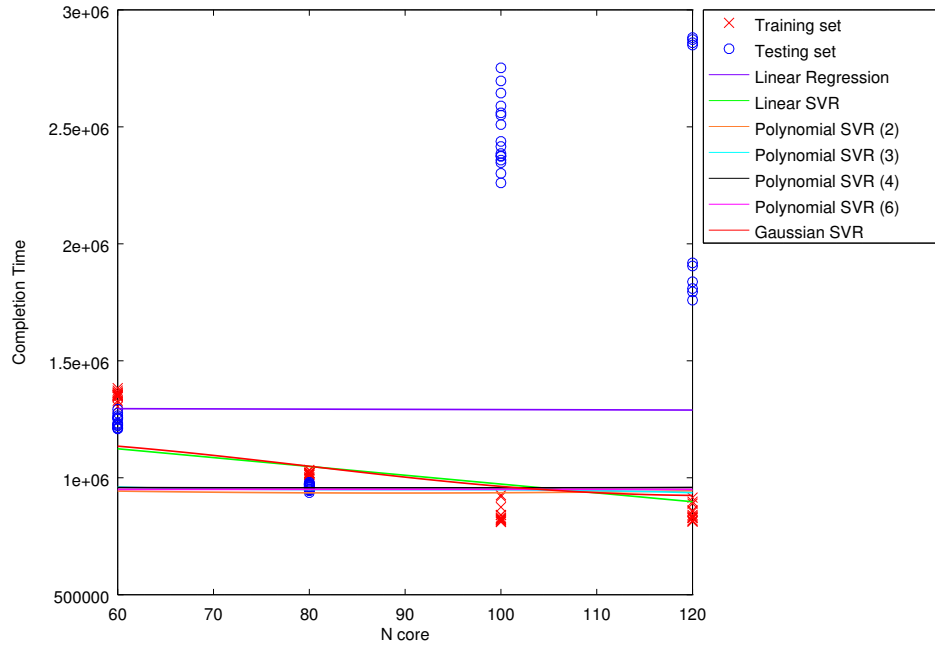


**Figure 51:** Completion time vs ncores for: R3 → R4 with datasize 750GB

### 3.3.4 R3 → R4 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	1.7878	-1.2231	784437	0.3797
Linear SVR	1.6361	0.0281	660983	0.2902
Polynomial SVR (2)	1.6549	0.0157	689481	0.3172
Polynomial SVR (3)	1.6757	0.0650	705047	0.3218
Polynomial SVR (4)	1.9237	0.1209	764929	0.3431
Polynomial SVR (6)	1.8627	0.0254	735220	0.3294
Gaussian SVR	1.5644	0.0772	640303	0.2875

**Table 52:** Results for R3 → R4-1000



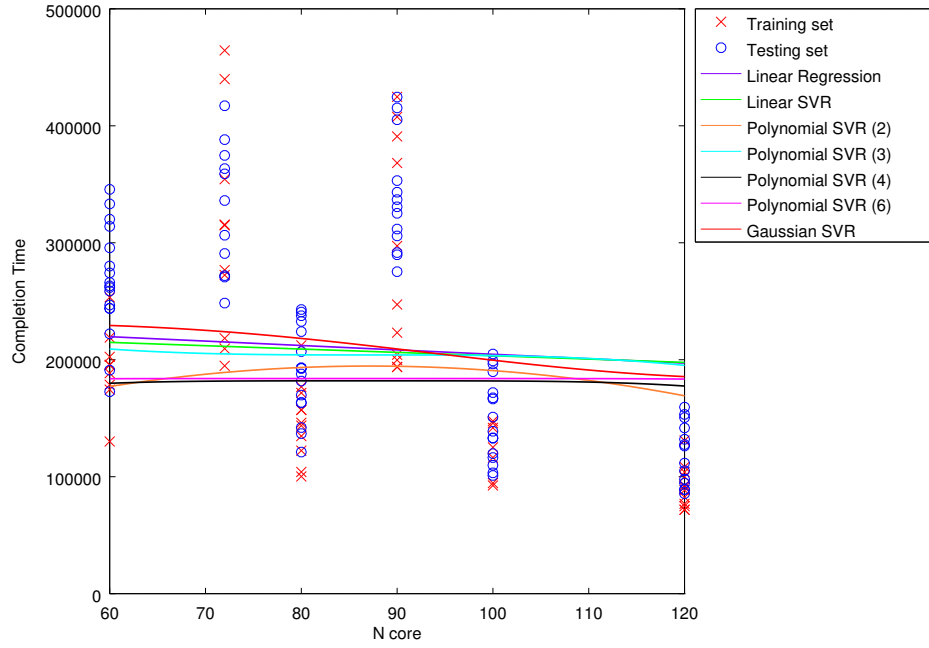
**Figure 52:** Completion time vs ncores for: R3 → R4 with datasize 1000GB

### 3.4 R4 → R3 (Slow - Slow)

#### 3.4.1 R4 → R3 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1117	0.9859	6402	0.0272
Linear SVR	0.1241	0.9882	8184	0.0379
Polynomial SVR (2)	1.0187	0.0102	84602	0.4395
Polynomial SVR (3)	0.7609	0.7115	53158	0.2639
Polynomial SVR (4)	1.0989	0.0178	85066	0.3771
Polynomial SVR (6)	1.0632	0.0244	82125	0.3725
Gaussian SVR	0.5420	0.8101	41963	0.2180

**Table 53:** Results for R4 → R3-250



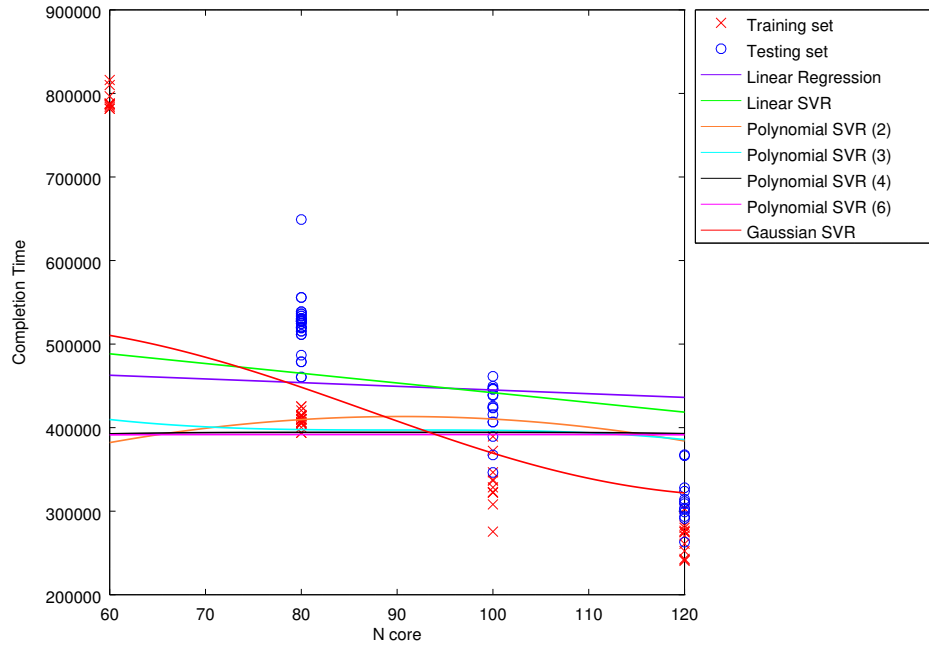
**Figure 53:** Completion time vs ncores for: R4 → R3 with datasize 250GB



### 3.4.2 R4 → R3 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1181	0.9592	13349	0.0297
Linear SVR	0.1281	0.9719	13961	0.0304
Polynomial SVR (2)	1.0739	0.0989	142216	0.3246
Polynomial SVR (3)	0.5917	0.7504	77226	0.1606
Polynomial SVR (4)	1.0211	0.2288	134494	0.3151
Polynomial SVR (6)	0.9462	0.0853	117567	0.2718
Gaussian SVR	0.3507	0.7951	46118	0.1032

**Table 54:** Results for R4 → R3-500

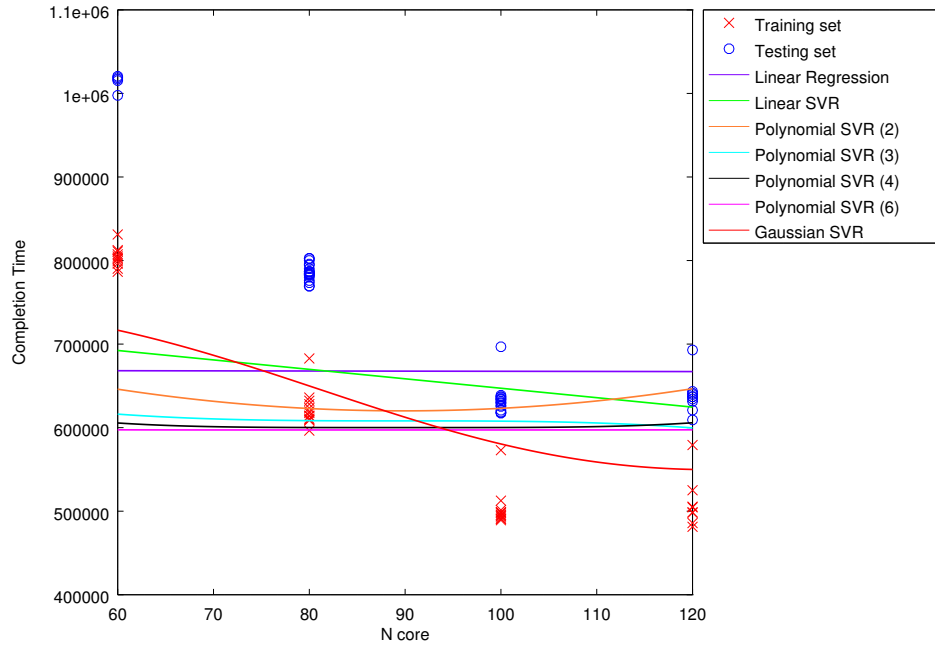


**Figure 54:** Completion time vs ncores for: R4 → R3 with datasize 500GB

### 3.4.3 R4 → R3 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1065	0.9853	13983	0.0192
Linear SVR	0.1572	0.9937	18494	0.0250
Polynomial SVR (2)	1.7890	0.8077	185642	0.2313
Polynomial SVR (3)	1.0593	0.8125	126124	0.1704
Polynomial SVR (4)	2.8642	0.6965	233765	0.2775
Polynomial SVR (6)	2.9392	0.5268	213611	0.2468
Gaussian SVR	0.8538	0.7386	85514	0.1055

**Table 55:** Results for R4 → R3-750

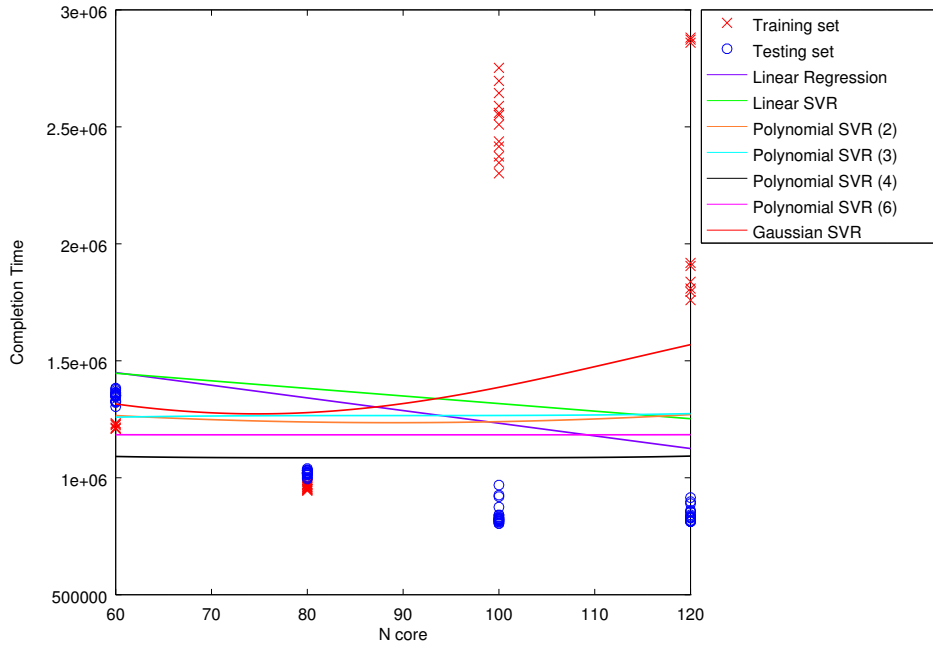


**Figure 55:** Completion time vs ncores for: R4 → R3 with datasize 750GB

### 3.4.4 R4 $\rightarrow$ R3 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2979	0.2999	163252	0.1736
Linear SVR	0.2150	0.9597	109503	0.1219
Polynomial SVR (2)	0.7033	0.6852	382105	0.4195
Polynomial SVR (3)	0.5805	0.7986	324151	0.3440
Polynomial SVR (4)	0.3638	0.7740	189564	0.2076
Polynomial SVR (6)	0.4766	0.6324	252222	0.2741
Gaussian SVR	1.0142	0.5173	574326	0.6115

**Table 56:** Results for R4  $\rightarrow$  R3-1000



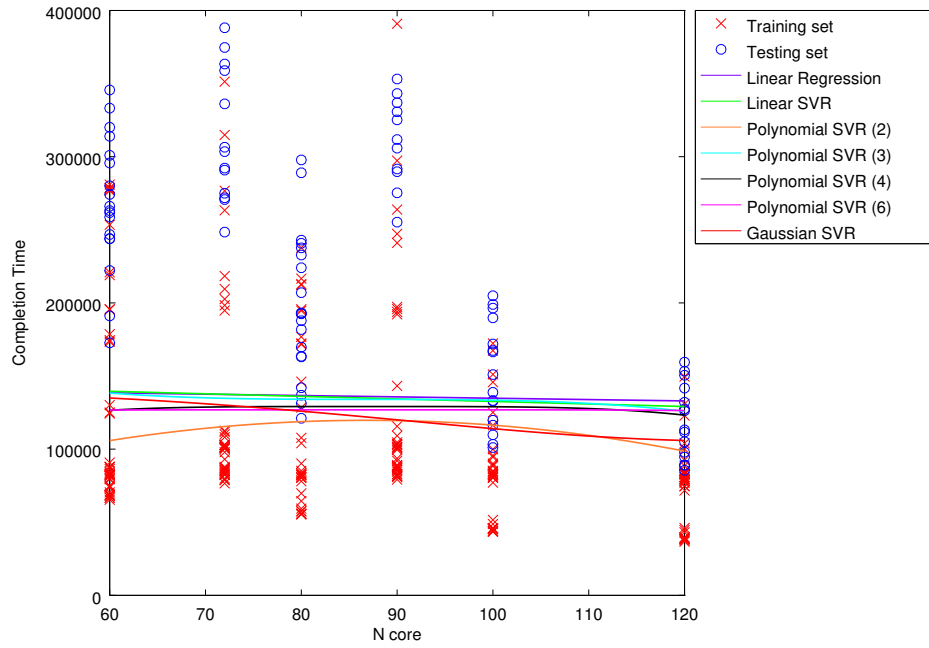
**Figure 56:** Completion time vs ncores for: R4  $\rightarrow$  R3 with datasize 1000GB

### 3.5 R1, R2, R4 $\rightarrow$ R3 (Fast and Slow - Average)

#### 3.5.1 R1,R2,R4 $\rightarrow$ R3 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1484	0.9772	9501	0.0466
Linear SVR	0.1560	0.9845	10574	0.0553
Polynomial SVR (2)	0.8548	0.7060	57023	0.2555
Polynomial SVR (3)	0.4469	0.8616	28093	0.1225
Polynomial SVR (4)	0.7404	0.7503	42222	0.1783
Polynomial SVR (6)	1.1648	0.5504	61124	0.2455
Gaussian SVR	0.3926	0.9405	24779	0.1084

**Table 57:** Results for R1,R2,R4  $\rightarrow$  R3-250

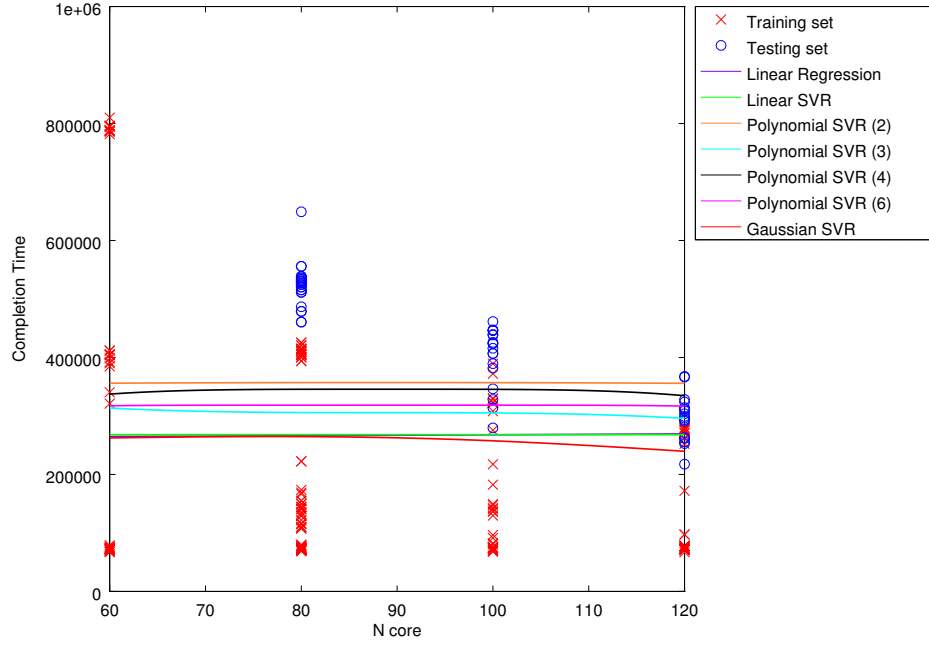


**Figure 57:** Completion time vs ncores for: R1,R2,R4  $\rightarrow$  R3 with datasize 250GB

### 3.5.2 R1,R2,R4 $\rightarrow$ R3 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0914	0.9687	13457	0.0317
Linear SVR	0.0833	0.9741	13357	0.0308
Polynomial SVR (2)	0.6833	0.6864	118761	0.2933
Polynomial SVR (3)	0.2660	0.8449	42425	0.0947
Polynomial SVR (4)	0.7291	0.6105	129234	0.3055
Polynomial SVR (6)	0.7959	0.4006	136789	0.3104
Gaussian SVR	0.2844	0.8990	43387	0.1147

**Table 58:** Results for R1,R2,R4  $\rightarrow$  R3-500

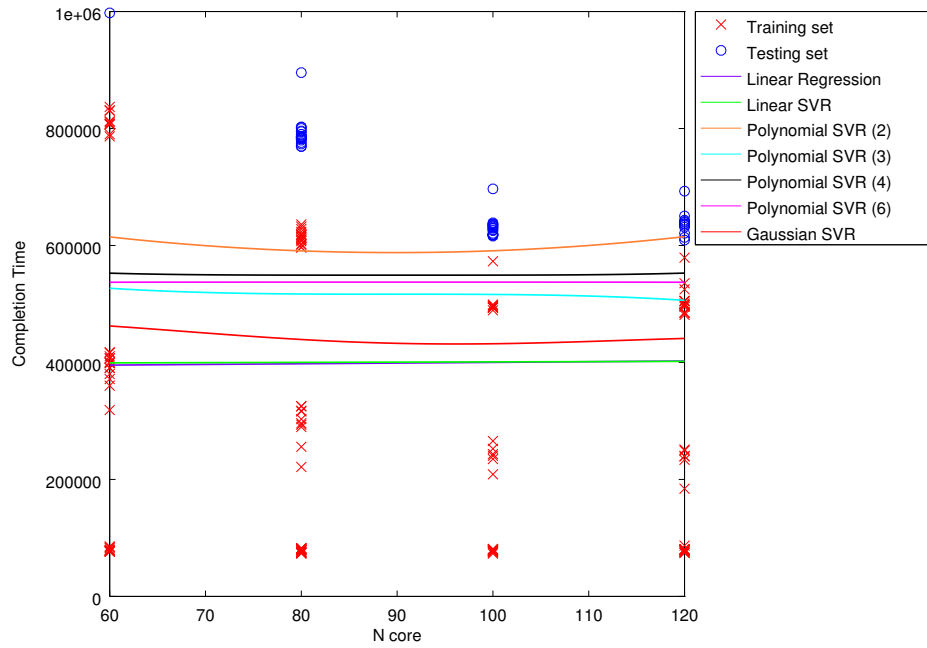


**Figure 58:** Completion time vs ncores for: R1,R2,R4  $\rightarrow$  R3 with datasize 500GB

### 3.5.3 R1,R2,R4 $\rightarrow$ R3 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0581	0.9694	13358	0.0186
Linear SVR	0.0432	0.9924	10296	0.0152
Polynomial SVR (2)	0.4219	0.7036	106082	0.1502
Polynomial SVR (3)	0.2763	0.7323	53615	0.0742
Polynomial SVR (4)	0.6688	0.4095	168830	0.2349
Polynomial SVR (6)	0.6903	0.2073	166482	0.2269
Gaussian SVR	0.3779	0.5523	82501	0.1133

**Table 59:** Results for R1,R2,R4  $\rightarrow$  R3-750

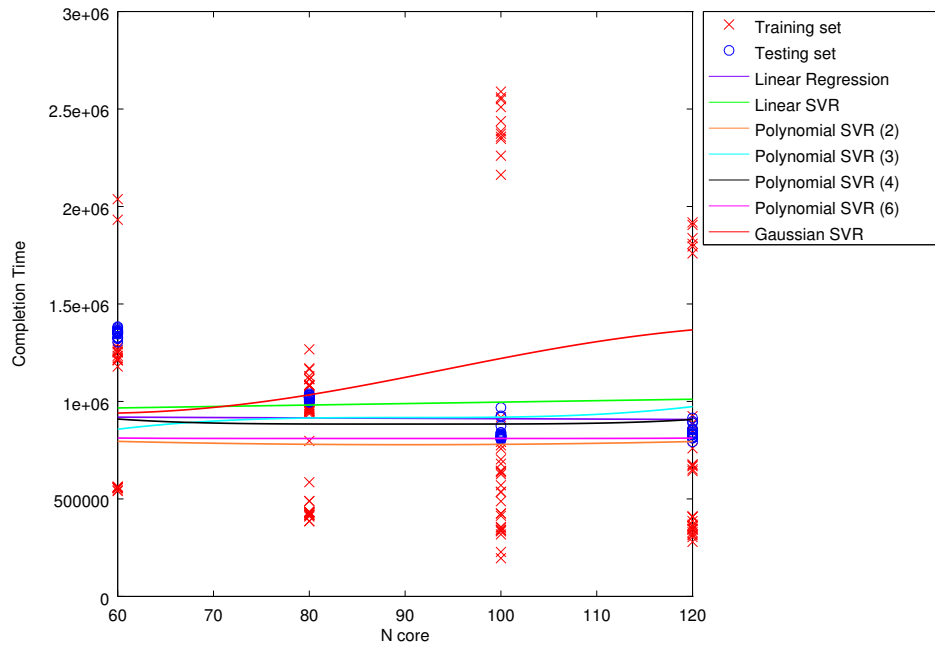


**Figure 59:** Completion time vs ncores for: R1,R2,R4  $\rightarrow$  R3 with datasize 750GB

### 3.5.4 R1,R2,R4 $\rightarrow$ R3 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.5423	-0.7126	260530	0.2553
Linear SVR	0.5775	0.9500	275655	0.2718
Polynomial SVR (2)	0.5045	0.7467	202083	0.1809
Polynomial SVR (3)	0.4542	0.0345	148622	0.1290
Polynomial SVR (4)	0.5079	0.0024	181230	0.1545
Polynomial SVR (6)	0.5769	0.0175	203255	0.1711
Gaussian SVR	0.3920	0.7004	168171	0.1877

**Table 60:** Results for R1,R2,R4  $\rightarrow$  R3-1000



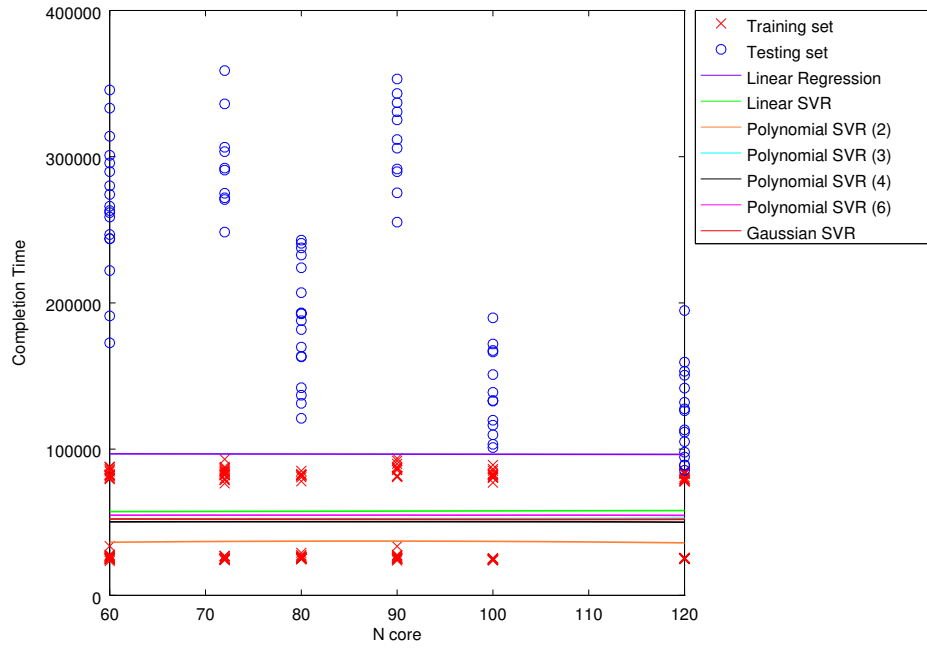
**Figure 60:** Completion time vs ncores for: R1,R2,R4  $\rightarrow$  R3 with datasize 1000GB

### 3.6 R2, R5 $\rightarrow$ R3 (Fast - Slow)

#### 3.6.1 R2,R5 $\rightarrow$ R3 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	2.9358	-8.9645	225067	1.0067
Linear SVR	2.0189	0.1000	154238	0.6860
Polynomial SVR (2)	2.3413	0.0143	179635	0.8056
Polynomial SVR (3)	2.0566	0.0575	157367	0.6993
Polynomial SVR (4)	2.1957	0.0474	167202	0.7392
Polynomial SVR (6)	2.1578	0.0358	163842	0.7215
Gaussian SVR	2.0866	0.2880	160725	0.7205

**Table 61:** Results for R2,R5  $\rightarrow$  R3-250



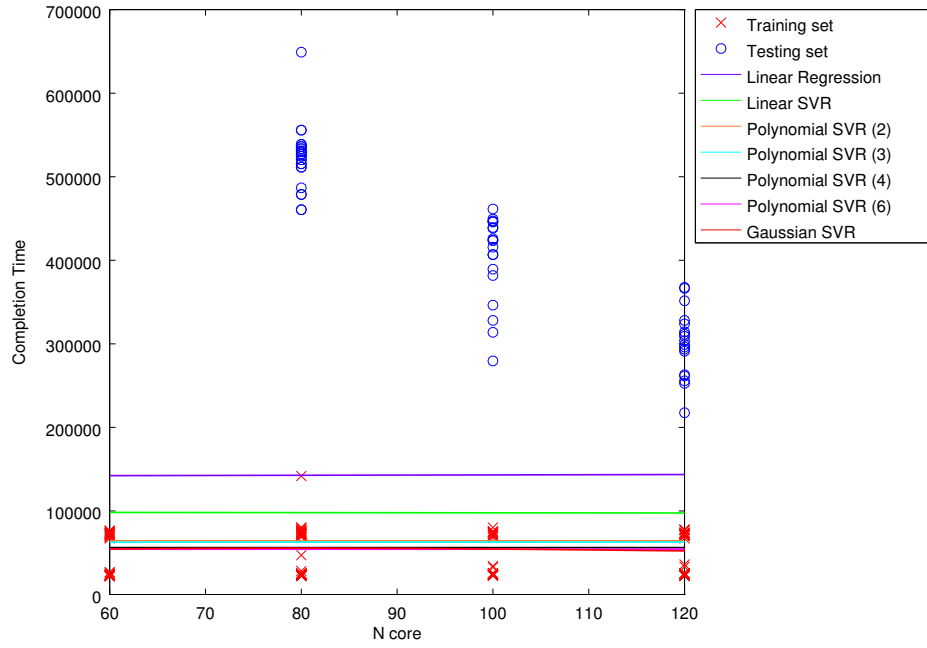
**Figure 61:** Completion time vs ncores for: R2,R5  $\rightarrow$  R3 with datasize 250GB



### 3.6.2 R2,R5 $\rightarrow$ R3 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	24.4317	-	4082755	9.5054
Linear SVR	1.3901	0.0020	217391	0.4766
Polynomial SVR (2)	2.3279	0.8694	385016	0.8855
Polynomial SVR (3)	2.1472	0.6028	358017	0.8305
Polynomial SVR (4)	2.2733	0.4990	377968	0.8744
Polynomial SVR (6)	2.2643	0.2486	376874	0.8730
Gaussian SVR	2.2414	0.1919	373980	0.8683

**Table 62:** Results for R2,R5  $\rightarrow$  R3-500

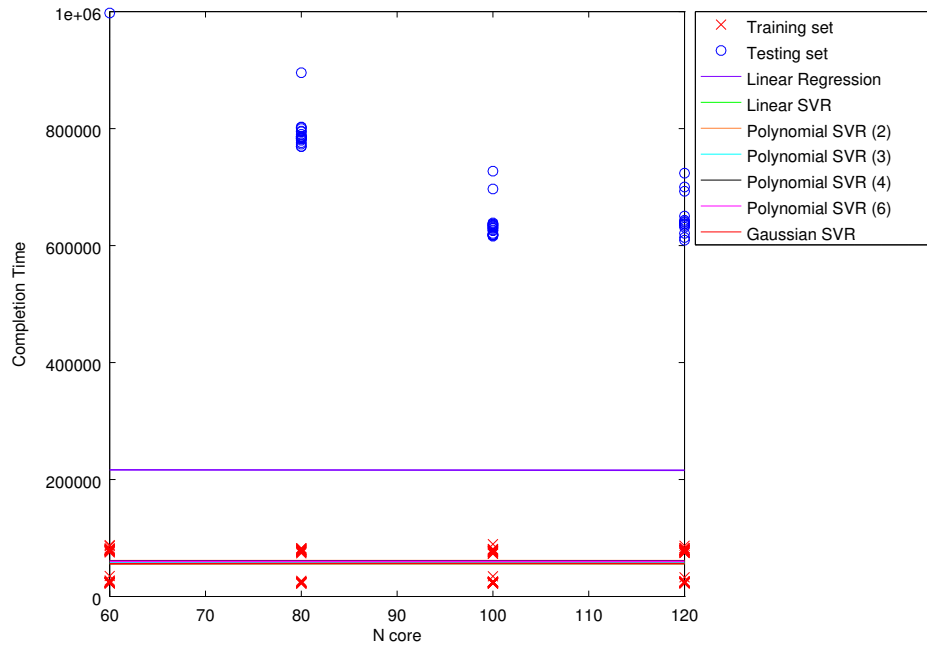


**Figure 62:** Completion time vs ncores for: R2,R5  $\rightarrow$  R3 with datasize 500GB

### 3.6.3 R2,R5 → R3 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	3.1762	-	905865	1.2847
Linear SVR	2.2593	0.0458	645146	0.9158
Polynomial SVR (2)	2.2682	0.1040	647657	0.9193
Polynomial SVR (3)	2.2586	0.3000	644988	0.9156
Polynomial SVR (4)	2.2552	0.0930	643893	0.9139
Polynomial SVR (6)	2.2549	0.0726	643840	0.9139
Gaussian SVR	2.2664	0.6737	647187	0.9187

**Table 63:** Results for R2,R5 → R3-750

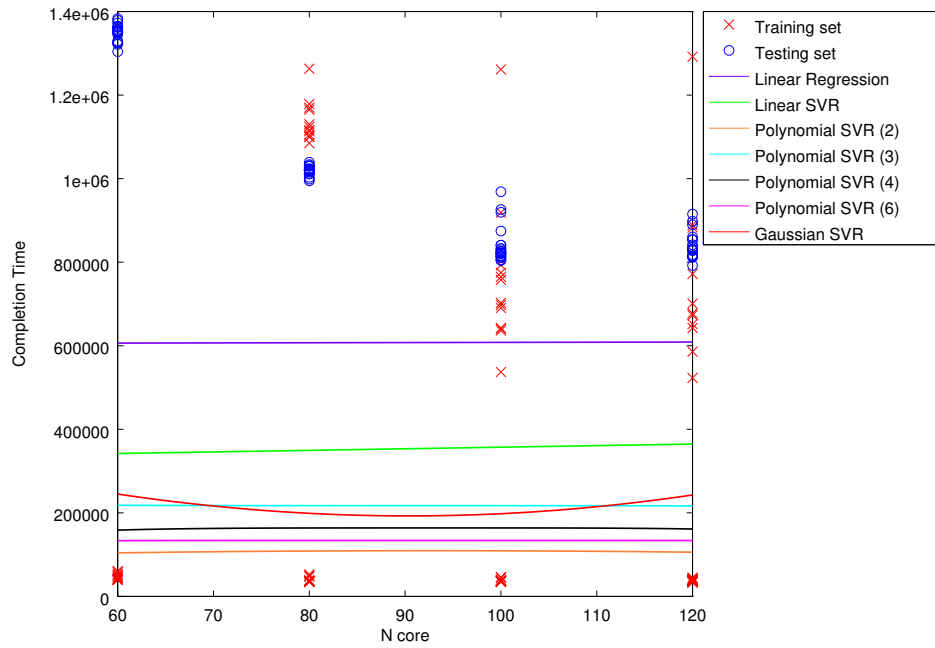


**Figure 63:** Completion time vs ncores for: R2,R5 → R3 with datasize 750GB

### 3.6.4 R2,R5 $\rightarrow$ R3 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.4825	-0.2162	225269	0.2231
Linear SVR	1.5082	0.8905	700319	0.6897
Polynomial SVR (2)	1.8532	0.0474	856486	0.8417
Polynomial SVR (3)	1.7507	0.0017	810643	0.7966
Polynomial SVR (4)	1.8048	0.1072	838182	0.8256
Polynomial SVR (6)	1.8737	0.3832	871983	0.8603
Gaussian SVR	1.5367	0.8611	723592	0.7227

**Table 64:** Results for R2,R5  $\rightarrow$  R3-1000



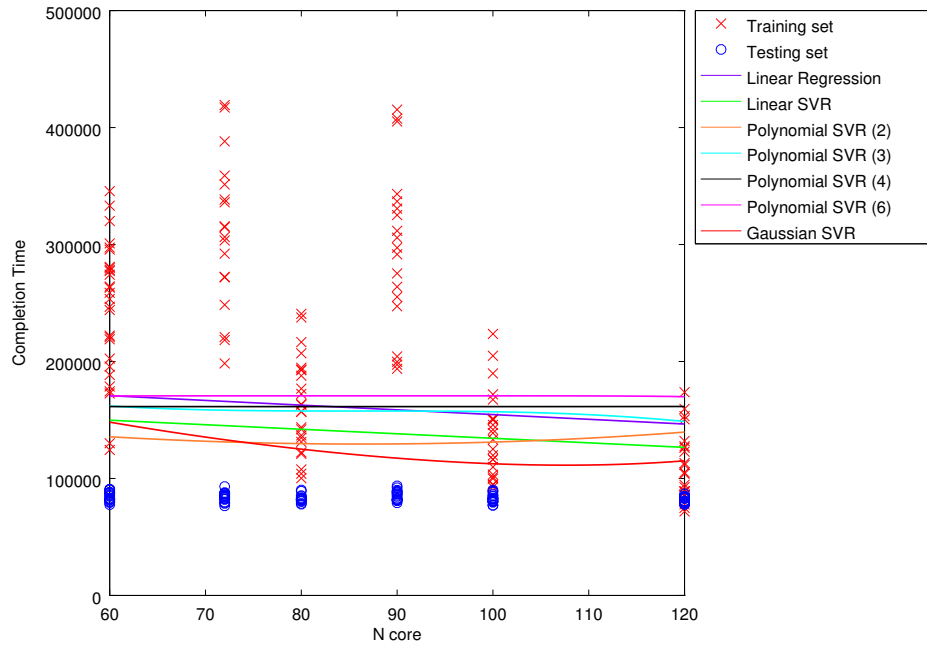
**Figure 64:** Completion time vs ncores for: R2,R5  $\rightarrow$  R3 with datasize 1000GB

### 3.7 R3, R4 $\rightarrow$ R2 (Fast - Slow)

#### 3.7.1 R3,R4 $\rightarrow$ R2 — Datasize 250GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.7320	-	67111	0.8046
Linear SVR	0.6184	0.3343	56625	0.6788
Polynomial SVR (2)	2.2958	0.1994	209908	2.5228
Polynomial SVR (3)	0.6067	0.0155	47847	0.5734
Polynomial SVR (4)	2.0229	0.0101	180533	2.1680
Polynomial SVR (6)	1.7103	0.0153	146100	1.7568
Gaussian SVR	1.0720	0.2857	98273	1.1766

**Table 65:** Results for R3,R4  $\rightarrow$  R2-250

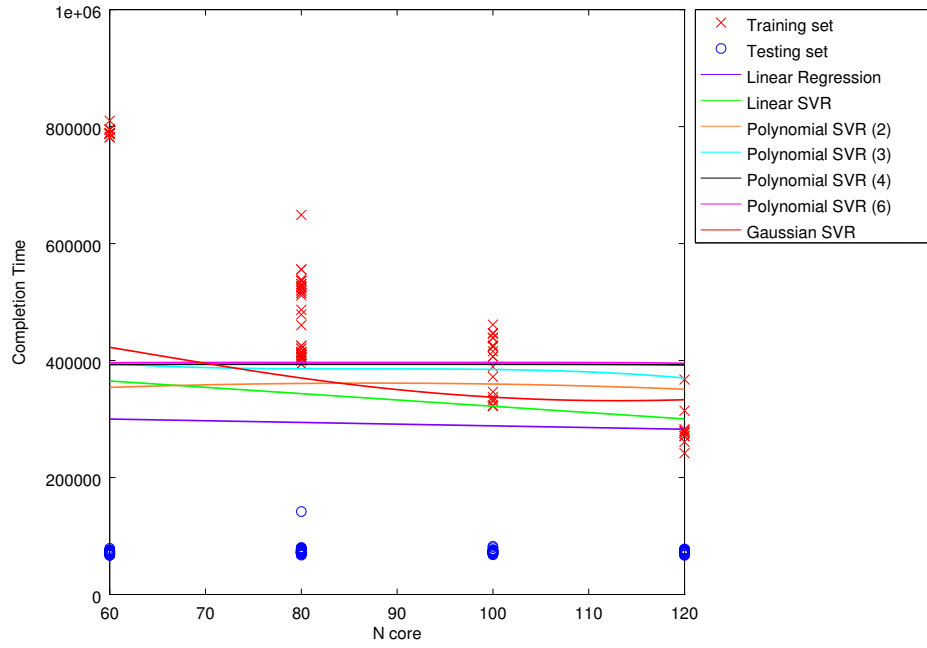


**Figure 65:** Completion time vs ncores for: R3,R4  $\rightarrow$  R2 with datasize 250GB

### 3.7.2 R3,R4 $\rightarrow$ R2 — Datasize 500GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.2274	-	50511	0.6952
Linear SVR	0.4418	0.0006	96855	1.3309
Polynomial SVR (2)	1.8749	0.0345	420366	5.7616
Polynomial SVR (3)	0.9866	0.0290	213001	2.9014
Polynomial SVR (4)	1.8453	0.0260	410413	5.6274
Polynomial SVR (6)	1.7365	0.0248	385605	5.2880
Gaussian SVR	1.4964	0.0442	337994	4.6265

**Table 66:** Results for R3,R4  $\rightarrow$  R2-500

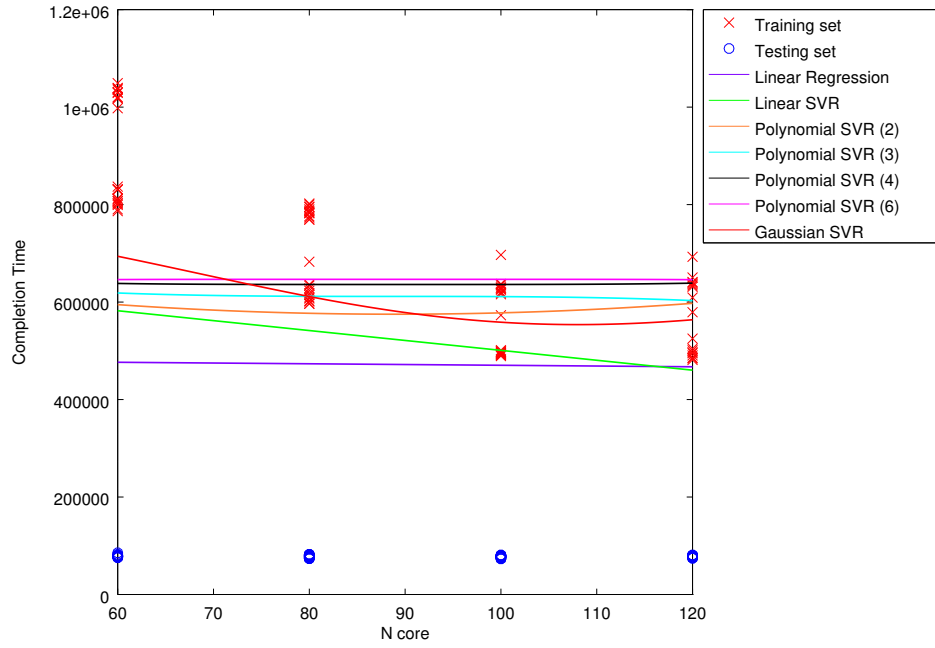


**Figure 66:** Completion time vs ncores for: R3,R4  $\rightarrow$  R2 with datasize 500GB

### 3.7.3 R3,R4 $\rightarrow$ R2 — Datasize 750GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.4540	-	146163	1.8730
Linear SVR	0.4483	0.0421	137083	1.7535
Polynomial SVR (2)	2.4760	0.1503	788379	10.1194
Polynomial SVR (3)	1.1909	0.3274	304007	3.8427
Polynomial SVR (4)	2.6951	0.2837	803028	10.3423
Polynomial SVR (6)	3.3215	0.2411	930987	11.9433
Gaussian SVR	1.9111	0.1845	615634	7.8865

**Table 67:** Results for R3,R4  $\rightarrow$  R2-750

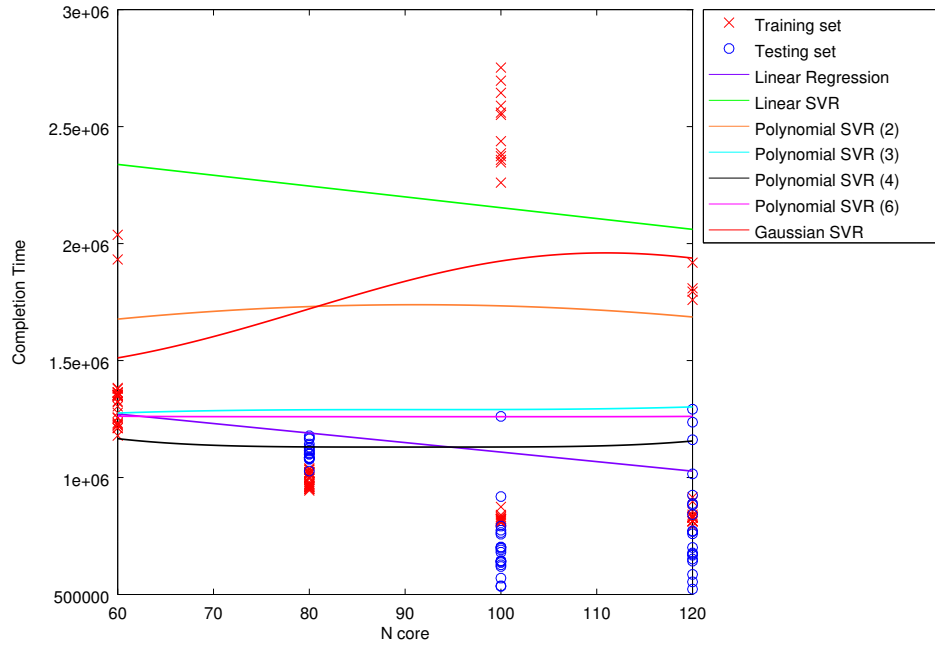


**Figure 67:** Completion time vs ncores for: R3,R4  $\rightarrow$  R2 with datasize 750GB

### 3.7.4 R3,R4 $\rightarrow$ R2 — Datasize 1000GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	11.2761	-	3410161	3.2357
Linear SVR	7.7162	0.9822	3657876	4.2191
Polynomial SVR (2)	8.0373	0.9441	2766694	2.6339
Polynomial SVR (3)	10.5398	0.8977	3750682	3.6335
Polynomial SVR (4)	10.3551	0.8329	3230811	2.9786
Polynomial SVR (6)	7.2376	0.6977	2298685	2.2241
Gaussian SVR	2.4222	0.5863	1124917	1.5115

**Table 68:** Results for R3,R4  $\rightarrow$  R2-1000



**Figure 68:** Completion time vs ncores for: R3,R4  $\rightarrow$  R2 with datasize 1000GB

S



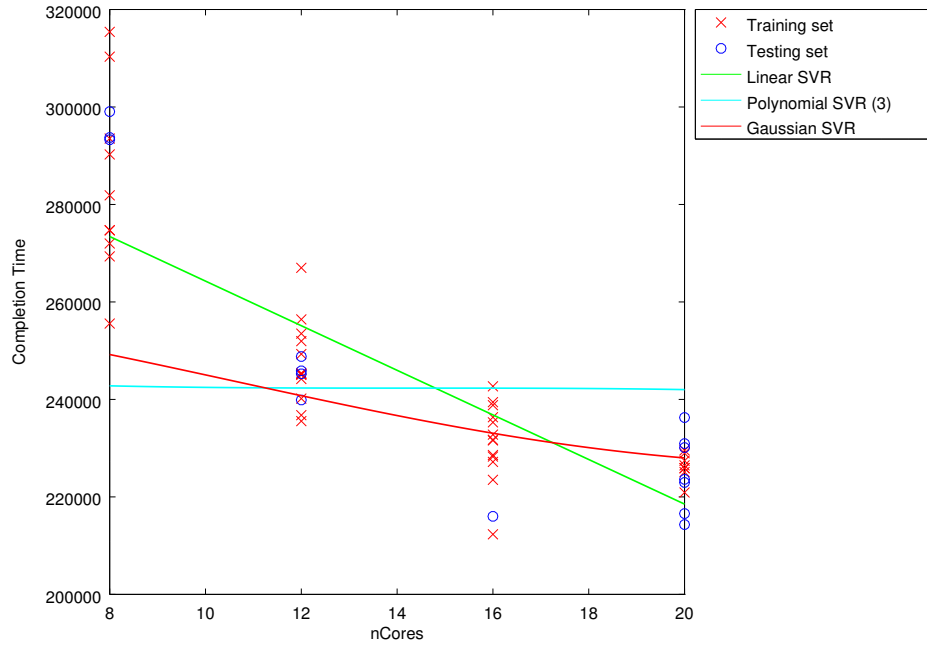
## 4 Fixed Datasize, all the features,

### 4.1 Query Q2

#### 4.1.1 Query Q2 — Datasize 40GB

Model	RMSE	$R^2$	Mean absolute error	Mean relative error
Linear regression	1.6276	-0.9999	32359	0.1341
Linear SVR	0.3547	0.9266	6650	0.0274
Polynomial SVR (2)	0.8078	0.6069	15924	0.0649
Polynomial SVR (3)	0.5883	0.7715	10921	0.0446
Polynomial SVR (4)	1.0508	0.1893	19767	0.0789
Polynomial SVR (6)	1.1443	0.0177	21436	0.0850
Gaussian SVR	0.5819	0.8864	10799	0.0422

**Table 69:** Results for Q2-40GB

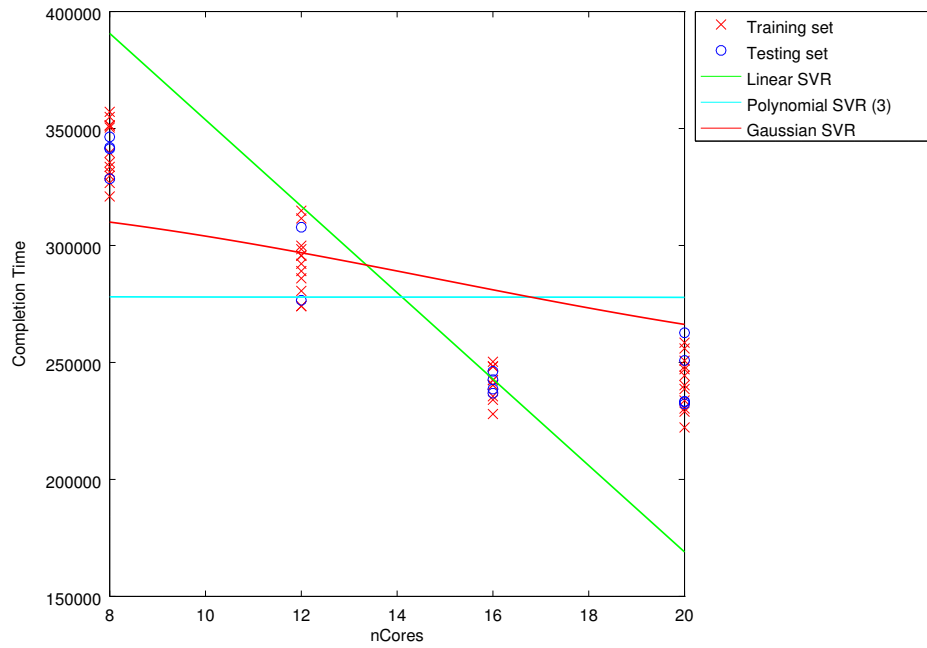


**Figure 69:** Completion time vs ncores for query Q2 with datasize 40

#### 4.1.2 Query Q2 — Datasize 50GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.7974	0.4045	26855	0.0982
Linear SVR	0.3127	0.9112	10937	0.0419
Polynomial SVR (2)	0.7898	0.4304	28540	0.1044
Polynomial SVR (3)	0.7750	0.4572	30331	0.1124
Polynomial SVR (4)	1.2471	0.2083	37425	0.1433
Polynomial SVR (6)	0.9696	0.2704	37310	0.1350
Gaussian SVR	0.5052	0.8391	17201	0.0663

**Table 70:** Results for Q2-50GB



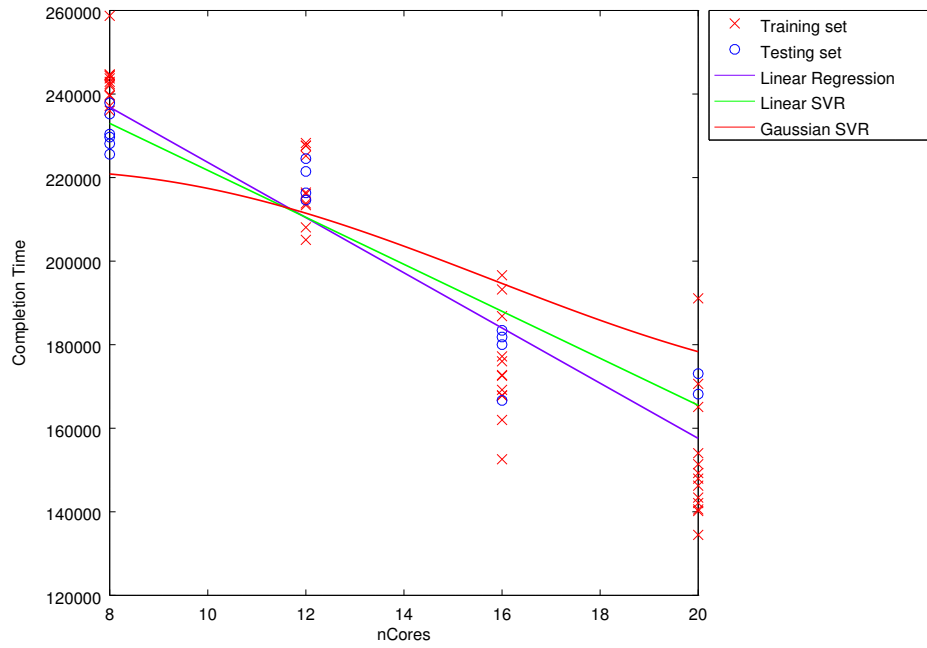
**Figure 70:** Completion time vs ncores for query Q2 with datasize 50

## 4.2 Query Q3

### 4.2.1 Query Q3 — Datasize 50GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1821	0.9385	5147	0.0238
Linear SVR	0.3589	0.8794	10758	0.0502
Polynomial SVR (2)	0.7194	0.0867	23497	0.1173
Polynomial SVR (3)	0.4076	0.7912	11214	0.0514
Polynomial SVR (4)	0.5708	0.5249	14385	0.0683
Polynomial SVR (6)	0.6185	0.4444	17186	0.0817
Gaussian SVR	0.1952	0.9502	5697	0.0269

**Table 71:** Results for Q3-50GB



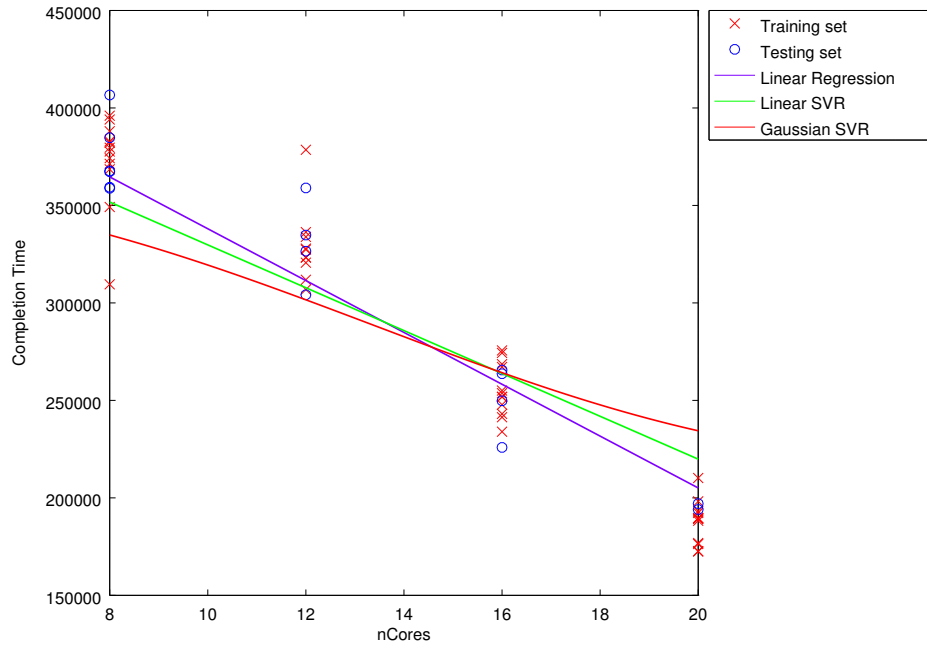
**Figure 71:** Completion time vs ncores for query Q3 with datasize 50

### 4.3 Query Q4

#### 4.3.1 Query Q4 — Datasize 50GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0990	0.9881	4688	0.0154
Linear SVR	0.1198	0.9849	6748	0.0217
Polynomial SVR (2)	0.7944	0.3575	53287	0.1731
Polynomial SVR (3)	0.4045	0.8300	25556	0.0869
Polynomial SVR (4)	0.7117	0.5033	42806	0.1298
Polynomial SVR (6)	0.7431	0.4922	45187	0.1372
Gaussian SVR	0.1867	0.9590	11207	0.0387

**Table 72:** Results for Q4-50GB



**Figure 72:** Completion time vs ncores for query Q4 with datasize 50

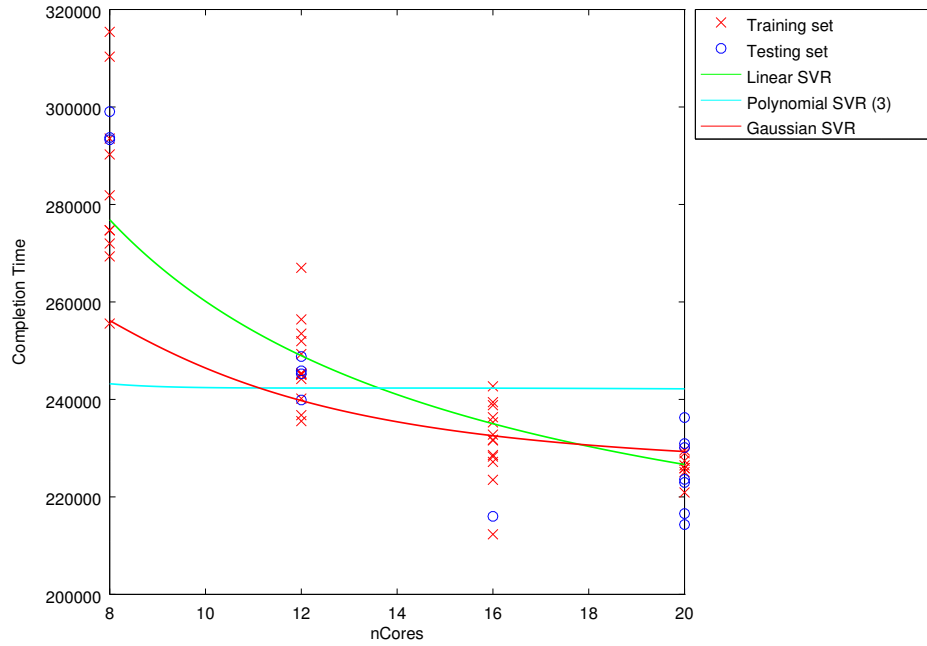
## 5 Fixed Datasize, all features with 1/nCores,

### 5.1 Query Q2

#### 5.1.1 Query Q2 — Datasize 40GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	1.1724	-0.0377	24035	0.1012
Linear SVR	0.3713	0.9310	7170	0.0293
Polynomial SVR (2)	0.7597	0.7219	15618	0.0647
Polynomial SVR (3)	0.5592	0.8204	10138	0.0414
Polynomial SVR (4)	0.9967	0.4517	18992	0.0756
Polynomial SVR (6)	1.1150	0.0733	20847	0.0824
Gaussian SVR	0.5455	0.9096	10367	0.0410

**Table 73:** Results for Q2-40GB

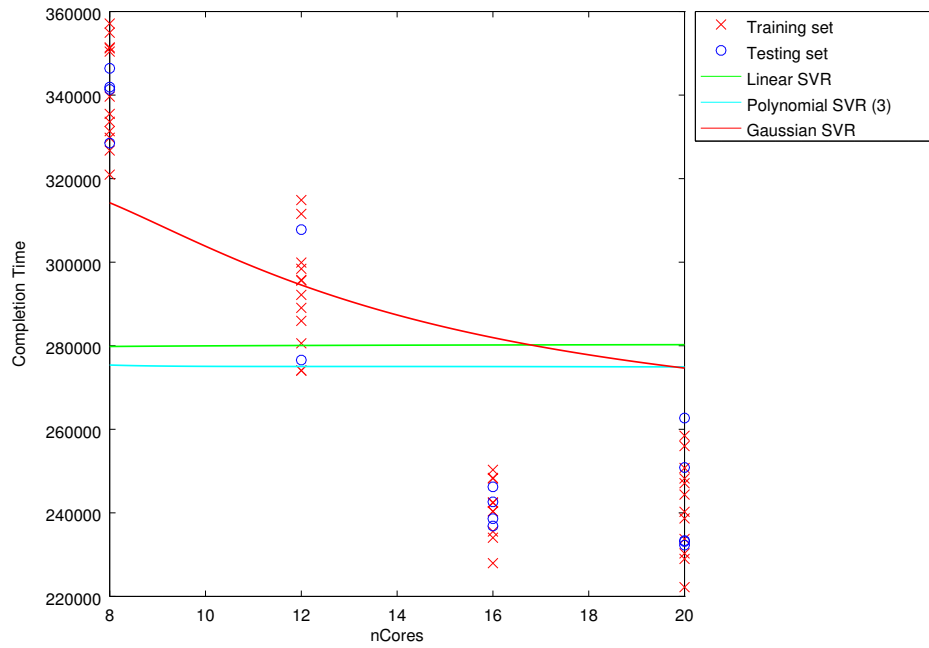


**Figure 73:** Completion time vs ncores for query Q2 with datasize 40

### 5.1.2 Query Q2 — Datasize 50GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.9401	0.1721	35276	0.1336
Linear SVR	0.4863	0.7826	15991	0.0631
Polynomial SVR (2)	0.7756	0.4662	28112	0.1025
Polynomial SVR (3)	0.7518	0.4994	26075	0.0987
Polynomial SVR (4)	1.1967	0.2245	36195	0.1382
Polynomial SVR (6)	1.1080	0.1845	37971	0.1412
Gaussian SVR	0.4857	0.8544	16546	0.0638

**Table 74:** Results for Q2-50GB



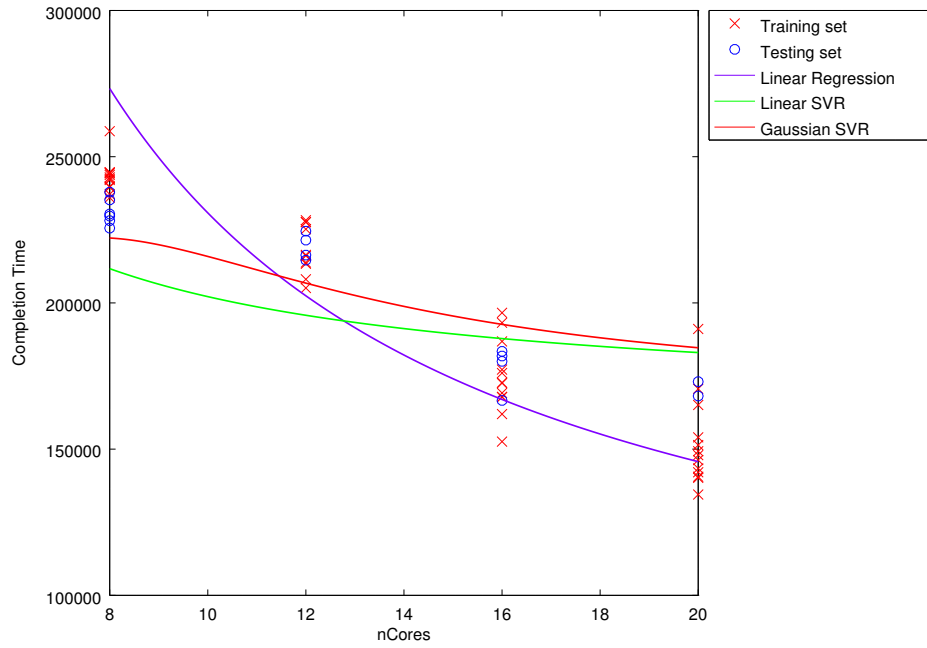
**Figure 74:** Completion time vs ncores for query Q2 with datasize 50

## 5.2 Query Q3

### 5.2.1 Query Q3 — Datasize 50GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.1821	0.9385	5147	0.0238
Linear SVR	0.3698	0.7939	9957	0.0472
Polynomial SVR (2)	0.6657	0.2370	22068	0.1086
Polynomial SVR (3)	0.4115	0.7344	10516	0.0487
Polynomial SVR (4)	0.6275	0.4275	16290	0.0778
Polynomial SVR (6)	0.6407	0.4513	16807	0.0790
Gaussian SVR	0.2121	0.9383	6227	0.0295

**Table 75:** Results for Q3-50GB



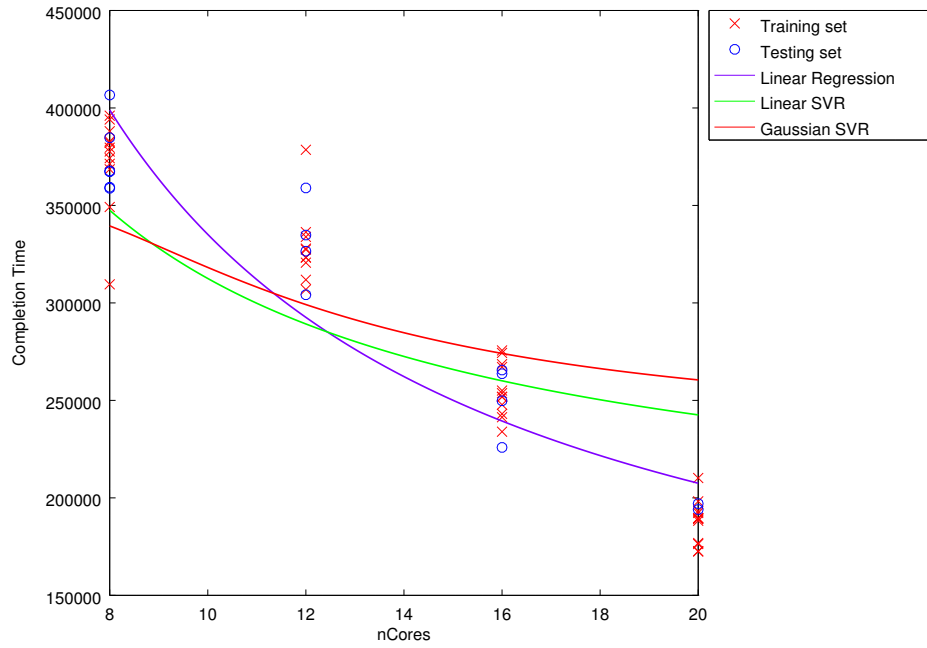
**Figure 75:** Completion time vs ncores for query Q3 with datasize 50

### 5.3 Query Q4

#### 5.3.1 Query Q4 — Datasize 50GB

Model	RMSE	R <sup>2</sup>	Mean absolute error	Mean relative error
Linear regression	0.0990	0.9881	4688	0.0154
Linear SVR	0.1853	0.9602	11231	0.0364
Polynomial SVR (2)	0.6609	0.5640	41416	0.1327
Polynomial SVR (3)	0.3972	0.8135	23263	0.0776
Polynomial SVR (4)	0.4891	0.7383	28158	0.0915
Polynomial SVR (6)	0.6421	0.6013	39270	0.1209
Gaussian SVR	0.2128	0.9463	12910	0.0451

**Table 76:** Results for Q4-50GB



**Figure 76:** Completion time vs ncores for query Q4 with datasize 50