

# 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>2</b>
1.1	Query Q2 . . . . .	2
1.1.1	Query Q2 — Datasize 40GB . . . . .	2
1.1.2	Query Q2 — Datasize 50GB . . . . .	3
1.2	Query Q3 . . . . .	4
1.2.1	Query Q3 — Datasize 50GB . . . . .	4
1.3	Query Q4 . . . . .	5
1.3.1	Query Q4 — Datasize 50GB . . . . .	5

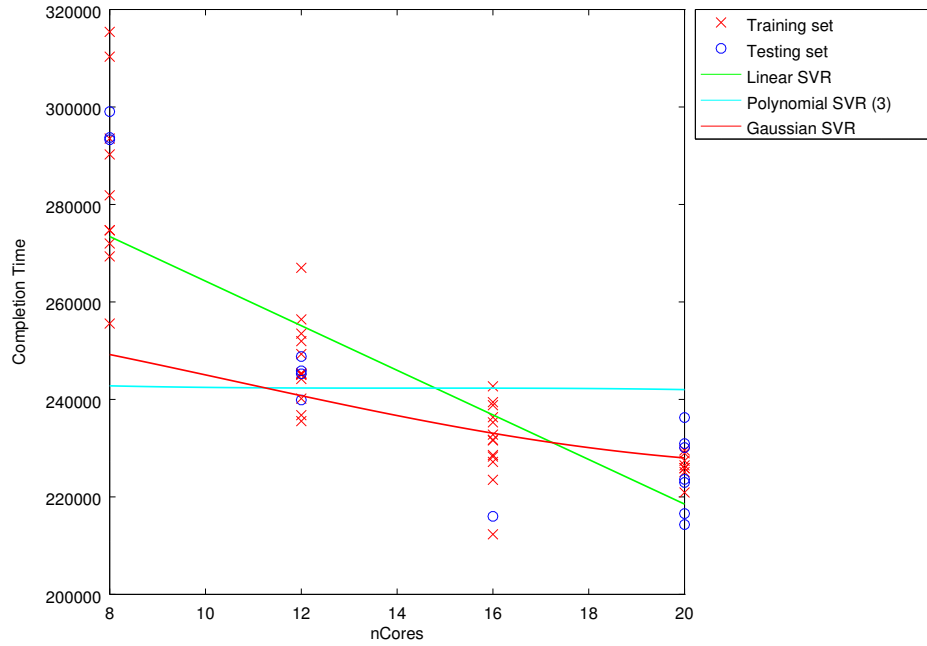
# 1 Fixed Datasize, all the features,

## 1.1 Query Q2

### 1.1.1 Query Q2 — Datasize 40GB

Model	RMSE	R <sup>2</sup>	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 1:** Results for Q2-40GB

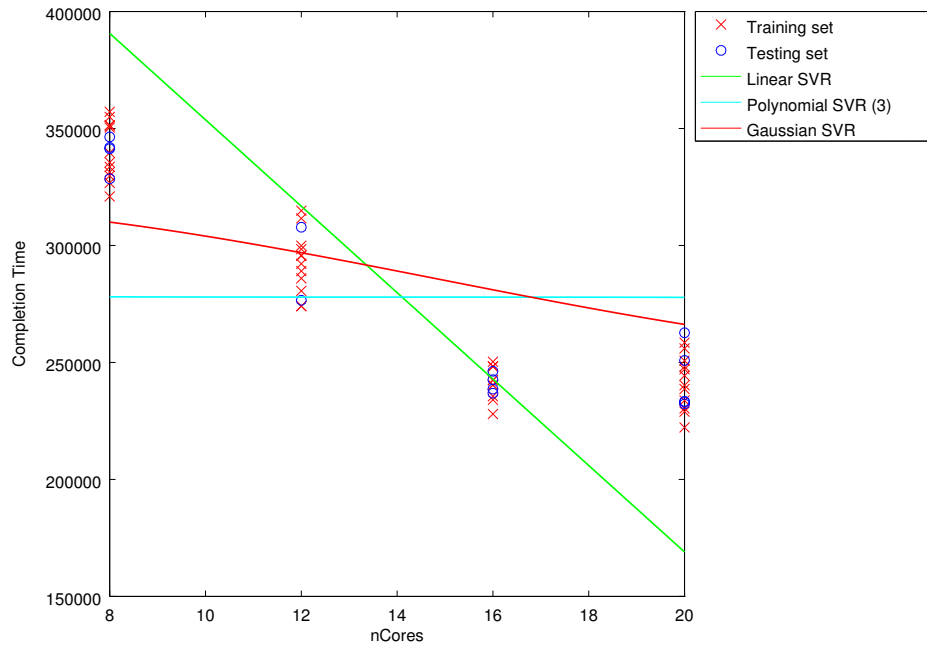


**Figure 1:** Completion time vs ncores for query Q2 with datasize 40

### 1.1.2 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 2:** Results for Q2-50GB



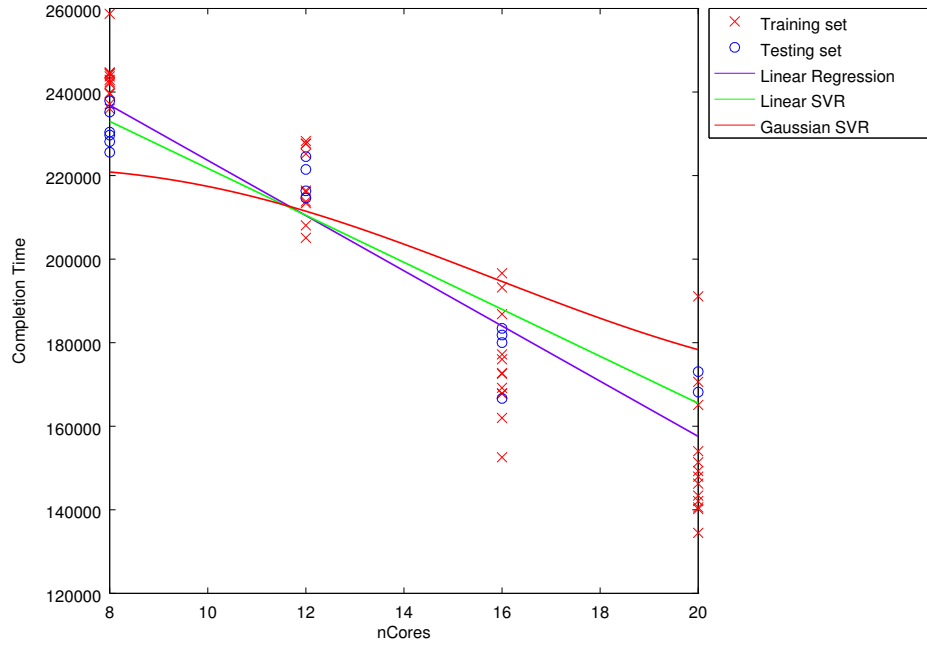
**Figure 2:** Completion time vs ncores for query Q2 with datasize 50

## 1.2 Query Q3

### 1.2.1 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 3:** Results for Q3-50GB



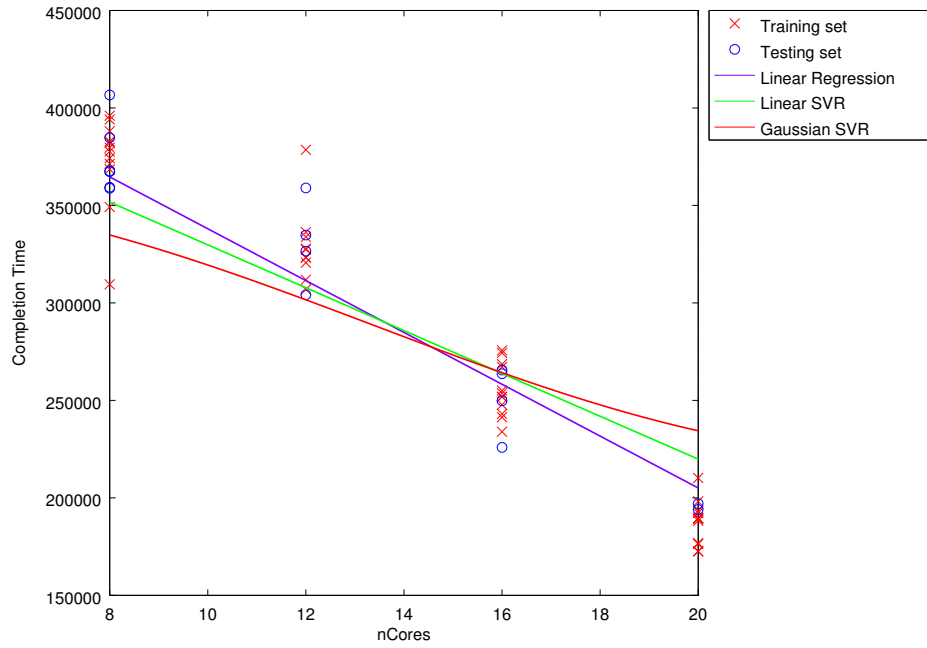
**Figure 3:** Completion time vs ncores for query Q3 with datasize 50

### 1.3 Query Q4

#### 1.3.1 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 4:** Results for Q4-50GB



**Figure 4:** Completion time vs ncores for query Q4 with datasize 50