Project Report

On

Predicting CPU Performance

With

Business Intelligence

Submitted by:

14012121009 (Malav Shastri)

14012121012 (Raj Parmar)

(Oct, 2016)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

U. V. PATEL COLLEGE OF ENGINEERING

GANPAT UNIVERSITY

Project Definition

Description: The CPU performs the relative performance of computer processing power on the basis of a number of some relative attributes which represents different computer configurations.

The classic way of dealing with continuous prediction is to write the outcome as a linear sum of the attribute values will appropriate weights, for example:

$$Y = B_0 + B_1*X1 + B_2*X2 + B_3*X3 + B_4*X4$$

Assumption: We assume that there exist a linear relation between dependent variable CPU Performance and independent variables such as Cycle Time, Main Memory, Cache, etc.

Expected Outcome:

- Create the CPU Database.
- Find the linear relation between CPU performance and various other attributes.
- Predict the CPU Performance using the found relation.
- Compare the result with actual values.
- Create Visualizations based on regression analysis.

Database

SQL Queries

```
SET SQL_MODE = "NO_AUTO_VALUE_ON_ZERO";
SET time_zone = "+00:00";

/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET
@OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8 */;

--
-- Database: `cpu`
--

-- --------------------------------------------------------

--
-- Table structure for table `performance`
--

CREATE TABLE IF NOT EXISTS `performance` (
  `sr` int(3) NOT NULL AUTO_INCREMENT,
  `vendor` varchar(12) DEFAULT NULL,
  `myct` int(4) DEFAULT NULL,
  `mmin` int(5) DEFAULT NULL,
  `mmax` int(5) DEFAULT NULL,
  `cach` int(3) DEFAULT NULL,
  `chmin` int(2) DEFAULT NULL,
  `chmax` int(3) DEFAULT NULL,
  `performance` int(4) DEFAULT NULL,
  `prp` decimal(9,6) DEFAULT NULL,
  PRIMARY KEY (`sr`)
) ENGINE=InnoDB  DEFAULT CHARSET=utf8 AUTO_INCREMENT=210 ;

--
-- Dumping data for table `performance`
--
```

INSERT INTO `performance` (`sr`, `vendor`, `myct`, `mmin`, `mmax`, `cach`, `chmin`, `chmax`, `performance`, `prp`) VALUES
(1, 'adviser', 125, 256, 6000, 256, 16, 128, 199, '262.542381'),
(2, 'amdahl', 29, 8000, 32000, 32, 8, 32, 253, '313.649173'),
(3, 'amdahl', 29, 8000, 32000, 32, 8, 32, 253, '313.649173'),
(4, 'amdahl', 29, 8000, 32000, 32, 8, 32, 253, '313.649173'),
(5, 'amdahl', 29, 8000, 16000, 32, 8, 16, 132, '188.990469'),
(6, 'amdahl', 26, 8000, 32000, 64, 8, 32, 290, '329.274266'),
(7, 'amdahl', 23, 16000, 32000, 64, 16, 32, 381, '442.153687'),
(8, 'amdahl', 23, 16000, 32000, 64, 16, 32, 381, '442.153687'),
(9, 'amdahl', 23, 16000, 64000, 64, 16, 32, 749, '653.033687'),
(10, 'amdahl', 23, 32000, 64000, 128, 32, 64, 1238, '949.271655'),
(11, 'apollo', 400, 1000, 3000, 0, 1, 2, 23, '-3.789983'),
(12, 'apollo', 400, 512, 3500, 4, 1, 6, 24, '-0.694251'),
(13, 'basf', 60, 2000, 8000, 65, 1, 8, 70, '60.387711'),
(14, 'basf', 50, 4000, 16000, 65, 1, 8, 117, '141.062101'),
(15, 'bti', 350, 64, 64, 0, 1, 4, 15, '-37.425287'),
(16, 'bti', 200, 512, 16000, 0, 4, 32, 64, '97.224060'),
(17, 'burroughs', 167, 524, 2000, 8, 4, 15, 23, '-13.505098'),
(18, 'burroughs', 143, 512, 5000, 0, 7, 32, 29, '20.457272'),
(19, 'burroughs', 143, 1000, 2000, 0, 5, 16, 22, '-11.204942'),
(20, 'burroughs', 110, 5000, 5000, 142, 8, 64, 124, '190.969902'),
(21, 'burroughs', 143, 1500, 6300, 0, 5, 32, 35, '43.504262'),
(22, 'burroughs', 143, 3100, 6200, 0, 5, 20, 39, '51.322434'),
(23, 'burroughs', 143, 2300, 6200, 0, 6, 64, 40, '92.555933'),
(24, 'burroughs', 110, 3100, 6200, 0, 6, 64, 45, '101.824820'),
(25, 'c.r.d', 320, 128, 6000, 0, 1, 12, 28, '10.239123'),
(26, 'c.r.d', 320, 512, 2000, 4, 1, 3, 21, '-19.459638'),
(27, 'c.r.d', 320, 256, 6000, 0, 1, 6, 28, '4.863405'),
(28, 'c.r.d', 320, 256, 3000, 4, 1, 3, 22, '-16.532230'),
(29, 'c.r.d', 320, 512, 5000, 4, 1, 5, 28, '2.712700'),
(30, 'c.r.d', 320, 256, 5000, 4, 1, 6, 27, '0.251277'),
(31, 'cdc', 25, 1310, 2620, 131, 12, 24, 102, '62.710931'),
(32, 'cdc', 25, 1310, 2620, 131, 12, 24, 102, '62.710931'),
(33, 'cdc', 50, 2620, 10480, 30, 12, 24, 74, '84.958258'),
(34, 'cdc', 50, 2620, 10480, 30, 12, 24, 74, '84.958258'),
(35, 'cdc', 56, 5240, 20970, 30, 12, 24, 138, '191.967464'),
(36, 'cdc', 64, 5240, 20970, 30, 12, 24, 136, '192.495152'),
(37, 'cdc', 50, 500, 2000, 8, 1, 4, 23, '-34.261751'),
(38, 'cdc', 50, 1000, 4000, 8, 1, 5, 29, '-12.727082'),

(39, 'cdc', 50, 2000, 8000, 8, 1, 5, 44, '27.939918'),
(40, 'cambex', 50, 1000, 4000, 8, 3, 5, 30, '-13.071756'),
(41, 'cambex', 50, 1000, 8000, 8, 3, 5, 41, '13.288244'),
(42, 'cambex', 50, 2000, 16000, 8, 3, 5, 74, '80.315244'),
(43, 'cambex', 50, 2000, 16000, 8, 3, 6, 74, '81.516413'),
(44, 'cambex', 50, 2000, 16000, 8, 3, 6, 74, '81.516413'),
(45, 'dec', 133, 1000, 12000, 9, 3, 12, 54, '54.025658'),
(46, 'dec', 133, 1000, 8000, 9, 3, 12, 41, '27.665658'),
(47, 'dec', 810, 512, 512, 8, 1, 1, 18, '2.630866'),
(48, 'dec', 810, 1000, 5000, 0, 1, 1, 28, '35.232858'),
(49, 'dec', 320, 512, 8000, 4, 1, 5, 36, '22.482700'),
(50, 'dec', 200, 512, 8000, 8, 1, 8, 38, '20.148759'),
.
.
.
.
.
.
.
.

/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;

Structure

#	Name	Type	Collation	Attributes	Null	Default	Extra
1	sr	int(3)			No	*None*	AUTO_INCREMENT
2	vendor	varchar(12)	utf8_general_ci		Yes	*NULL*	
3	myct	int(4)			Yes	*NULL*	
4	mmin	int(5)			Yes	*NULL*	
5	mmax	int(5)			Yes	*NULL*	
6	cach	int(3)			Yes	*NULL*	
7	chmin	int(2)			Yes	*NULL*	
8	chmax	int(3)			Yes	*NULL*	
9	performance	int(4)			Yes	*NULL*	
10	prp	decimal(9,6)			Yes	*NULL*	

SQL Output

sr	vendor	myct	mmin	mmax	cach	chmin	chmax	performance	prp
1	adviser	125	256	6000	256	16	128	199	262.542381
2	amdahl	29	8000	32000	32	8	32	253	313.649173
3	amdahl	29	8000	32000	32	8	32	253	313.649173
4	amdahl	29	8000	32000	32	8	32	253	313.649173
5	amdahl	29	8000	16000	32	8	16	132	188.990469
6	amdahl	26	8000	32000	64	8	32	290	329.274266
7	amdahl	23	16000	32000	64	16	32	381	442.153687
8	amdahl	23	16000	32000	64	16	32	381	442.153687
9	amdahl	23	16000	64000	64	16	32	749	653.033687
10	amdahl	23	32000	64000	128	32	64	1238	949.271655
11	apollo	400	1000	3000	0	1	2	23	-3.789983
12	apollo	400	512	3500	4	1	6	24	-0.694251
13	basf	60	2000	8000	65	1	8	70	60.387711
14	basf	50	4000	16000	65	1	8	117	141.062101
15	bti	350	64	64	0	1	4	15	-37.425287
16	bti	200	512	16000	0	4	32	64	97.224060
17	burroughs	167	524	2000	8	4	15	23	-13.505098
18	burroughs	143	512	5000	0	7	32	29	20.457272
19	burroughs	143	1000	2000	0	5	16	22	-11.204942
20	burroughs	110	5000	5000	142	8	64	124	190.969902
21	burroughs	143	1500	6300	0	5	32	35	43.504262
22	burroughs	143	3100	6200	0	5	20	39	51.322434
23	burroughs	143	2300	6200	0	6	64	40	92.555933
24	burroughs	110	3100	6200	0	6	64	45	101.824820
25	c.r.d	320	128	6000	0	1	12	28	10.239123
26	c.r.d	320	512	2000	4	1	3	21	-19.459638
27	c.r.d	320	256	6000	0	1	6	28	4.863405

sr	vendor	myct	mmin	mmax	cach	chmin	chmax	performance	prp
28	c.r.d	320	256	3000	4	1	3	22	-16.532230
29	c.r.d	320	512	5000	4	1	5	28	2.712700
30	c.r.d	320	256	5000	4	1	6	27	0.251277
31	cdc	25	1310	2620	131	12	24	102	62.710931
32	cdc	25	1310	2620	131	12	24	102	62.710931
33	cdc	50	2620	10480	30	12	24	74	84.958258
34	cdc	50	2620	10480	30	12	24	74	84.958258
35	cdc	56	5240	20970	30	12	24	138	191.967464
36	cdc	64	5240	20970	30	12	24	136	192.495152
37	cdc	50	500	2000	8	1	4	23	-34.261751
38	cdc	50	1000	4000	8	1	5	29	-12.727082
39	cdc	50	2000	8000	8	1	5	44	27.939918
40	cambex	50	1000	4000	8	3	5	30	-13.071756
41	cambex	50	1000	8000	8	3	5	41	13.288244
42	cambex	50	2000	16000	8	3	5	74	80.315244
43	cambex	50	2000	16000	8	3	6	74	81.516413
44	cambex	50	2000	16000	8	3	6	74	81.516413
45	dec	133	1000	12000	9	3	12	54	54.025658
46	dec	133	1000	8000	9	3	12	41	27.665658
47	dec	810	512	512	8	1	1	18	2.630866
48	dec	810	1000	5000	0	1	1	28	35.232858
49	dec	320	512	8000	4	1	5	36	22.482700
50	Dec	200	512	8000	8	1	8	38	20.148759

.

.

.

.

NOTE: Some Rows Skipped From Output.

Regression Analysis

In statistical modeling, regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors').

In our problem, dependent variable (that we want to predict) is Performance (PMP) and independent variables (predicates) are Cycle Time (MYCT), Main Memory Minimum (MMIN), Main Memory Maximum (MMAX), Cache (CACH), Channels Minimum (CHMIN) and Channels Maximum (CHMAX). First of all, let us write the general linear relation between these variables as follows:

$$PMP = B_0 + B_1*MYCT + B_2*MMIN + B_3*MMAX + B_4*CACH + B_5*CHMIN + B_6*CHMAX$$

Here, B_0 is the intercept and B_1, B_2, etc. are coefficients of predicates. We need to find the values of intercept and coefficients. We are going to use Microsoft Excel for performing the regression analysis on our data. During the analysis choose dependent variable as the Performance and all other numeric columns as dependent variable.

The results of the analysis are as following:

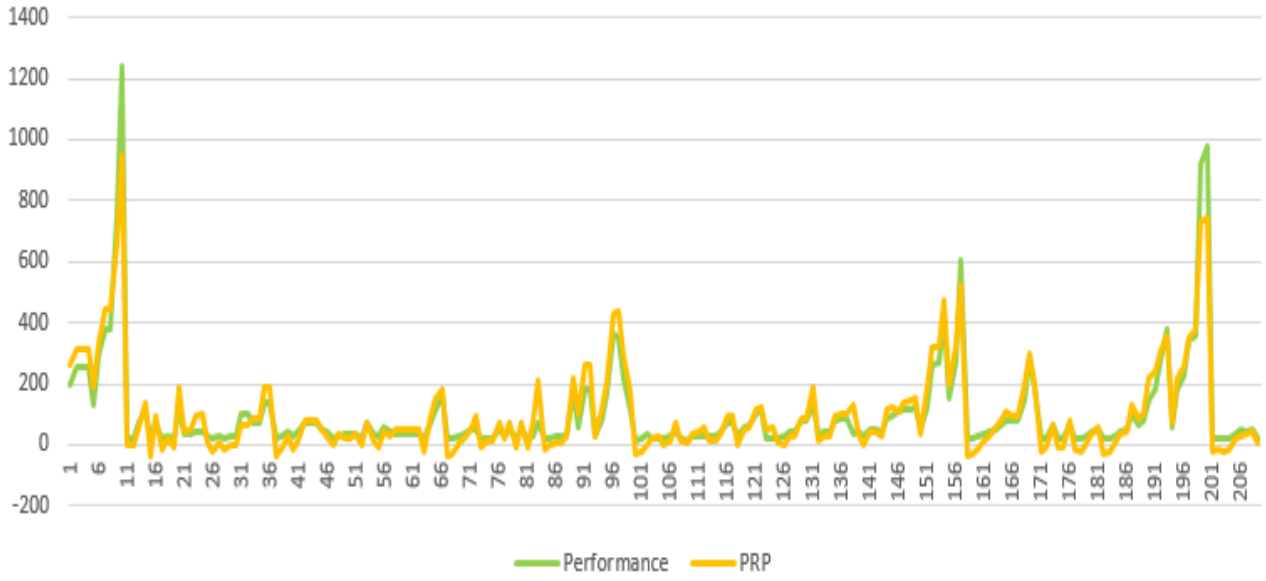
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.954394769							
R Square	0.910869375							
Adjusted R Square	0.90822193							
Standard Error	46.88350245							
Observations	209							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regression	6	4537541.534	756256.9223	344.0561033	3.9E-103			
Residual	202	444008.686	2198.062802					
Total	208	4981550.22						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-66.4813843	6.287510725	-10.57356198	4.442E-21	-78.879	-54.0838	-78.879	-54.0838
X Variable 1	0.065961068	0.01369191	4.817521161	2.84924E-06	0.038964	0.092958	0.038964	0.092958
X Variable 2	0.014307068	0.001427731	10.02084622	1.91934E-19	0.011492	0.017122	0.011492	0.017122
X Variable 3	0.006590128	0.000501599	13.13822924	6.65484E-29	0.005601	0.007579	0.005601	0.007579
X Variable 4	0.494468141	0.109073389	4.533352672	9.93561E-06	0.2794	0.709537	0.2794	0.709537
X Variable 5	-0.17233733	0.668742492	-0.25770358	0.796898025	-1.49095	1.146274	-1.49095	1.146274
X Variable 6	1.201169127	0.171969159	6.984793855	4.04051E-11	0.862084	1.540254	0.862084	1.540254

Therefore our final linear regression equation becomes:

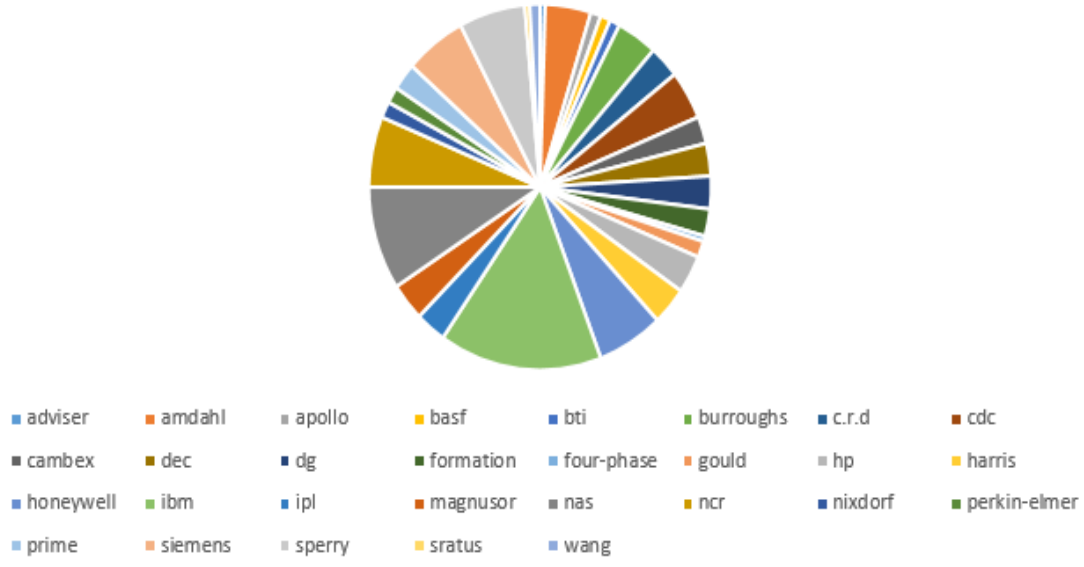
PMP = **-66.4813** + **0.0659***MYCT + **0.0143***MMIN + **0.0065***MMAX + **0.4944***CACH - **0.1723***CHMIN + **1.2011***CHMAX

Visualizations

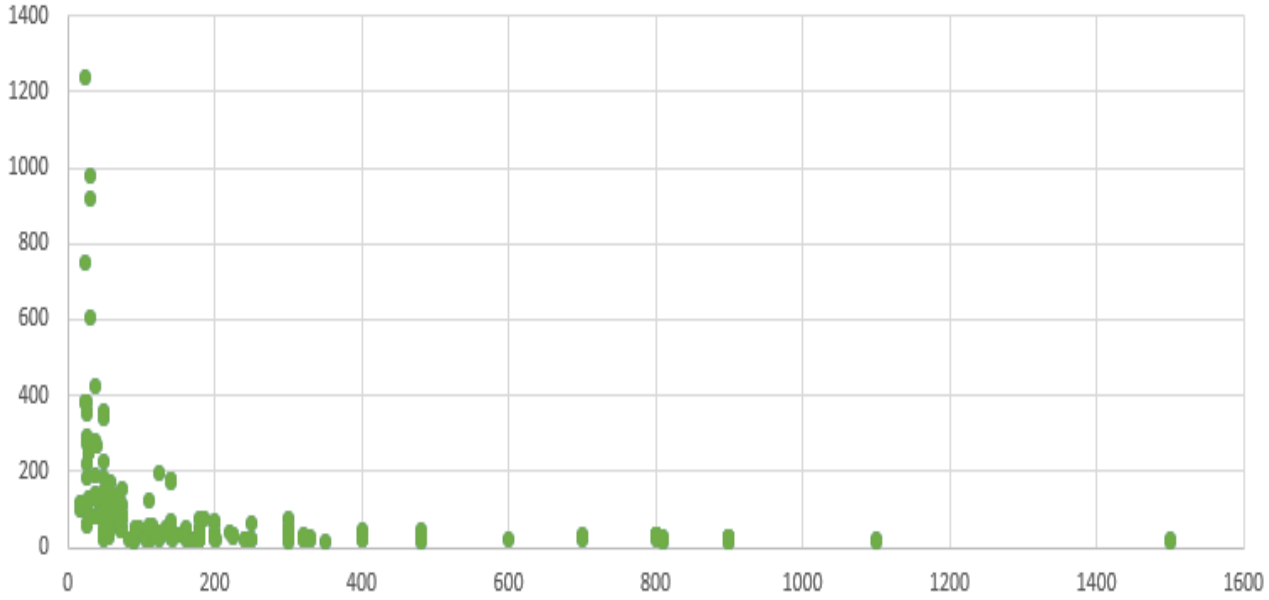
- **CPU Performance vs Calculated Performance**



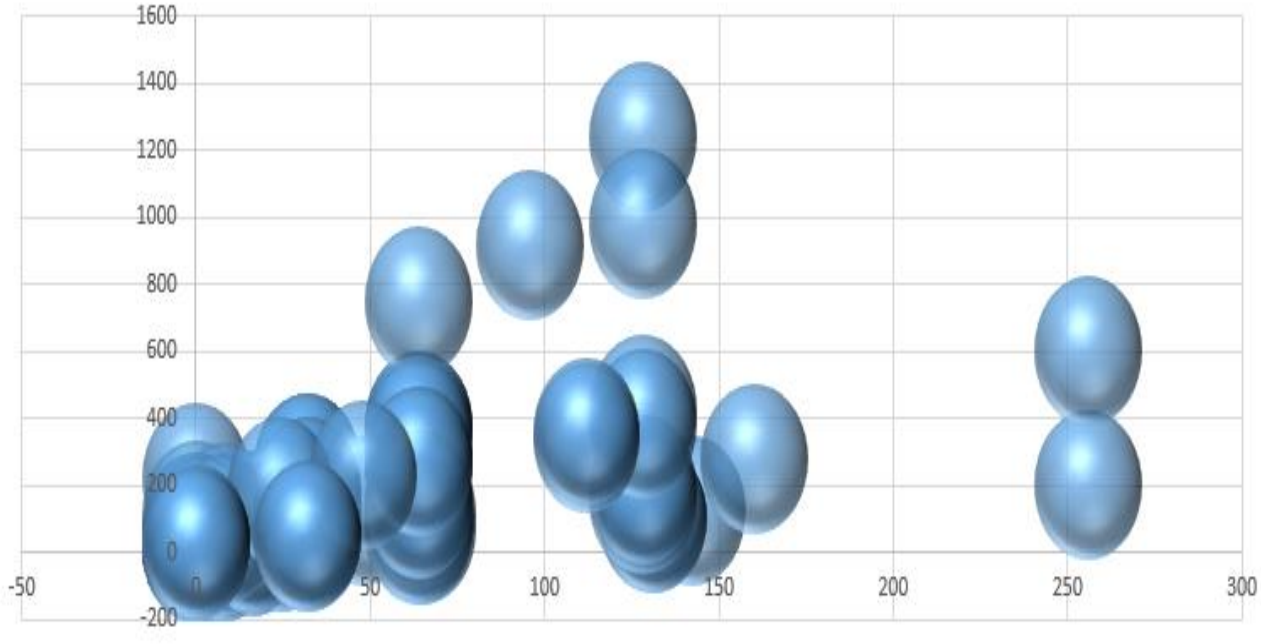
- **Market share of Vendor**



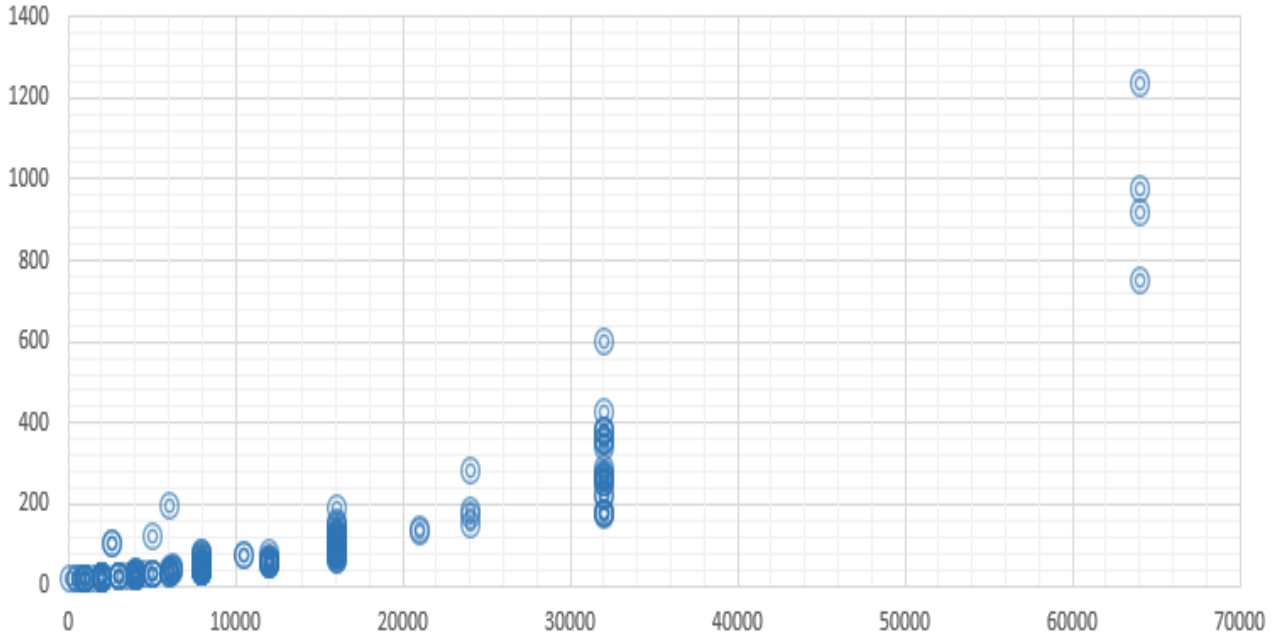
- **Cycle time vs CPU performance**



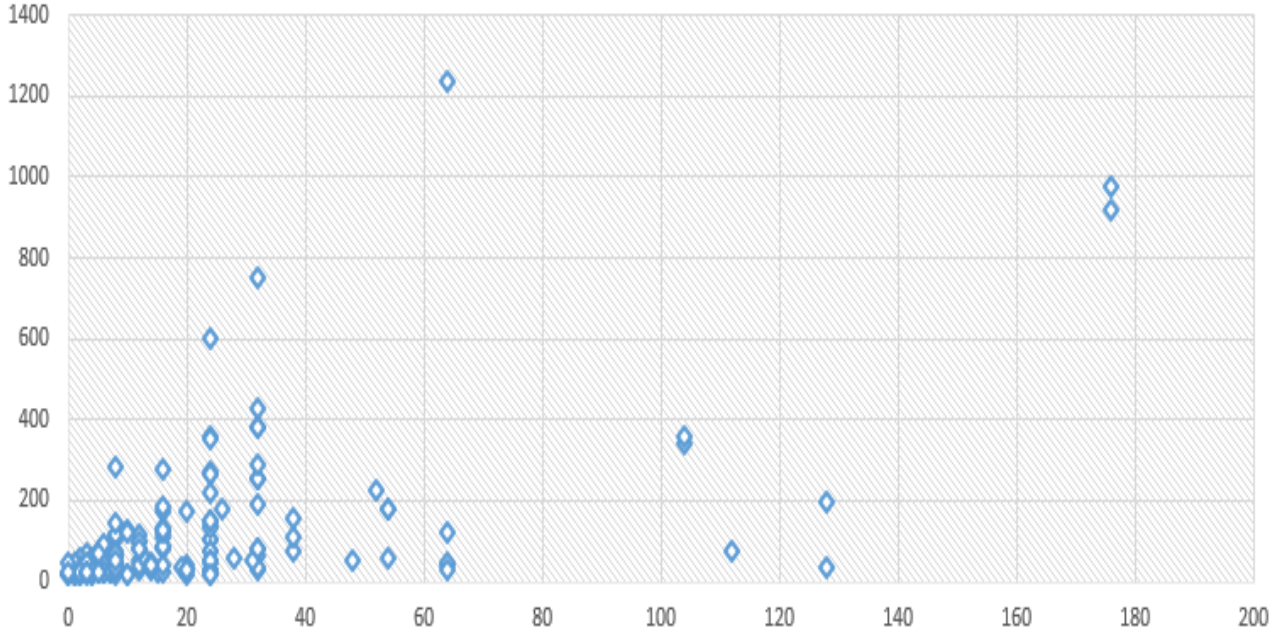
- **Cach Memory vs CPU Performance**



- **Maximum Main Memory vs CPU Performance**



- **Maximum Available Channels vs CPU Performance**



References

Software

- Pentaho: Data Integration, Business Analytics and Big Data
- PHP Server Side Scripting Language
- MySQL RDBMS
- Apache Http Server
- JQuery
- Bootstrap

Websites

- http://cpu20performance.mybluemix.net/ **(Project Site)**
- http://www.technologyforge.net/Datasets/
- http://cameron.econ.ucdavis.edu/excel/ex61multipleregression.html/
- http://community.pentaho.com/ctools/knowledge/cde-tutorial/
- https://console.ng.bluemix.net/home/

Project Guide

A. Saai Sanjeev Achaarya

Date: