

Experimental Study of an IP Model for the View Selection Problem

by Zohreh Asgharzadeh Talebi, Rada Chirkova, and Yahya Fathi

NC State University, 2005

In this document we present an experimental study of an IP model for the view selection problem; the IP model was introduced in “A Formal Model for the Problem of View Selection for Aggregate Queries,” by J. Li, Z. Asgharzadeh Talebi, R. Chirkova, and Y. Fathi, published in the proceedings of the 9th East-European Conference on Advances in Databases and Information Systems (ADBIS), 2005; the paper is available at <http://www4.ncsu.edu/~rychirko/Papers/LiACF05.pdf>

Experiments and Observations:

All the bellow experiments have been done on the lattice of 7 attributes where sizes are in number of rows except when mentioned. Also when not mentioned, the storage limits were set as the size of the top view plus half of the summation of the sizes of input queries.

In order to compare HHRU and IP algorithms in solving the view selection problem, first we chose the input queries randomly and each time we increased the number of input queries and compared the results of the algorithms both from the quality(cost value) point of view and execution time. The answers were compared with the LP lower bound too. The results are summarized in the next two pages.

As we can see, in almost all cases, the execution time when solving with IP method is much less in comparison to the one when HHRU algorithm is used. Yet, the execution time for solving the problem with IP and LP are almost the same. (When We mention LP we mean that we relaxed the constraint that $x[i]$ should be 0 or 1 to $0 \leq x[i] \leq 1$.)

	#input queries	#candidate views for being materialized	time needed to generate AMPL input(S)	time needed to solve IP	time needed to solve IPR	time needed to solve with HHRU	IP Answer	IPR Answer	HHRU Answer	IP/IPR	HHRU/IPR	HHRU/IP
1	2	64	<1	0.03	0.03	40	300314	300310	300314	1.000013	1.000013	1
2	3	58	<1	0.02	0.02	<1	565830	565812	565830	1.000032	1.000032	1
3	4	68	<1	0.02	0.03	4	405052	394099	405052	1.027793	1.027793	1
4	5	64	<1	0.04	0.02	1	925292	925155	925292	1.000148	1.000148	1
5	6	52	<1	0.03	0.05	1	1469930	1469930	1469934	1	1.000003	1.000003
6	7	51	<1	0.04	0.02	1	1596748	1596710	1596748	1.000024	1.000024	1
7	8	70	<1	0.03	0.02	2	942652	942651	942652	1.000001	1.000001	1
8	9	70	<1	0.09	0.01	11	1987860	1987640	1987864	1.000111	1.000113	1.000002
9	10	69	<1	0.07	0.04	2	2043735	2043470	2043735	1.00013	1.00013	1
10	12	58	<1	0.06	0.07	1	3242040	3242020	3242040	1.000006	1.000006	1
11	15	73	<1	0.03	0.04	1	3875900	3875660	3875903	1.000062	1.000063	1.000001
12	20	76	<1	0.07	0.04	1	4994736	4994520	4994736	1.000043	1.000043	1
13	30	109	1	0.11	0.07	3	6850848	6850790	6850848	1.000008	1.000008	1
14	50	112	1	13	0.11	117	11242300	11242300	11242304	1	1	1
15	60	115	1	0.1	0.09	5	13001276	13001276	13001276	1	1	1
16	70	117	1	0.16	0.1	5	15267900	15267900	15267919	1	1.000001	1.000001
17	100	121	1	0.22	11	5363	21006900	21006900	21006947	1	0.000002	0.000002
18	127	127	1	0.23	0.13	59188	27191700	27191690	27191691	1	1	1

Input queries for the above instances:

Instance	Input Queries	Storage Space
1	4, 101	449967
2	40 68 89	582556
3	10 48 57 96	484095
4	5 21 24 99 104	762306
5	34 37 41 55 69 95	1034768
6	23 24 49 76 108 118 122	1098122
7	28 40 44 60 65 95 96 109	771138
8	51 58 64 92 112 113 115 121 124	1293558
9	28 30 36 40 47 70 74 89 100 113	1321216
10	19 23 30 35 51 68 84 85 88 94 115 124	1920803
11	5 9 13 22 30 31 43 47 62 72 92 100 105 117 123	2237482
12	10 17 21 23 25 26 43 44 58 69 70 76 83 84 89 90 100 110 115 126	2796908
13	1 4 5 11 14 15 18 26 31 37 38 45 46 48 50 51 60 69 70 71 79 84 86 91 92 94 97 106 113 118	3725130
14	5 7 12 13 16 18 21 24 26 33 34 36 38 39 40 42 46 53 55 56 60 62 63 68 69 72 76 77 78 79 82 86 90 92 93 95 99 102 105 106 107 108 109 113 116 117 118 121 124 125	5920880
15	2 3 6 7 10 11 12 15 16 17 20 25 28 30 31 33 34 36 38 39 40 44 46 47 49 51 53 60 61 63 66 71 78 79 83 84 85 86 93 94 96 97 98 101 102 105 106 109 111 112 113 115 116 117 118 119 121 123 124 126	6800427
16	3 4 10 11 13 16 18 19 20 23 26 27 28 29 30 34 36 37 38 40 42 43 45 47 49 50 52 53 56 57 58 60 61 62 63 70 71 72 74 76 77 78 80 81 84 86 87 88 89 91 92 93 94 97 102 105 107 108 109 113 114 116 117 118 119 120 121 124 125 126	7933740
17	1 2 4 6 7 8 9 10 12 13 15 16 18 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 38 39 40 41 42 44 45 46 47 48 50 51 53 55 58 59 60 61 62 63 64 65 67 68 69 71 72 73 74 75 76 77 78 79 80 81 83 84 85 88 89 90 91 93 94 95 96 97 98 99 100 102 103 104 106 107 108 109 110 112 113 114 115 116 117 119 120 122 123 125 126	10803261
18	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127	13895635

We also saw that the execution time of HHRU algorithm does not necessarily increase as the number of input queries increases. We guessed that this is because of the random way of choosing the input queries. As a result, we did another set of experiments in which, instead of generating new set of input queries for each instance, we kept the previous input queries and just add some more to them. We can see in the following instances that again, the execution time of the HHRU algorithm is not necessarily increasing, as the number of input queries increase. For example, in the case of 100 input queries, the execution time is just 7 seconds, however this time is 704 seconds when the number of input queries is 70 and 2608 when the number of input queries is 127.

This should not be surprising! The reason is that in the case of 100 input queries, the storage limit is filled very fast because big size queries are picked in the first iterations of HHRU algorithm.

instance#	1	2	3	4	5	6	7	8	9
#input queries	2	3	4	5	6	7	8	9	10
#candidate views for being materialized	20	42	42	44	53	58	58	58	58
execution time of generating AMPL input (S)	0	0	0	0	0	0	0	0	0
execution time of IP(S)	0.08	0.06	0.05	0.04	0.03	0.04	0.02	0.07	0.04
execution time of IPR(S)	0.05	0.03	0.03	0.03	0.05	0.04	0.02	0.06	0.03
execution time of HHRU(S)	1	7	1	0	0	1	1	1	1
IP Answer	599275	605128	904589	1204403	1308945	1340597	1637139	1934831	2234292
IPR Answer	599275	604778	904585.1	1204395	1308938	1340590	1637137	1934558	2234260
HHRU Answer	599275	605128	904589	1204403	1308945	1340597	1637139	1934831	2234292
IP/IPR	1	1.000579	1.000004	1.000007	1.000005	1.000005	1.000001	1.000141	1.000014
HHRU/IPR	1	1.000579	1.000004	1.000007	1.000005	1.000005	1.000001	1.000141	1.000014
stlimit	599447	602197	752104	902011	954282	970108	1118379	1267048	1416898
IP used stlimit	599275	303627	604775	604775	709317	740969	1037511	1035389	1334850
HHRU used stlimit	599275	303627	604775	604775	709317	740969	1037511	1035389	1334850
B&B nodes	0	0	0	0	0	0	0	5	0
Dual simplex	13	18	16	17	19	21	24	29	30
MIP simplex iteration	3	20	16	17	19	21	24	37	30

continue:

instance#	10	11	12	13	14	15	16	17	18
#input queries	12	15	20	30	50	60	70	100	127
#candidate views for being materialized	64	67	93	98	115	120	122	125	127
Time needed to generate AMPL input (S)	0	0	0	1	0	0	1	0	0
Time needed to solve with IP(S)	0.03	0.05	0.06	0.09	0.1	0.14	0.19	0.2	0.23
Time needed to solve with IPR(S)	0.05	0.05	0.05	0.03	0.09	0.11	0.07	0.18	0.13
Time needed to solve with HHRU(S)	1	1	2	3	877	287	704	7	59188
IP Answer	2800100	3319357	4518417	6511499	10822208	12478218	15142765	21511463	27191691
IPR Answer	2800062	3319304	4518413	6511498	10822180	12478210	15142744	21511461	27191690
HHRU Answer	2800100	3319357	4518417	6511499	10822213	12478223	15142772	21511468	27191691
IP/IPR	1.000013	1.000016	1.000001	1	1.000003	1.000001	1.000001	1	1
HHRU/IPR	1.000013	1.000016	1.000001	1	1.000003	1.000001	1.000002	1	1.001678
stlimit	1699781	1959408	2559013	3555552	5710869	6538886	7871153	11055518	13895635
IP used stlimit	1600844	1820287	2419761	3513401	5427100	6439776	7609233	10955068	13540885
HHRU used stlimit	1600844	1820287	2419761	3513401	5427105	6439781	7609240	10656125	13704118
B&B nodes	0	0	0	0	0	0	0	0	0
Dual simplex	39	47	51	69	131	147	168	207	229
MIP simplex iteration	40	52	51	69	134	148	165	204	229

The index of input queries for the above instances are chosen in the following order:

{69, 86, 12, 95, 91, 49, 50, 76, 124, 122, 68, 51, 99, 52, 14, 32, 19, 111, 43, 119, 34, 107, 78, 120, 121, 18, 123, 15, 97, 54, 96, 5, 118, 7, 33, 13, 104, 83, 37, 87, 21, 67, 1, 77, 30, 106, 60, 23, 41, 24, 58, 112, 115, 42, 17, 64, 9, 88, 31, 79, 8, 108, 84, 105, 45, 73, 113, 74, 39, 29, 38, 27, 55, 61, 10, 114, 53, 103, 26, 66, 125, 11, 70, 72, 63, 20, 36, 109, 4, 102, 89, 82, 93, 40, 110, 98, 28, 80, 25, 81, 94, 71, 62, 56, 65, 47, 44, 48, 90, 3, 116, 101, 6, 126, 75, 92, 2, 117, 59, 100, 16, 57, 35, 22, 85, 46}

For example, for the 10th instance which has 12 input queries, the input queries are the first 12 index of the above set: 69, 86, 12, 95, 91, 49, 50, 76, 124, 122, 68, 51

In the following table, the details for each instance is shown:

instance#1	index of input queries: 69 86
selected views with IP	127 86
selected views with HHRU	127 86
instance#2	index of input queries: 69 86 12
selected views with IP	127 127 12
selected views with HHRU	127 127 12
instance#3	index of input queries: 69 86 12 95
selected views with IP	127 86 12 127
selected views with HHRU	127 86 12 127
instance#4	index of input queries: 69 86 12 95 91
selected views with IP	127 86 12 127 127
selected views with HHRU	127 86 12 127 127
instance#5	index of input queries: 69 86 12 95 91 49
selected views with IP	127 86 12 127 127 49
selected views with HHRU	127 86 12 127 127 49
instance#6	index of input queries: 69 86 12 95 91 49 50
selected views with IP	127 118 12 127 127 49 50
selected views with HHRU	127 86 12 127 127 49 50
instance#7	index of input queries: 69 86 12 95 91 49 50 76
selected views with IP	127 118 12 127 127 49 50 76
selected views with HHRU	127 86 12 127 127 49 50 76
instance#8	index of input queries: 69 86 12 95 91 49 50 76 124
selected views with IP	127 127 12 127 127 49 50 76 124
selected views with HHRU	127 127 12 127 127 49 50 76 124

continue:

instance#9	index of input queries: 69 86 12 95 91 49 50 76 124 122
selected views with IP	127 118 12 127 127 49 50 76 124 127
selected views with HHRU	127 86 12 127 127 49 50 76 124 127
instance#10	index of input queries: 69 86 12 95 91 49 50 76 124 122 68 51
selected views with IP	127 118 12 127 127 49 50 76 124 127 100 127
selected views with HHRU	127 86 12 127 127 49 50 76 124 127 68 127
instance#11	index of input queries: 69 86 12 95 91 49 50 76 124 122 68 51 99 52 14
selected views with IP	127 118 12 127 127 49 50 76 124 127 100 127 127 52 14
selected views with HHRU	127 86 12 127 127 49 50 76 124 127 68 127 127 52 14
instance#12	index of input queries: 69 86 12 95 91 49 50 76 124 122 68 51 99 52 14 32 19 111 43 119
selected views with IP	127 118 12 127 127 49 50 108 124 122 68 51 127 52 14 32 51 127 127 127
selected views with HHRU	127 86 12 127 127 49 50 76 124 122 68 51 127 52 14 32 51 127 127 127
instance#13	index of input queries:69 86 12 95 91 49 50 76 124 122 68 51 99 52 14 32 19 111 43 119 34 107 78 120 121 18 123 15 97 54
selected views with IP	127 118 12 127 127 49 50 108 124 122 68 51 127 52 14 32 51 127 127 127 34 127 110 120 121 18 127 127 97 54
selected views with HHRU	127 86 12 127 127 49 50 76 124 122 68 51 127 52 14 32 51 127 127 127 34 127 78 120 121 18 127 127 97 54
instance#14	index of input queries: 69 86 12 95 91 49 50 76 124 122 68 51 99 52 14 32 19 111 43 119 34 107 78 120 121 18 123 15 97 54 96 5 118 7 33 13 104 83 37 87 21 67 1 77 30 106 60 23 41 24
selected views with IP	127 118 12 127 127 49 50 127 124 122 100 51 127 52 14 32 51 127 127 127 34 127 110 120 121 18 127 127 97 54 96 37 118 127 33 13 104 127 37 127 21 127 1 108 30 106 60 127 41 24
selected views with HHRU	127 118 12 127 127 49 50 76 124 122 68 51 127 52 14 32 51 127 127 127 34 127 78 120 121 18 127 127 97 54 96 37 118 127 33 13 104 127 37 127 53 127 1 127 30 106 60 127 41 24

continue:

instance#15	index of input queries: 69 86 12 95 91 49 50 76 124 122 68 51 99 52 14 32 19 111 43 119 34 107 78 120 121 18 123 15 97 54 96 5 118 7 33 13 104 83 37 87 21 67 1 77 30 106 60 23 41 24 58 112 115 42 17 64 9 88 31 79
selected views with IP	127 118 12 127 127 49 50 108 124 122 100 51 127 52 14 32 51 127 127 127 34 127 110 120 121 18 127 127 97 54 96 37 118 127 33 13 104 127 37 127 21 127 1 127 30 106 60 55 41 24 58 112 127 42 17 96 9 120 127 127
selected views with HHRU	127 118 12 127 127 49 50 76 124 122 68 51 127 52 14 32 51 127 127 127 34 127 78 120 121 18 127 127 97 54 96 37 118 23 33 13 104 127 37 127 53 127 1 127 30 106 60 23 41 24 58 112 127 42 17 96 9 120 127 127
instance#16	index of input queries: 69 86 12 95 91 49 50 76 124 122 68 51 99 52 14 32 19 111 43 119 34 107 78 120 121 18 123 15 97 54 96 5 118 7 33 13 104 83 37 87 21 67 1 77 30 106 60 23 41 24 58 112 115 42 17 64 9 88 31 79 8 108 84 105 45 73 113 74 39 29
selected views with IP	127 118 12 127 127 49 50 108 124 122 100 51 127 52 14 32 51 127 127 127 34 127 110 120 121 18 127 127 97 54 96 37 118 127 33 45 104 127 37 127 21 127 1 127 30 106 60 55 41 24 58 112 127 42 17 96 9 120 127 127 8 108 116 105 45 105 113 106 55 29
selected views with HHRU	127 118 12 127 127 49 50 108 124 122 68 51 127 52 14 32 51 127 127 127 34 127 78 120 121 18 127 127 97 54 96 37 118 55 33 45 104 127 37 127 53 127 1 127 30 106 60 55 41 24 58 112 127 42 17 96 9 120 127 127 8 108 84 105 45 105 113 106 55 61

continue:

instance#17	index of input queries: 69 86 12 95 91 49 50 76 124 122 68 51 99 52 14 32 19 111 43 119 34 107 78 120 121 18 123 15 97 54 96 5 118 7 33 13 104 83 37 87 21 67 1 77 30 106 60 23 41 24 58 112 115 42 17 64 9 88 31 79 8 108 84 105 45 73 113 74 39 29 38 27 55 61 10 114 53 103 26 66 125 11 70 72 63 20 36 109 4 102 89 82 93 40 110 98 28 80 25 81
selected views with IP	69 118 12 127 127 49 50 108 124 122 100 51 127 52 14 32 51 127 43 127 34 127 110 120 121 18 127 127 97 54 96 5 118 39 33 45 104 127 37 127 21 127 1 125 30 106 60 55 41 24 58 112 127 42 17 96 9 120 127 127 8 108 116 105 45 105 113 106 39 61 38 127 55 61 10 114 53 127 26 98 125 43 102 104 127 20 36 125 4 102 121 114 125 40 110 98 28 112 25 113
selected views with HHRU	69 118 12 127 127 49 50 108 124 122 68 51 127 52 14 32 51 127 43 127 34 127 110 120 121 18 127 127 97 54 96 5 118 39 33 45 104 127 37 127 53 127 1 125 30 106 60 55 41 24 58 112 127 42 17 96 9 120 127 127 8 108 84 105 45 105 113 106 39 61 38 127 55 61 10 114 53 127 26 98 125 43 102 104 127 20 36 125 4 102 121 114 125 40 110 98 28 112 25 113
instance#18	index of input queries: 69 86 12 95 91 49 50 76 124 122 68 51 99 52 14 32 19 111 43 119 34 107 78 120 121 18 123 15 97 54 96 5 118 7 33 13 104 83 37 87 21 67 1 77 30 106 60 23 41 24 58 112 115 42 17 64 9 88 31 79 8 108 84 105 45 73 113 74 39 29 38 27 55 61 10 114 53 103 26 66 125 11 70 72 63 20 36 109 4 102 89 82 93 40 110 98 28 80 25 81 94 71 62 56 65 47 44 48 90 3 116 101 6 126 75 92 2 117 59 100 16 57 35 22 85 46
selected views with IP	101 118 12 127 127 49 50 108 124 122 100 51 115 52 14 32 51 127 43 127 34 127 110 120 121 18 127 127 97 54 96 5 118 39 33 13 104 127 37 127 21 127 1 125 30 106 60 55 41 24 58 112 115 42 17 96 9 120 127 127 8 108 116 105 45 105 113 106 39 61 38 59 55 61 10 114 53 127 26 98 125 43 102 104 127 20 36 125 4 102 121 114 125 40 110 98 28 112 25 113 126 127 62 56 97 127 44 48 122 35 116 101 6 126 127 124 2 117 59 100 16 57 35 22 117 46
selected views with HHRU	101 118 12 127 127 49 50 108 124 122 100 51 115 52 14 32 51 127 43 127 34 127 110 120 121 18 127 127 97 54 96 5 118 39 33 13 104 115 37 127 21 115 1 125 30 106 60 55 41 24 58 112 115 42 17 96 9 120 127 127 8 108 116 105 45 105 113 106 39 61 38 59 55 61 10 114 53 127 26 98 125 43 102 104 127 20 36 125 4 102 121 114 125 40 110 98 28 112 25 113 126 127 62 56 97 127 44 48 122 35 116 101 6 126 127 124 2 117 59 100 16 57 35 22 117 46

In most instances, HHRU answer/IP answer ratio is close to 1, but there are some instances on real data that this ratio is very big. In the following instances, we show this.

Instances in which HHRU algorithm works weak in comparison to IP:

instance#1	Storage limit	cost	used space	queries		
	300328			4	8	16
IP Answer		514	514	4	8	16
HHRU Answer		299828	299861	127	8	16

$$\frac{HHRU_{cost}}{IP_{cost}} = 583.323$$

Selected views with HHRU : 24, 16, 8, 127 (view 24 is not used for answering any query.)

instance#2	Storage limit	cost	used space	queries		
	300327			4	8	32
IP Answer		514	300328	4	8	32
HHRU Answer		299827	299849	127	8	32

$$\frac{HHRU_{cost}}{IP_{cost}} = 584.4581$$

Selected views with HHRU : 40, 32, 8, 127 (view 40 is not used for answering any query.)

instance#3	Storage limit	cost	used space	queries		
	299858			8	16	32
IP Answer		16	299830	8	16	32
HHRU Answer		132	299858	56	56	56

$$\frac{HHRU_{cost}}{IP_{cost}} = 8.25$$

Selected views with HHRU algorithm: 56, 127 (view 127 is not used for answering any query.)

The reason of these results is that HHRU picks the common parent view first, so big amount of storage limit is filled very fast and it can not change this choice later on. (This is one of the problems of greedy algorithms.)

In the above instances, storage spaces were computed as the summation of the sizes of all input queries, plus the top view size. These instances bring a general set of problems that will have the same behavior in a sense that HHRU cost is much bigger than the optimal one.¹

¹this is discussed in the paper, mentioned at the beginning of this report

Instances with all views as input queries:

In the following instances, all the nodes in a lattice were chosen to be input queries.

3 attribute lattice		cost	time	storage limit
	IP	1024859	0.02	661734
	HHRU	1024859	0	661734
4 attribute lattice		cost	time	storage limit
	IP	2499932	0.07	1399783
	HHRU	2499932	0	1399783
5 attribute lattice		cost	time	storage limit
	IP	5273778	0.04	2786631
	HHRU	5273778	1	2786631
6 attribute lattice		cost	time	storage limit
	IP	10709709	0.09	5504695
	HHRU	10709709	1	5504695
7 attribute lattice		cost	time	storage limit
	IP	27191691	0.029	13895600
	HHRU	27191691	59188	13895600

Input queries and selected views by each algorithm for each instance are shown in the following table:

instance1 input queries	1 2 3 4 5 6 7
selected views with IP	1 2 7 4 7 6 7
selected views with HHRU	1 2 7 4 7 6 7
instance2 input queries	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
selected views with IP	1 2 15 4 5 6 15 8 9 10 15 12 15 14 15
selected views with HHRU	1 2 15 4 5 6 15 8 9 10 15 12 15 14 15
instance3 input queries	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
selected views with IP	1 1 31 4 5 6 31 8 9 10 31 12 31 14 31 16 17 18 31 20 21 22 31 24 25 26 31 28 31 30 31
selected views with HHRU	1 1 31 4 5 6 31 8 9 10 31 12 31 14 31 16 17 18 31 20 21 22 31 24 25 26 31 28 31 30 31
instance4 input queries	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63
selected views with IP	1 2 51 4 37 6 55 8 9 10 63 12 61 14 63 16 17 18 51 20 53 22 55 24 25 26 63 28 61 30 63 32 33 34 51 36 37 38 55 40 41 42 63 44 61 46 63 48 49 50 51 52 53 54 55 56 57 58 63 60 61 62 63
selected views with HHRU	1 2 51 4 37 6 55 8 9 10 63 12 61 14 63 16 17 18 51 20 53 22 55 24 25 26 63 28 61 30 63 32 33 34 51 36 37 38 55 40 41 42 63 44 61 46 63 48 49 50 51 52 53 54 55 56 57 58 63 60 61 62 63
instance5 input queries	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126
selected views with IP	1 2 3 4 5 6 39 8 9 10 11 12 13 14 127 16 17 18 51 20 21 22 55 24 25 26 27 28 61 30 127 32 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 96 65 66 127 100 101 70 71 72 105 106 75 108 125 110 127 80 81 82 127 116 85 118 127 120 89 90 127 92 93 94 127 96 97 98 115 100 101 102 103 104 105 106 127 108 109 110 127 112 113 114 115 116 117 118 127 120 121 122 127 124 125 126
selected views with HHRU	1 2 35 4 5 6 39 8 9 10 43 12 13 14 127 16 17 18 19 20 21 22 55 24 25 26 59 28 29 30 127 32 39 40 41 42 43 44 45 46 127 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 127 96 97 98 115 100 101 102 127 104 105 106 127 108 125 110 127 112 113 114 115 116 117 118 127 120 121 122 127 124 125 126 127 96 97 98 115 100 101 102 127 104 105 106 127 108 125 110 111 112 113 114 115 116 117 118 119 120 121 122 127 124 125 126

As we can see, in all cases, the IP and HHRU answers are the same.

LPRelaxed

In the following experiments, we relaxed one of the IP model constraints that force the top view to be materialized (In HHRU algorithm, top view should always be materialized).

IP1: Model that forces the top view to be materialized

IP2: Model that does not force the top view to be materialized

In most cases, when the top view is in the IP optimal answer, HHRU and IP have the same answers, but in the cases where top view is not in the IP optimal answer, the IP and HHRU answers are significantly different:

	storage limit*	used storage space	cost	queries				
	1112258			21	53	88	102	116
selected views with IP1 model		914119	1213067	53	53	88	127	116
selected views with IP2 model		913664	1212612	53	53	88	102	116
selected views with HHRU algorithm		914119	1213067	53	53	88	127	116
	storage limit*	used storage space	cost	queries				
	516123			2	10	17	22	56
selected views with IP1 model		516117	516117	2	10	17	127	56
selected views with IP2 model		322853	322853	2	10	17	22	56
selected views with HHRU algorithm		516117	516117	2	10	17	127	56

	storage limit*	used storage space	cost	queries						
	1314129			13	29	30	100	113	116	
selected views with IP1 model		1214277	1514091	127	127	38	100	113	116	
selected views with IP2 model		1400849	1513927	29	29	30	100	113	116	
selected views with HHRU algorithm		1214277	1514091	127	127	38	100	113	116	
	storage limit*	used storage space	cost	queries						
	370977			40	50	52	88	104	112	
selected views with IP1 model		364837	364837	40	50	52	127	104	112	
selected views with IP2 model		106220	106220	40	50	52	120	104	112	
selected views with HHRU algorithm		346823	388020	40	127	52	120	120	112	
	storage limit*	used storage space	cost	queries						
	378180			34	60	72	80	88	96	112
selected views with IP1 model		371773	375586	34	60	72	112	127	96	112
selected views with IP2 model		113156	116969	34	60	72	112	88	96	112
selected views with HHRU algorithm		365411	410421	127	60	120	112	120	96	112
	storage limit*	used storage space	cost	queries						
	394090			34	60	72	96	104	112	120
selected views with IP1 model		85597	167991	34	60	120	96	120	112	120
selected views with IP2 model		71959	140715	34	60	104	96	104	112	120
selected views with HHRU algorithm		85597	167991	34	60	120	96	120	112	120

	storage limit*	used storage space	cost	queries							
	405781			34	60	72	80	88	96	112	120
selected views with IP1 model		85597	171804	34	60	120	112	120	96	112	120
selected views with IP2 model		113156	158166	34	60	72	112	120	96	112	120
selected views with HHRU algorithm		85597	171804	34	60	120	112	120	96	112	120
	storage limit*	used storage space	cost	queries							
	362002			44	48	50	52	56	60	64	72
selected views with IP1 model		360985	360985	44	48	127	52	56	60	64	72
selected views with IP2 model		92823	92823	44	48	50	52	56	60	64	72
selected views with HHRU algorithm		368048	368048	60	48	127	52	56	60	64	72

: stlimit=top view size +.67(summation of the sizes af input queries)

Comparing IP and Ulman performance when there are ancestor/predecessor relationships among input queries:

In the following experiments, the input queries are chosen in a way that there is ancestor/predecessor relationships among input queries. As we can see, in almost all cases (except for one case) the IP and Ulman answers are the same:

In the following table, we can see the results of the experiments which were done over the input queries shown in figure1.

algorithm	storage limit	cost	used space	execution time(S)	Queries						
					84	99	100	111	118	119	126
IP	300000	2098698	299814	0	127	127	127	127	127	127	127
HHRU		2098698	299814	0	127	127	127	127	127	127	127
IP	599275	2047390	573974	0.03	116	127	116	127	127	127	127
HHRU		2047390	573974	0	116	127	116	127	127	127	127
IP	599628	2047390	573974	0.06	116	127	116	127	127	127	127
HHRU		2047390	573974	0	116	127	116	127	127	127	127
IP	599600	2047390	573974	0.05	116	127	116	127	127	127	127
HHRU		2047390	573974	0	116	127	116	127	127	127	127
IP	1000000	2039224	839968	0.05	116	127	100	127	127	127	127
HHRU		2039224	839968	0	116	127	100	127	127	127	127
IP	1200000	2038871	1139429	0.06	116	127	100	127	118	127	127
HHRU		2038871	1139429	1	116	127	100	127	118	127	127
IP	1500000	1439215	1139429	0.07	116	127	100	127	118	127	126
HHRU		1439215	1139429	0	116	127	100	127	118	127	126
IP	2000000	2038842	1739028	0.04	84	99	100	127	118	127	126
HHRU		2038842	1739028	1	84	99	100	127	118	127	126

Number of nodes that can answer these input queries are 28.

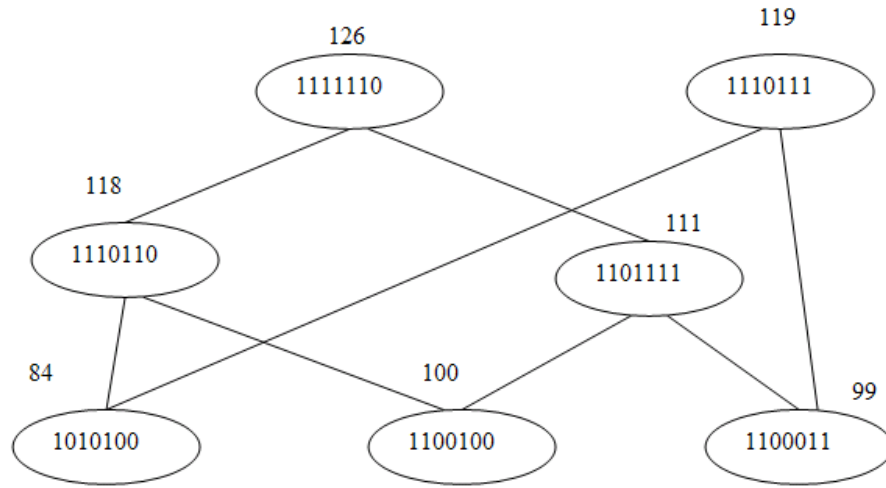


Figure 1: ancestor/predecessor relationships among input queries

In the following table, we can see the results of the experiments which were done over the input queries shown in figure2.

method	storage limit	cost	used space	execution time(S)	Queries			
					16	20	24	28
IP	300000	599664	299850	0.04	16	127	24	127
HHRU		599664	299850	7	16	127	24	127
IP	301000	599664	299850	0.04	16	127	24	127
HHRU		599664	299850	51	16	127	24	127
IP	301314	599664	299850	0.05	16	127	24	127
HHRU		599664	299850	64	16	127	24	127
IP	302000	301350	301350	0.02	16	20	24	127
HHRU		301350	301350	28	16	20	24	127
IP	316314	66000	316314	0.04	28	28	28	28
HHRU		66000	316314	1	28	28	28	28
IP	317850	18036	317850	0.04	16	20	24	28
HHRU		18036	317850	0	16	20	24	28

Number of nodes that can answer these input queries are 64.

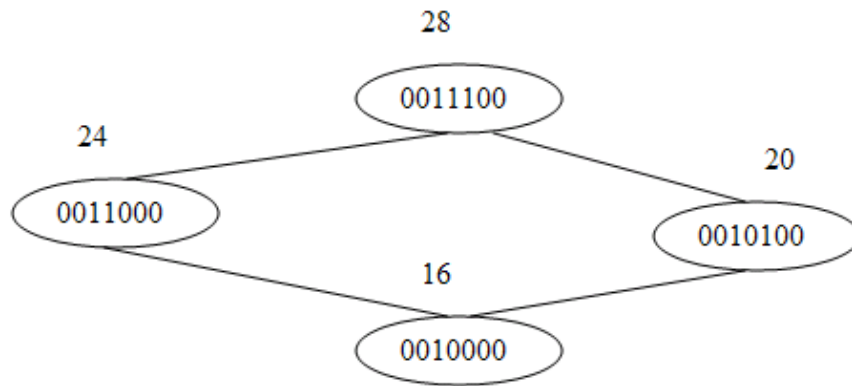


Figure 2: ancestor/predecessor relationships among input queries

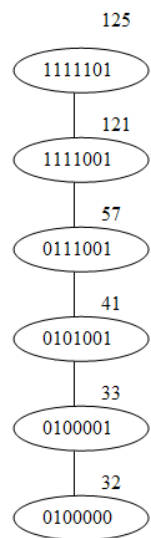


Figure 3: ancestor/predecessor relationships among input queries

In the following table, we can see the results of the experiments which were done over the input queries shown in figure3.

method	storage limit	cost	used space	execution time(S)	Queries					
					32	33	41	57	121	125
IP	300000	1499072	299816	0.04	32	127	127	127	127	127
HHRU		1499072	299816	13	32	127	127	127	127	127
IP	400000	1276185	376743	0.02	32	33	127	127	127	127
HHRU				more than 10 min						
IP	500000	1276185	376743	0.05	32	33	127	127	127	127
HHRU				more than 10 min						
IP	600000	1276185	376743	0.04	32	33	127	127	127	127
HHRU				more than 10 min						
IP	700000	1203205	640067	0.06	32	33	57	57	127	127
HHRU		1203205	640067	1	32	33	57	57	127	127
IP	8000000	1203205	640067	0.03	32	33	57	57	127	127
HHRU		1203205	640067	1	32	33	57	57	127	127
IP	900000	1191708	891894	0.02	32	33	41	57	127	127
HHRU		1191708	891894	1	32	33	41	57	127	127
IP	1000000	1191708	891894	0.06	32	33	41	57	127	127
HHRU		1191708	891894	0	32	33	41	57	127	127
IP	1100000	1191708	891894	0.04	32	33	41	57	127	127
HHRU		1191708	891894	1	32	33	41	57	127	127
IP	1200000	1191380	1191380	0.05	32	33	41	57	121	127
HHRU		1191380	1191380	0	32	33	41	57	121	127
IP	1300000	1191380	1191380	0.06	32	33	41	57	121	127
HHRU		1191380	1191380	0	32	33	41	57	121	127
IP	1400000	1191380	1191380	0.06	32	33	41	57	121	127
HHRU		1191380	1191380	0	32	33	41	57	121	127
IP	1500000	1191378	1491192	0.06	32	33	41	57	121	125
HHRU		1191378	1491192	1	32	33	41	57	121	125

Number of nodes that can answer these input queries are 64.

In the following table, we can see the results of the experiments which were done over the input queries shown in figure4.

method	storage limit	cost	used space	execution time(S)	Queries				
					105	112	117	121	125
IP	300000	1499070	299814	0	127	127	127	127	127
HHRU		1499070	299814	0	127	127	127	127	127
IP	400000	1203069	303627	0.05	127	112	127	127	127
HHRU		1203069	303627	3	127	112	127	127	127
IP	500000	1203069	303627	0.04	127	112	127	127	127
HHRU		1203069	303627	7	127	112	127	127	127
IP	600000	1203069	303627	0.04	127	112	127	127	127
HHRU		1203069	303627	11	127	112	127	127	127
IP	700000	1202413	603113	0.04	121	112	127	121	127
HHRU		1202413	603113	0	121	112	127	121	127
IP	8000000	1202413	603113	0.07	121	112	127	121	127
HHRU		1202413	603113	0	121	112	127	121	127
IP	900000	1202413	603113	0.04	121	112	127	121	127
HHRU		1202413	603113	0	121	112	127	121	127
IP	1000000	1202291	902447	0.02	105	112	127	121	127
HHRU		1202291	902447	1	105	112	127	121	127
IP	1100000	1202291	902447	0.04	105	112	127	121	127
HHRU		1202291	902447	1	105	112	127	121	127
IP	1200000	1202291	902447	0.03	105	112	127	121	127
HHRU		1202291	902447	1	105	112	127	121	127
IP	1300000	1202285	1202285	0.03	105	112	117	121	127
HHRU		1202285	1202285	1	105	112	117	121	127
IP	1400000	1202285	1202285	0.05	105	112	117	121	127
HHRU		1202285	1202285	1	105	112	117	121	127
IP	1500000	1202285	1202285	0.04	105	112	117	121	127
HHRU		1202285	1202285	1	105	112	117	121	127

Number of nodes that can answer these input queries are 20.

In the following table, we can see the results of the experiments which were done over the input queries shown in figure5.

method	storage limit	cost	used space	execution time(S)	Queries		
					64	66	124
IP	300000	899442	299814	0	127	127	127
HHRU		899442	299814	0	127	127	127
IP	400000	602152	302338	0.02	64	127	127
HHRU		602152	302338	5	64	127	127
IP	500000	602152	302338	0.01	64	127	127
HHRU		602152	302338	10	64	127	127
<i>IP</i>	<i>600000</i>	<i>600314</i>	<i>600314</i>	<i>0.02</i>	<i>64</i>	<i>66</i>	<i>127</i>
<i>HHRU</i>		<i>602152</i>	<i>302338</i>	<i>16</i>	<i>64</i>	<i>127</i>	<i>127</i>
IP	700000	600314	600314	0.04	64	66	127
HHRU		600314	600314	1	64	66	127
IP	800000	600314	600314	0.04	64	66	127
HHRU		600314	600314	1	64	66	127
IP	900000	600314	600314	0.03	64	66	127
HHRU		600314	600314	1	64	66	127
IP	1000000	600286	897653	0.05	64	66	124
HHRU		600286	897653	1	64	66	124

Number of nodes that can answer these input queries are 64.

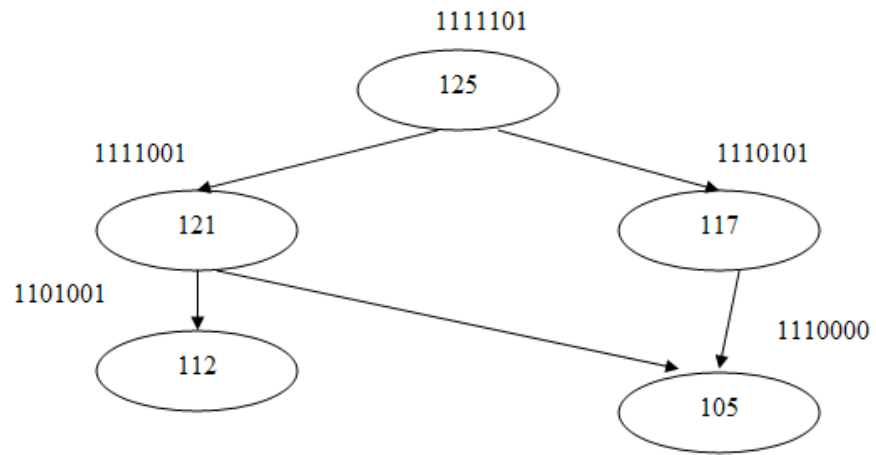


Figure 4: ancestor/predecessor relationships among input queries

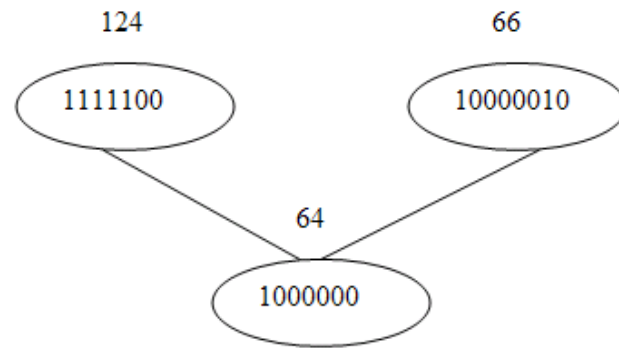


Figure 5: ancestor/predecessor relationships among input queries

The above experiments were done on the 7-attribute lattice where sizes are number of rows. As we see, in most cases, HHRU performs optimally. However, we were suspected that when the sizes are measured in bytes, HHRU algorithm performs worse. Following are some instances which supports this suspicion.

Again there are ancestor/predecessor relationships among input queries, but we can find in couple instances, IP and HHRU solution is different.

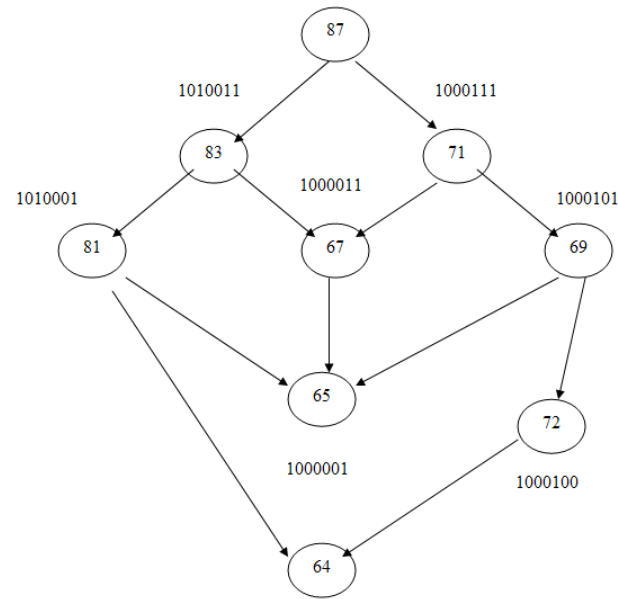


Figure 6: Input queries from 7-attribute lattice where sizes are in number of bytes

In the following table, we can see the results of the experiments which were done over the input queries shown in figure 6 from a 7-attribute lattice where sizes are in number of bytes.

store limit	HHRU cost	IP cost	Ratio
2500000	16877309	16877309	1
3000000	16877309	16877309	1
3500000	15363542	15363542	1
4000000	13279521	13279521	1
4500000	12679913	12679913	1
5000000	12679913	12665201	1.00116
5500000	11451625	11451625	1
6000000	11451625	11451625	1
6500000	11151630	10866317	1.02626
7000000	10552183	10552183	1
7500000	10552183	10552183	1
8000000	10252188	10252188	1
8500000	10252188	10252188	1
9000000	10252188	9952338	1.03013
9500000	9952342	9952338	1
10000000	9952342	9952338	1
10500000	9652524	9652519	1
11000000	9652518	9652518	1
11500000	9652518	9652519	1
12000000	9352705	9352705	1

In the following table, we can see the results of the experiments which were done over the input queries shown in figure7 from a 13-attribute lattice where sizes are in number of bytes.

store limit	HHRU cost	IP cost	Ratio
5000000	37776564	37776564	1
6000000	34998788	34998788	1
7000000	30281214	30281214	1
8000000	26983260	26983260	1
9000000	25404740	25404740	1
10000000	25184256	25184256	1
11000000	25184256	24461558	1.02954423
12000000	24205404	24205404	1
13000000	24205404	23768275	1.01839128
14000000	23262222	23262222	1
15000000	23262222	23262222	1
16000000	22839628	22748988	1.00398435
17000000	22662506	22662506	1
18000000	22662506	22149272	1.0231716
19000000	22239912	22149272	1.00409223
20000000	22239912	22149272	1.00409223
21000000	21834878	21834878	1
23000000	21535069	21535069	1
24000000	21535064	21535064	1
25000000	21535064	21535064	1
26000000	21235255	21235255	1

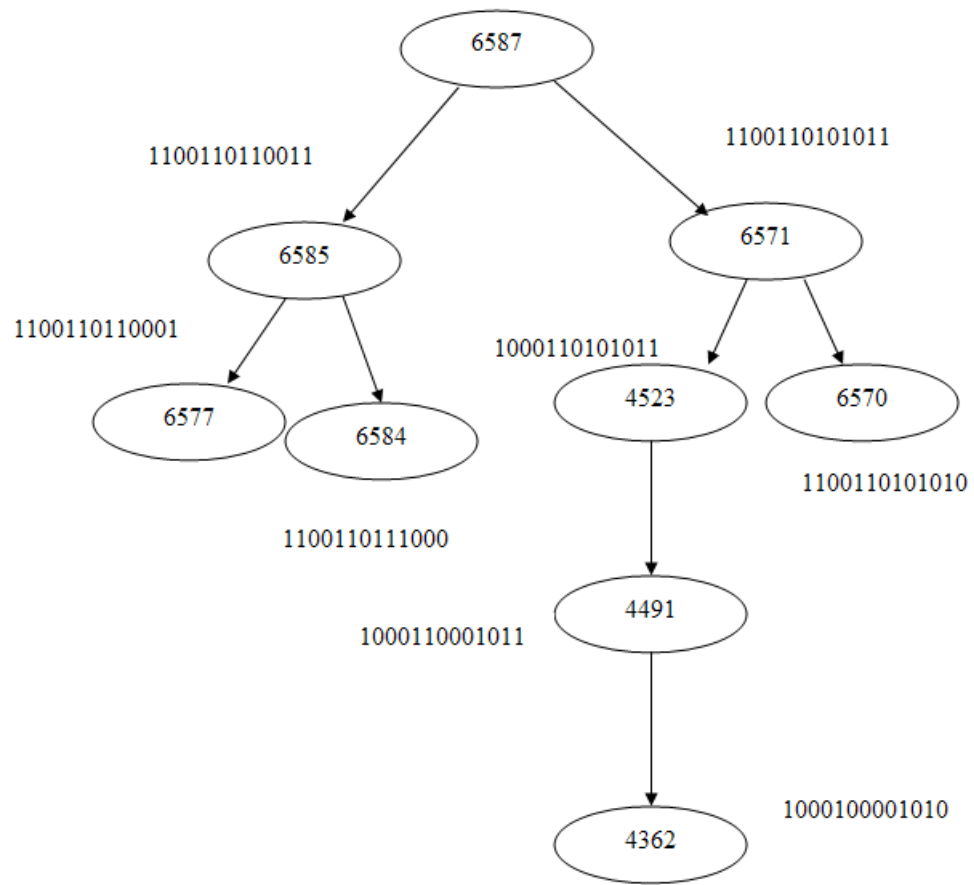


Figure 7: Input queries from 13-attribute lattice where sizes are in number of bytes