

# Data Quality for Semantic Interoperable Electronic Health Records

*Shivani Batra & Shelly Sachdeva*

Jaypee Institute of Information Technology University,  
Sector-128, Noida, India

**Workshop on Quality in Databases (QDB 2016) in conjunction with VLDB 2016**

# Presentation Outline

- Motivation
- Standardized Electronic Health Records (EHRs)
- Layered Approach
- Archetype Authoring Process
- Proposed Approach
- AEAV- Archetype based Entity Attribute Value model
- Physical Storage
- Timeliness
- Conclusions

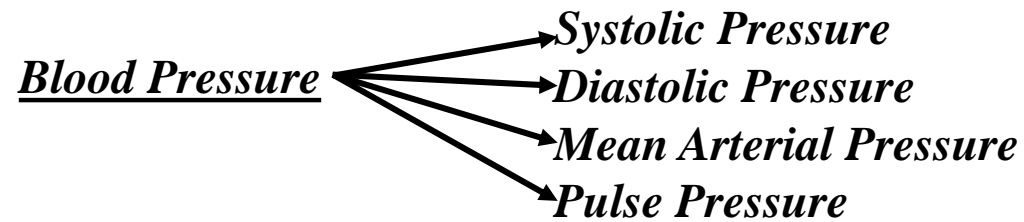
# Motivation

- Data misinterpretation

Hospital Name	Body Temperature
APPOLO	104
FORTIS	40

# Motivation

- Data misinterpretation
- Distinct set of attributes for same medical concept



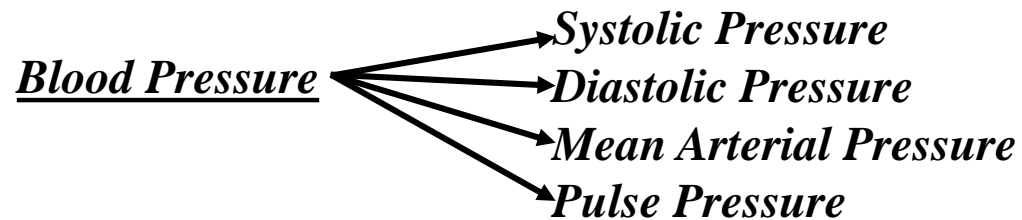
Hospital A
<i>Systolic</i> <i>Diastolic</i>

Hospital B
<i>Mean Arterial</i>

Hospital C
<i>Systolic</i> <i>Diastolic</i> <i>Mean Arterial</i>

# Motivation

- Data misinterpretation
- Distinct set of attributes for same medical concept
- Distinct local schema



Hospital A
<i>Systolic</i>
<i>Diastolic</i>

Hospital B
<i>Mean Arterial</i>

Hospital C
<i>Systolic</i>
<i>Diastolic</i>
<i>Mean Arterial</i>

.....	Systolic	Diastolic	.....	.....

.....	Mean Arterial	.....	.....	.....

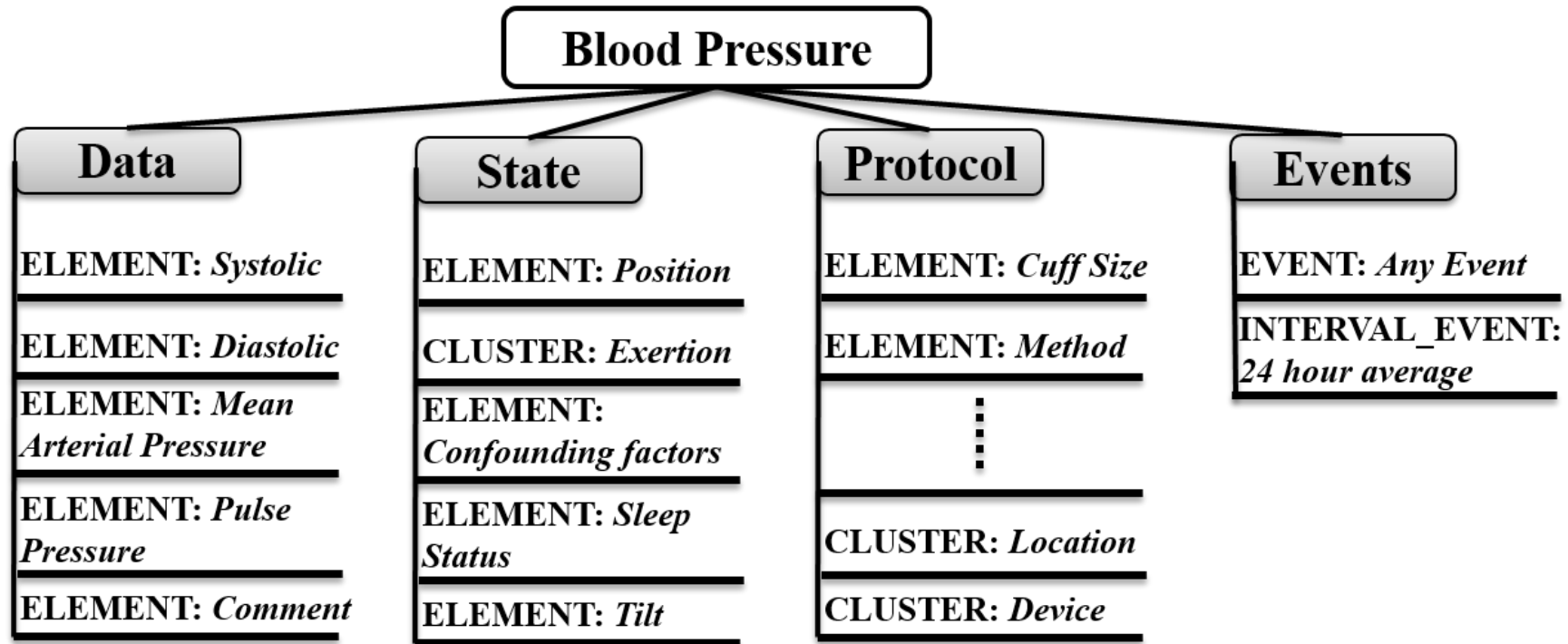
.....	Systolic	Diastolic	Mean Arterial	.....

# Motivation

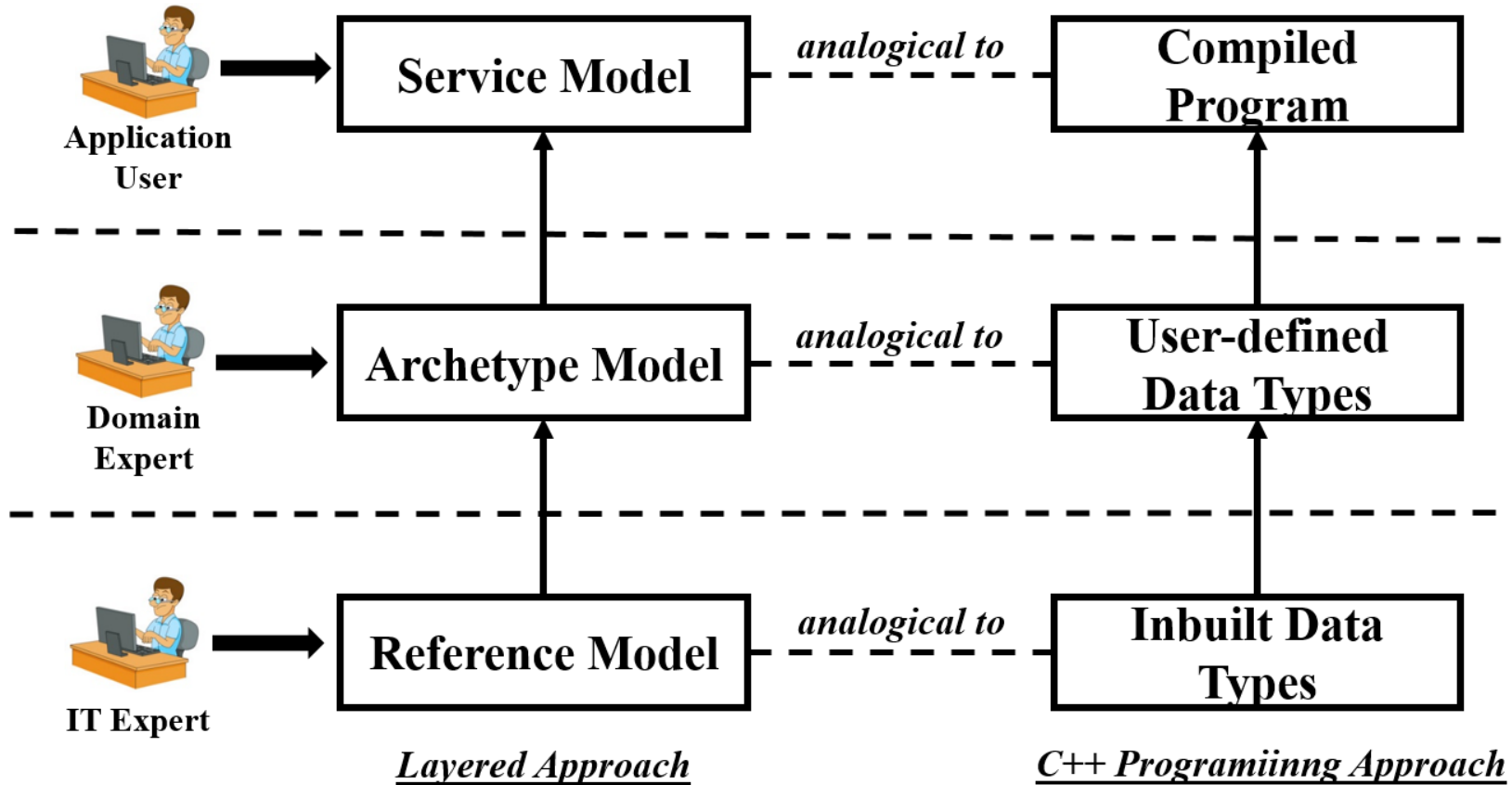
- Data misinterpretation
- Distinct set of attributes for same medical concept
- Distinct local schema } Generic Structure
- Timeliness } Physical Organization

} Standardization

# Standardized Electronic Health Records

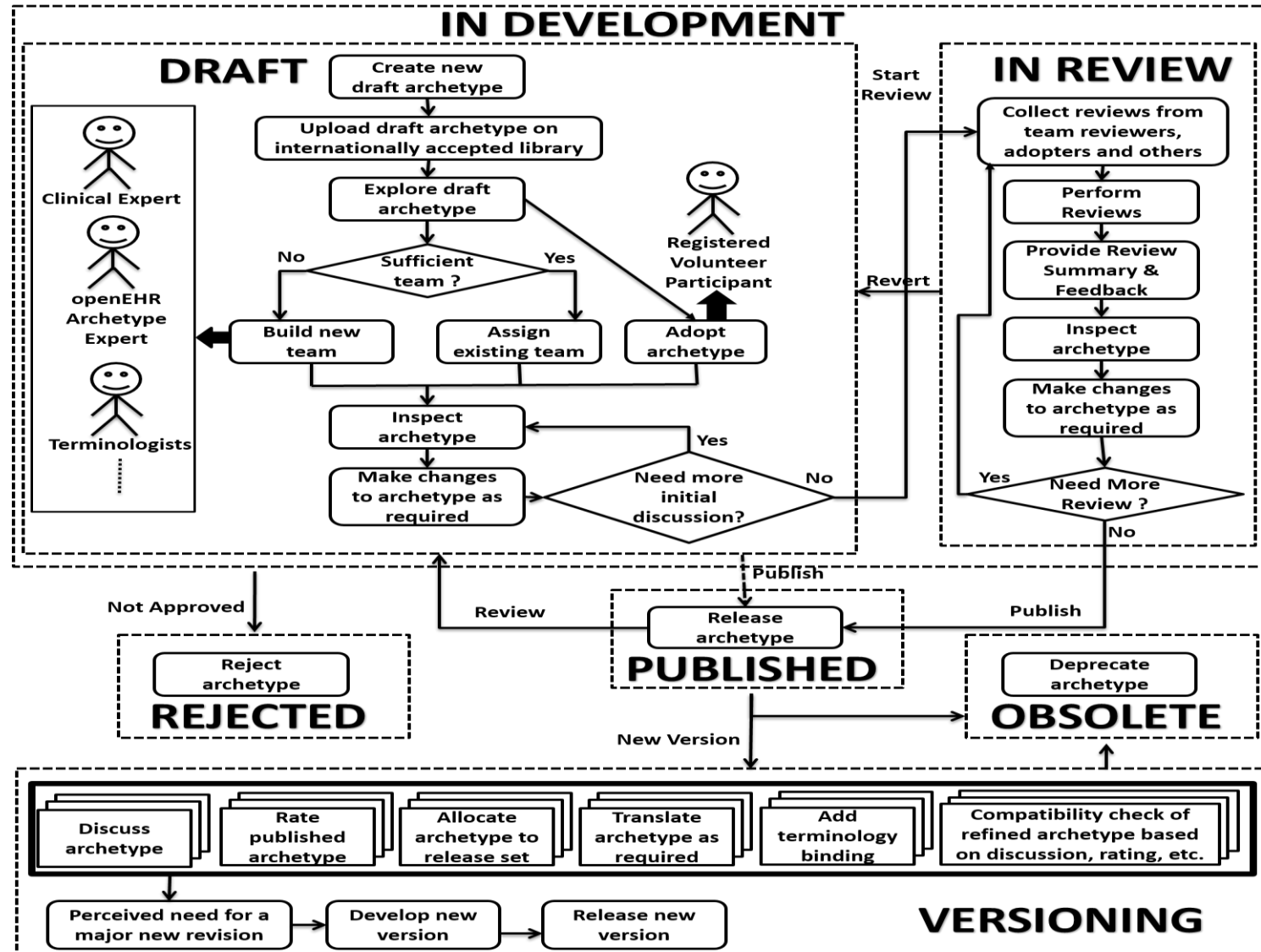


# Layered Approach for Standardization





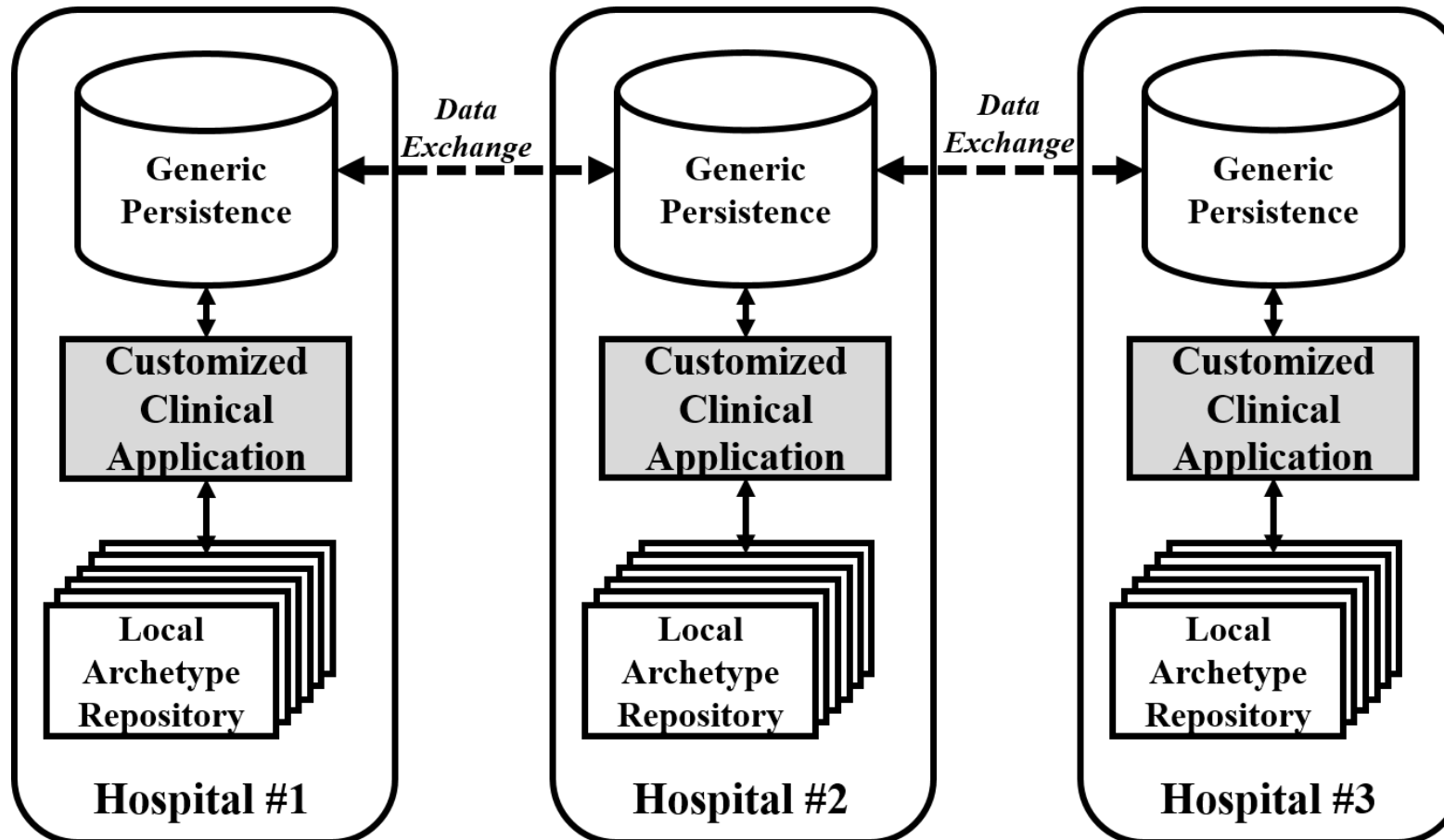
# Archetype Authoring Process



# Solution to Challenges Identified

#	Challenge	Proposed Solution
1	Data Misinterpretation	<ul style="list-style-type: none"><li>• Solved through adoption of archetype based system.</li><li>• Using archetypes aids in capturing maximum possible information about medical concept.</li><li>• Archetypes provide links to standard medical terminologies such as, SNOMED-CT and LOINC.</li></ul>
2	Distinct set of attribute for same medical concept	<ul style="list-style-type: none"><li>• Archetypes define standard set of attribute for a medical concept.</li><li>• Archetypes following one standard can be transformed to archetype following another standard using online tools such as, POSEACLE convertor.</li></ul>

# Proposed Approach



# EAV- Entity Attribute Value model

Relational Model

Entity	Attribute1	Attribute2
Entity_id1	1	2
Entity_id2	3	null
Entity_id3	null	4

EAV Model

Entity	Attribute	Value
Entity_id1	Attribute1	1
Entity_id1	Attribute2	2
Entity_id2	Attribute1	3
Entity_id3	Attribute2	4
Entity_id3	Attribute3	5

# AEAV- Archetype based Entity Attribute Value model

Entity	Archetype_Name	Attribute_Name	Value_int
1	Blood Pressure	Systolic	105
1	Blood Pressure	Diastolic	85
2	Body Weight	Weight	70
3	Body Mass Index	Body Mass Index	40
2	Blood Pressure	Systolic	125
2	Blood Pressure	Diastolic	100
4	Body Weight	Weight	65

Entity	Archetype_Name	Attribute_Name	Value_string
2	Blood Pressure	Comment	High
1	Blood Pressure	Comment	Normal

Mapping Attribute Name

Attribute_Name	Coded Value
Systolic	001
Diastolic	002
Mean Arterial	003
Pulse Pressure	004
Comment	005

Mapping Archetype Name

Archetype_Name	Coded Value	Index
Blood Pressure	001	
Body Weight	002	
Body Mass Index	003	

Attribute_Name	Coded Value
Weight	001
Comment	002

Attribute_Name	Coded Value
Body Mass Index	001

1. Convert numeric code of Archetype\_Name into equivalent 8 bit binary code.
2. Append '00000000', i.e. eight 0 bits to the end of 8 bit Archetype\_Name code to make it a 16 bit code.
3. Convert the 16 bit code into an equivalent decimal and replace existing Archetype\_Name value with this new value.
4. Add decimal values of Archetype\_Name and Attribute\_Name columns and replace Archetype\_Name and Attribute\_Name columns are with one column named ArchAtt containing this summation value.

Entity	Archetype_Name	Attribute_Name	Value_int
1	001	001	105
1	001	002	85
2	002	001	70
3	003	001	40
2	001	001	125
2	001	002	100
4	001	001	65

Entity	ArchAtt	Value_int
1	257	105
1	258	85
2	513	70
3	769	40
2	257	125
2	258	100
4	257	65

Entity	Archetype_Name	Attribute_Name	Value_string
2	001	005	High
1	001	005	Normal

Entity	ArchAtt	Value_string
2	261	High
1	261	Normal

# AEAV- Archetype based Entity Attribute Value model

Entity	Archetype _Name	Attribute _Name	Value_ int
1	001	001	105
1	001	002	85
2	002	001	70
3	003	001	40
2	001	001	125
2	001	002	100
4	001	001	65



Entity	ArchAtt	Value_int
1	257	105
1	258	85
2	513	70
3	769	40
2	257	125
2	258	100
4	257	65

Entity	Archetype _Name	Attribute _Name	Value_ string
2	001	005	High
1	001	005	Normal



Entity	ArchAtt	Value_string
2	261	High
1	261	Normal

Entity	ArchAtt	Value_Int
1	257	105
2	257	125
4	257	65
1	258	85
2	258	100
2	513	70
3	769	40

# Solution to Challenges Identified

#	Challenge	Proposed Solution
1	<b>Data Misinterpretation</b>	<ul style="list-style-type: none"> <li>Solved through adoption of archetype based system.</li> <li>Using archetypes aids in capturing maximum possible information about medical concept.</li> <li>Archetypes provide links to standard medical terminologies such as, SNOMED-CT and LOINC.</li> </ul>
2	<b>Distinct set of attribute for same medical concept</b>	<ul style="list-style-type: none"> <li>Archetypes define standard set of attribute for a medical concept.</li> <li>Archetypes following one standard can be transformed to archetype following another standard using online tools such as, POSEACLE convertor.</li> </ul>
3	<b>Distinct Local Schema</b>	<ul style="list-style-type: none"> <li>Proposed generic schema, AEAV handles this issue.</li> <li>Schema is capable to capture all existing and future data requirements without making any changes in schema.</li> </ul>
4	<b>Sparseness</b>	<ul style="list-style-type: none"> <li>AEAV doesn't store any null value.</li> <li>AEAV reduces space by eliminating need of storing long archetype and attributes names.</li> </ul>

# Physical Organization (Row & Column)

Entity	ArchAtt	Value_Int
1	257	105
2	257	125
4	257	65
1	258	85
2	258	100
2	513	70
3	769	40

Row  
Organization

1	257	105	2	257	125	4	257	65	1	258	85	2	258	100	2	513	70	3	769	40
---	-----	-----	---	-----	-----	---	-----	----	---	-----	----	---	-----	-----	---	-----	----	---	-----	----

Column  
Organization

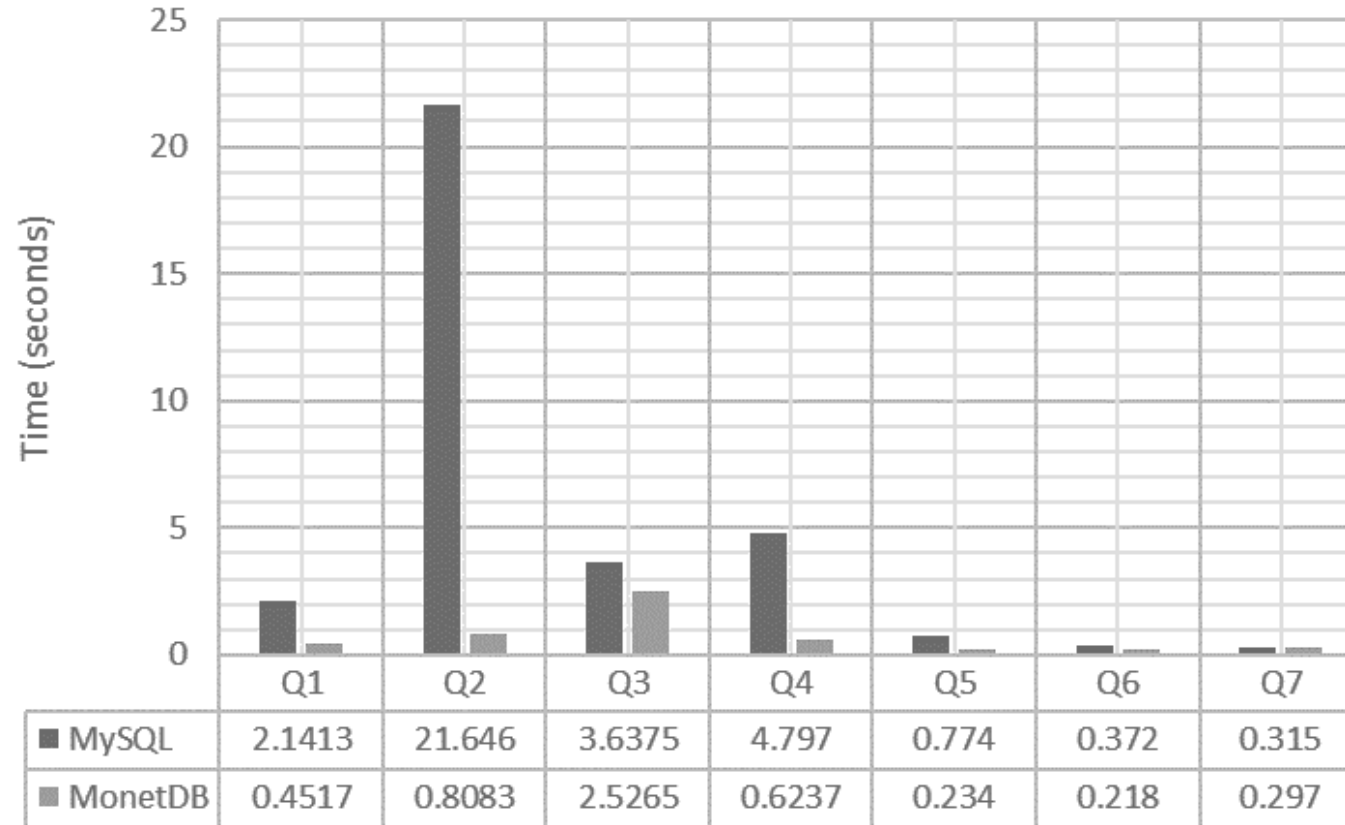
1	2	4	1	2	2	3	257	257	257	258	258	513	769	105	125	65	85	100	70	40
---	---	---	---	---	---	---	-----	-----	-----	-----	-----	-----	-----	-----	-----	----	----	-----	----	----



# Timeliness

ID	Task	Task Description	Time Taken (seconds)	
			Row-Oriented Storage	Column-Oriented Storage
Q1	Extracting Complete Column Details	Extracting details of Systolic pressure	0.377	0.359
		Extracting Systolic pressure, Diastolic pressure and overall interpretation of all patients	0.618	0.419
		Extracting ALP, AST, ALT, Albumin and Globulins of all patients	5.429	0.577
Q2	Extracting Complete Row Details	Extracting data of all patients	63.684	1.482
		Extracting data of all patients having Total T3 greater than 2	0.499	0.374
		Extracting data of all patients having Systolic pressure greater than 100, Diastolic pressure less than 100 and overall interpretation as Hypotension	0.755	0.569
Q3	Extracting Selected Column Details of Selected Rows	Extracting Systolic pressure, Diastolic pressure and overall interpretation of all patients having Patient ID greater than 4500 and Systolic pressure greater than 100	2.805	1.091
		Extracting ALP, AST, ALT, Albumin and Globulins of all patients having Patient ID less than 5000 and AST greater than 100	4.47	3.962
Q4	Performing Statistical Analysis	Extracting the average value of albumin among people tested for Liver	1.36	0.687
		Extracting number of patients tested for BP and diagnosed with Hypotension	0.396	0.232
		Group the patients tested for Liver according to Albumin values	12.635	0.952
Q5	Adding data	Insert data of one patient	0.774	0.234
Q6	Deleting data	Delete data of one patient	0.372	0.218
Q7	Modifying data	Update data of one patient	0.315	0.297

# Timeliness



# Conclusions

- supports syntactic, structural and semantic interoperability,
- refers a generic schema capable of capturing all current and future data requirements without making any changes in schema,
- eliminates the need of storing null values to save storage space,
- supports storage of heterogeneous data,
- improves search efficiency by utilizing optimization techniques of MonetDB.

# References

1. Atalag, K. and Bilgen, S. Multi-Level Modeling and the Role of Archetypes in the Design of Health Information Systems: A Modeling Example in Endoscopy. *HIBIT '07 Proceedings of the International Symposium on Health Informatics and Bioinformatics*, May2007.
2. Atalag, K. and Yang, H.Y. From openEHR domain models to advanced user interfaces: a case study in endoscopy. *In Health Informatics New Zealand Conference*, November 2010.
3. Beale, T. and Frankel, H. *The openEHR Reference Model Extract Information Model*. The openEHR release, 1, 2007.
4. Beale, T. and Heard, S. *openEHR architecture overview*. openEHR Foundation. London, UK, 2008.
5. Beale, T. and Heard, S. *The openEHR archetype model archetype definition language ADL 1.4*. openEHR release, 1(2), 2008.
6. CEN European committee for Standardization: [www.cen.eu](http://www.cen.eu) [online] (Accessed 01/16).
7. Clinical Knowledge Manager: <http://www.openehr.org/ckm/> [online] (Accessed 01/16).
8. Costa, C.M., Menárguez-Tortosa, M. and Fernández-Breis, J.T. Clinical data interoperability based on archetype transformation. *Journal of biomedical informatics*, 44(5), 2011, 869-880.
9. Dinu, V. and Nadkarni, P. Guidelines for the effective use of entity–attribute–value modeling for biomedical databases. *International journal of medical informatics*, 76(11), 2007, 769-779.
10. Health Level 7 International - Homepage: [www.hl7.org](http://www.hl7.org) [online] (Accessed 02/16).
11. International Health Terminology Standards Development Organization. Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). Available from: <http://www.ihtsdo.org/snomed-ct/> [online] (Accessed 04/16).
12. ISO organization: [www.iso.org](http://www.iso.org) [online] (Accessed 01/16).
13. Logical Observation Identifiers Names and Codes – LOINC: <https://loinc.org/> [online] (Accessed 04/16).
14. openEHR Foundation: [www.openehr.org](http://www.openehr.org) [online] (Accessed 01/16).
15. Patrick, J., Ly, R. and Truran, D. Evaluation of a persistent store for openEHR. *HIC 2006 and HINZ 2006: Proceedings*, 2006, 83-89.
16. Sachdeva, S. and Bhalla, S. Semantic interoperability in standardized electronic health record databases. *Journal of Data and Information Quality (JDIQ)*, 3(1), 2012, 1-37.
17. Sachdeva, S., Yaginuma, D., Chu, W., & Bhalla, S. Dynamic generation of archetype-based user interfaces for queries on electronic health record databases. *In Databases in Networked Information Systems*, Springer Berlin Heidelberg, 2011, 109-125.
18. UCI Machine Learning Repository: Liver Disorders Data Set: <https://archive.ics.uci.edu/ml/datasets/Liver+Disorders> (Accessed 06/15).
19. UCI Machine Learning Repository: Thyroid Disease Data Set: <https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease> (Accessed 06/15).

**Thanks ...**