

©2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Petro-data cluster mining - knowledge building analysis of complex petroleum systems

Shastri L NIMMAGADDA* and Heinz DREHER**

Schlumberger, P O Box 9056, East Ahmadi, Kuwait*

School of Information Systems, Curtin Business School**

Curtin University of Technology, Perth, WA, Australia

Abstract- Large volumes of historical petroleum data are available and presently unused primarily because of lack of knowledge. Initially, conceptual data models are derived and warehoused for data mining. For this purpose, petroleum system analysis and knowledge mapping of geological structure, reservoir, well and oil/gas production data are done concentrating on the key issues of geological-structure, reservoir and production data dimensions. Clustering is a data-mining tool for categorizing and analyzing groups of these data dimensions having similar attribute characteristics or properties. Using data warehousing, mining and interpretation strategies, petro-clustering is designed for understanding petroleum systems. Knowledge acquired on petroleum data clusters enhances understanding of relationships among petroleum data attributes, which can optimize economics of oil and gas exploration and development in the petroleum bearing basins.

I. INTRODUCTION

Clustering is a useful technique [1, 2] for the discovery of patterns from oil and gas exploration and production data instances. The goal of clustering is to discover both the dense and sparse regions of plotted business datasets. Data clustering is studied in statistics, machine learning, and database communities with diverse emphases. Several approaches of data mining do not adequately consider the fact that the dataset can be too large to fit into the main memory of a computer. Data mining algorithms need to be computationally efficient in addition to being memory-space efficient. The basic principle of clustering [1] hinges on a concept of distance metric or similarity metric. Since petroleum data are real numbers for statistical applications and pattern recognition, a large class of metrics exists and special purpose metrics can be defined depending on the requirement. In a typical petroleum database, the number of attributes could be very large, while the size of an average transaction or its attribute value is small. Furthermore, basins [3, 4] having similar oil-play, for example, similar reservoir patterns belonging to a single local cluster, contribute to large scale spatial extents of reservoir attributes within the petroleum system and thus enhancing knowledge on reservoir presence at large scale.

An attempt has been made analyzing the petroleum systems of different oil-bearing sedimentary basins of Western Australian, Indian and Middle Eastern basins by means of clustering of oil-play factors. Several authors have already made substantial contributions on the applicability and

feasibility of cluster analysis in various industrial situations [2, 5].

Rule mining and decision-tree analysis are alternate tools for mining the heterogeneous exploration and production petroleum data, but definite advantage of clustering is representation of different data dimensions in terms of visualization, which is useful, in the enhancement of data view interpretation. Hierarchical and partition cluster techniques in particular, have strength in uncovering predefined number of clusters.

II. PROBLEM DEFINITION AND ISSUES

Petroleum industries handle complex and large volumes of multi-dimensional data with numerous entities or objects and attributes with multiple associations. Data warehouse is constructed with a concept of multidimensional relational [1] and hierarchical database structures. Several data dimensions are integrated within a warehouse environment for data mining purposes. During the data mining stage, each relational dataset is grouped with similar properties; a group is considered as a cluster. It is intended simplifying large coarse-grained relational data instances of petroleum data into more fine-grained clusters, so that the volume of data views generated in response to queries within short time periods are timely and precise as input to quality business decisions. In order to achieve this, one needs rigorous mining of existing historical exploration and production data resulting in the accretion of more reserves without the need for additional surveys or drilling. Un-organized volumes of massive storage of oil and gas business data can hide undiscovered, or unknown knowledge and intelligence, which if properly organized and mined, can interpret the patterns, correlations and trends among exploration, drilling and production data as well as oil play factor analysis into meaningful geological and business information.

The aim of the present study is to investigate and examine scalable and high-dimensional data clusters based on similarity of characteristics and interpret them to build knowledge of petroleum systems. Each petroleum system is narrated with oil-play factors. The main emphasis is on ability in identifying and interpreting different clusters of petroleum exploration and production data, separating outliers by shapes of clusters. Another significant issue is examining types of attributes of heterogeneous petroleum data that can

be handled by cluster algorithms. Most of the Western Australian basins [5] are prospective and productive compared to other basins (Australia). Historical exploration and production data in Carnarvon, Bonaparte, Browse and Perth basins (Australia) hold multi-pay horizons in many wells. Similarly, Arabian Gulf (onshore) and Cambay Onshore Basins (Indian Sub-continent) possess multiple pay horizons, interpreted in many wells and surveys. At places, quality of oil play factors of a prospect in one basin may be poor when compared to other basins. Understanding productivity of the prospect in a basin is a significant problem. Data warehousing and mining of different fields or prospects of different basins is a significant step to solving the oilfield yield problem. At this time little effort has been put in integrating, mining and interpreting historical data in terms of data mining (especially in terms of cluster analysis) to facilitate analyzing the petroleum plays of different fields and basins and predict more reservoirs from the existing data volumes.

Another significant issue is plotting variable attribute dimensions with X and Y coordinates together with the similarity property discovered during data mining. This property, the similarity between two attributes is plotted in the third dimension facilitating the explicit visualization of clusters using graphic objects known as bubbles or clouds. For example, oil-play entities or objects and their attributes, which may have similar characteristics in different oil or gas fields of different basins, when clustered, can reveal geologically scientific and commercial value information.

The present study is intended to provide an insight to all petroleum business professionals interested in exploring oil-play data patterns from large databases, using clustering techniques as demonstrated for example by [6, 7]. This will also facilitate explorers or investors for updating their knowledge, if investment has to be made in the oil and gas exploration companies.

III. DESCRIPTION OF DATA TYPES

The basic idea of clustering petroleum business data is to integrate and merge similar properties of oil-play factors which otherwise remain undiscovered and hence unintelligible. Benefits of this approach include interoperability and knowledge reuse. As stated earlier, clustering, in the present context is broadly described as identification of similar property or characteristics petroleum business data and investigating their relationships. More specifically, oil and gas data entities, possess in general spatio-temporal dimension. Spatial data are in the form of X, Y, Z coordinates. Historical data are periodic in nature. Authors have described details of data types in [5]. Exploration, drilling, production and marketing are key activities of any petroleum producing company. Exploration, as a super-entity has several sub-type entities or objects, such as geology, geophysics, well logging, reservoir, logistics and inventory. Similar sub-type entities or objects can be derived from drilling, production and marketing operational data. An

ontological framework [5] is derived keeping in view an overall petroleum exploration business data. Different data visualization software is used to present the extracted data views for interpretation. The methodology and integrated framework used in the current study are discussed in the following section.

IV. METHODOLOGY

Entities and attributes of petroleum database are considered as multiple dimensions [1]. Multidimensional data modeling is carried out for warehousing large size petroleum data. Ontology is used for organizing and conceptualizing volume of attributes, which further facilitates clustering algorithms to group and categorize *similarity* property attributes. Data structures are denormalized [8] to arrive at finer groups that may provide more knowledgeable and intelligent clusters. However, the inclusion of irrelevant attributes can be futile to a successful clustering outcome because they negatively affect proximity measures and eliminate clustering tendency. A sound Exploratory Data Analysis (EDA) is required as given in [9]. EDA eliminates inappropriate attributes and reduces cardinality of the retained categorical attributes. EDA provides attribute selection, but cluster-specific attributes are yet to be invented - attribute scaling is viewed as the continuation of attribute selection. Applications that derive their data from measurements may have an associated amount of noise, which can be viewed as legitimate records having abnormal behavior even if clustering techniques do not distinguish between noise and the abnormalities that fit into clusters. Pre-processing of data, such as partitioning or data summarization, may help in minimizing the outliers.

Clustering of petroleum data is a method by which large sets of exploration and production data are grouped into clusters of smaller sets of similar data. The examples shown in the following sections demonstrate the clustering into balls/bubbles of similar and dissimilar colors. Authors are interested to group of different balls of similar colors into particular set of groups. Clustering implies grouping of exploration and production data or dividing a large set into smaller data sets with some similarity. The similarity property of attributes of petroleum business data entities or objects is the criterion used in the present studies by plotting similar attribute values from different oil fields and basins in the same scale. Exploration and production characteristics of different wells in different fields of different basins can be revealed if plotted in the same scale to permit their visualization along with their densities (distributions) and magnitudes. The relative shift or position of clusters shows which cluster characterizes and contributes more to oil-play of a petroleum system and thus informs the exploration and development of an oil field.

Multidimensional and object oriented data schemas are deployed keeping in view what is expected for interpreting finer clusters of petroleum business data. In general, relationships among the common attributes of petroleum-plays data are denormalized, so that the final data clusters

become finer. In the present study, data relevant to performances of petroleum systems have been organized in their appropriate data structures, for the purpose of fine-grained data mining processes [10, 11]. The multidimensional model consists of several fact tables (surveys, wells and permit facts or oil-play factors) surrounded by hundreds of dimension tables [5] in several star-schemas. Being a fact constellation schema, high-level granularity [11] has been maintained in order to derive fine-grained queries. Combination of star, snowflake and fact constellation schemas definitely with normalized and denormalized relationships can expect to generate fine-grained clusters from large volume of data warehouse. The above *surveys*, *wells* and *permits* schemas, though represented as star schemas [1, 5], are also designed for fact constellation schemas.

The models discussed so far [5] include well, lease, seismic, culture, asset and production data for petroleum industry. Our approach is to integrate all data through warehouse metadata model [1, 5]. In this integration process, comprehensive well data model incorporates tables for tops, cores, tests, production, logs run, deviation surveys and well location. Similarly, *seismic surveys* model uses records of the acquisition and processing history of seismic lines along with the location data. An exploration ontology model, representing surveys, wells, and survey/well permit data entities, has been described in a hierarchical structure as shown in Fig.1.

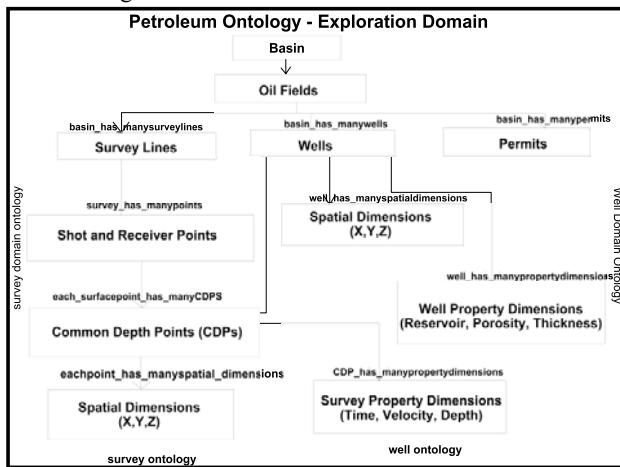


Fig.1: An ontology model showing surveys, wells, permits, and hierarchies

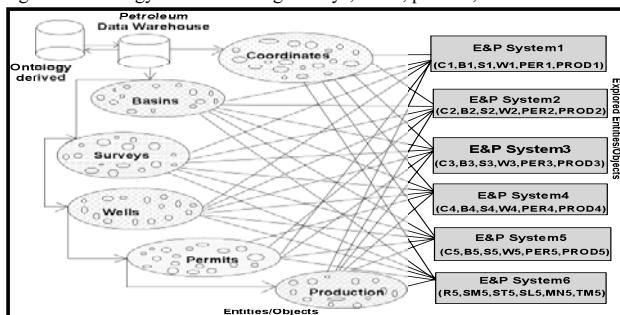


Fig.2: Clustering Process Model for Exploration and Production Systems

A process model deduced from the ontology structure model has been illustrated in the Fig.2 narrating different exploration and development entities, each described with coordinates, basins, surveys, wells, permits and production sub types. In the present study, our scope is to demonstrate building relations among geological structure, reservoir and production dimensions and visualize data views for renewed interpretation.

Relationships among the sub-type data dimensions are conceptualized through ontology [5]. A total framework, describing ontology, integrating data through relational and hierarchical data modeling is described in Fig.3. Surveys, basins, oil and gas fields, reservoirs, structures and production data instances are mapped through the warehouse model. Process of integrating data, and describing in several hierarchies, is given in Fig.4. Several data views are extracted from the warehoused exploration and production data, for mining and interpreting them in a knowledge domain. All data instances, visualized as data views, are represented into certain groups or categories, narrating the clusters. Mining of several data instance clusters and how they are interpreted in a knowledge domain is treated in the following sections.

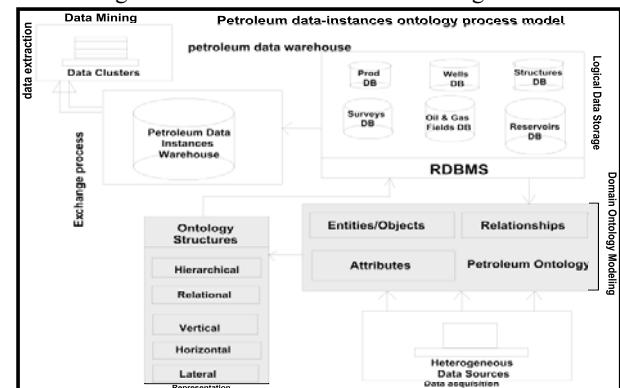


Fig.3: Framework for Warehousing & Mining petroleum data

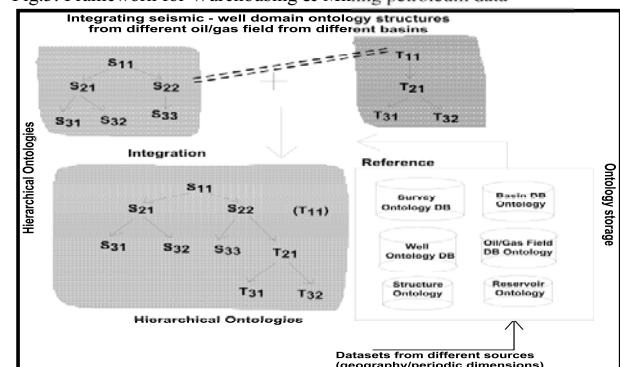


Fig.4: Data integration process model

V. INTERPRETATION OF CLUSTERS FOR KNOWLEDGE BUILDING

The results of clustering must be appropriately interpreted in the wider context of the application. Clustering of different petroleum data attributes may be considered as an initial data exploration tool before the design of further decision-tree or

classification system or constructing an association rule involving a fixed number of classes. In all these cases, the clusters must be suitably transformed or interpreted. The most common representation of clustering is by centroids. This is effective for compact and isotropic clusters but not for elongated or anisotropic clusters. In certain applications, the extremities of a cluster are used to form the conjunctive expressions in rule sets. *Clustering* in our present study is division of petroleum data into smaller groups of similar objects. Each group consists of objects possessing similarities and dissimilar to objects of other groups. Representing data by fewer clusters necessarily diminishes fine-grained details, but achieves simplification. Hidden patterns from clusters are interpreted in terms of variations in distances or concepts of oil-play data. Clustering in the present context, is said to be unsupervised learning of hidden petroleum system data.

The data mining process starts with the assessment of whether any cluster trend has meaning and interpretation, and correspondingly includes appropriate selection, and in many cases feature construction. Validation and evaluation of the resulting clusters must be assessed with their interpretability and visualization. Interpretability depends on the size of data (mostly numerical) and method used for building cluster. Dense areas around centroids score well. Cluster validation and their visualization issues have been addressed by [12, [13]. When two partitions [1, 11, 12] are visualized and interpreted as different sub-sets, each is implied as separable granularity and interpretable cluster. The distance between any two centroids normalized by corresponding clusters' radii (standard deviation) and averaged (with cluster weights) is a reasonable choice of coefficient of separation.

Clustering of onshore oil and gas field data: In one of the medium sized onshore fields in the South-east Asia, there are multi-stacked reservoirs, producing from this field creating surprises. more than 500 wells have been drilled on this single field. Geological structures and reservoirs and their areal extents are uncertain and unknown in the petroleum system. Many wells are wet or abandoned thus costing the industry millions of dollars. We attempt to use data warehousing and mining methodologies to understand data trends in the known oil accumulations and predict the currently unknown. From a well location map, showing existing well placement, metadata is created using a warehousing modeling approach. Hierarchical ontology is used for building relationships among geological structures and reservoirs with their production data instances. These data are integrated in a warehouse environment and mined to create interpretable data views.

Several data views have been extracted for interpreting data trends among several attributes of structure, reservoirs and production entities. These have been shown in Fig.5. Reservoir, geological structure, and production data attributes have been plotted with several bubbles on a single scale to explore data trends. Distribution of bubbles, size of bubbles, and direction of bubbles are key criteria to understanding and exploring data entities and attributes.

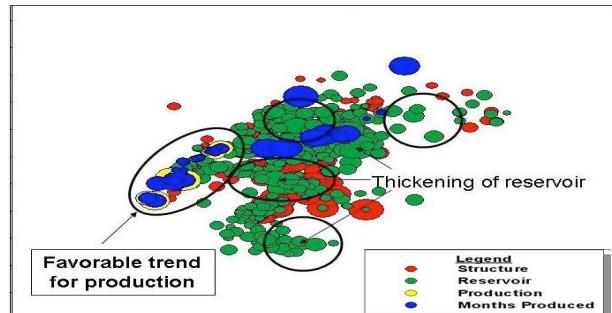


Fig.5: Structure, reservoir and production data entities at producing horizon-1

These bubbles are clustered in 2D X, Y plane. Bubbles are closely clustered because of the fact that data are from single field. However, size of bubble differentiates the attribute strength. Clusters of bubbles are plotted for different data instances for structure, reservoir and production data attributes to explore data patterns. Interesting trends are explored, with good fit among producing structure and reservoir attributes. Some of these data are fitted with linear trends, especially with reservoir thickness and porosity (reservoir properties), inferring relationships among porosity properties among multiple fields, resolving the issues related to structure and reservoir. This has significance in knowledge building and conceptualizing relationships among complex petroleum systems. A prerequisite for achieving this knowledge depends on effective data integration and denormalizing relationships [11] during fine-grained multi-dimensional data structuring process (at warehouse modeling stage).

Clustering of multiple oil and gas fields in Western Australian basins: As shown in Fig.6, period, number of surveys conducted, and number of drilled-well dimensions plotted on one scale, narrate interesting trends of clusters in circles and ellipsoids in the Canning and Carnarvon basins.

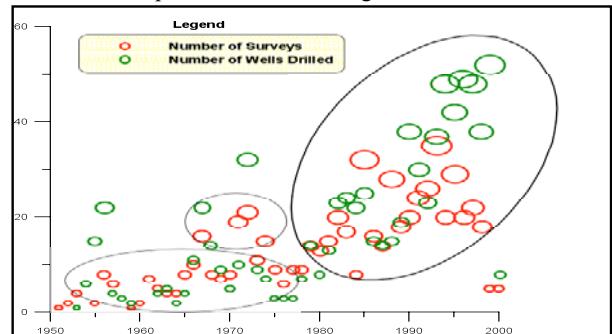


Fig.6: Data views extracted from warehouse, showing clusters of bubbles attributed to several petroleum surveys and wells (Carnarvon basin, WA)

The number of wells or surveys conducted in these basins is steady up to 1980 and shows a cluster of surveys conducted, as unrelated to the number of wells drilled. Clusters of surveys and wells drilled are highly variable during 1990; maximum wells drilled with corresponding number of surveys, suggesting that they relate each other and is interpreted to be that *wells-drilled* is dependent upon the *surveys* conducted. Size of bubbles of these clusters particularly in later years is decreasing.

In the Carnarvon basin (Fig.6), the clusters of number of surveys conducted and wells drilled reveal an exponentially increasing trend over the period. Fig.6 shows that bigger sized bubbles are interpreted in a big cluster in the increasing periods inferring the number of surveys and wells drilled are correlated in the sense that wells drilled in the Western Australian basins depend on the number of surveys conducted. Another interesting interpretation is that where the red and green bubbles are clustered together suggests more wells drilled based on surveys data.

In the case of Perth basin, as shown in Figs. 7, surveys and number of wells drilled are decreasing over time, but these clusters are in increasing order with decreasing periods. As shown in Fig.7a, clusters gathered among number of surveys and wells attributes appear to be matching each other in the increasing periods interpreted to be dependent on each other. As clusters bounded by circles are exponentially increasing with increasing bubble sizes and with period as well. This is interpreted as all the producing wells perfectly match all the drilled wells in all key basins of Western Australia. Wherever oil wells are drilled quite a few gas wells are also discovered, showing multiple reservoirs (Fig.7b), providing good relationship between oil and gas producing wells.

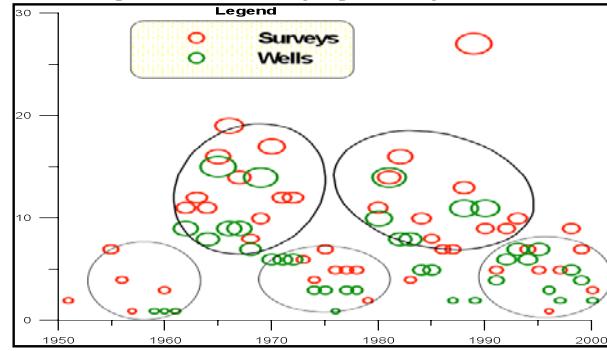


Fig.7a: Data views extracted from warehouse, showing clusters of bubbles attributed to several petroleum surveys and wells (Perth basin, WA)

A similar situation is observed (Fig.7c) between number of structures interpreted and number of hydrocarbon producing wells. As shown in Fig.7d, when clusters associated with number of oil, gas and condensate wells attributes plotted together with total number of wells drilled, it is interesting to observe that cluster associated with number of drilled wells attribute is separated, thereby leaving an impression of cluster associated with number of drilled wells attribute has least relationship with producing wells data attribute clusters.

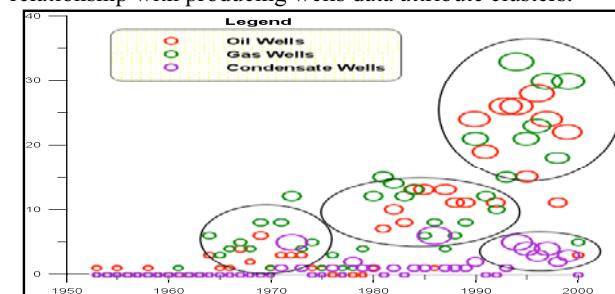


Fig.7b: Data views extracted from warehouse, showing clusters of bubbles attributed to several petroleum surveys and wells (Perth basin, WA)

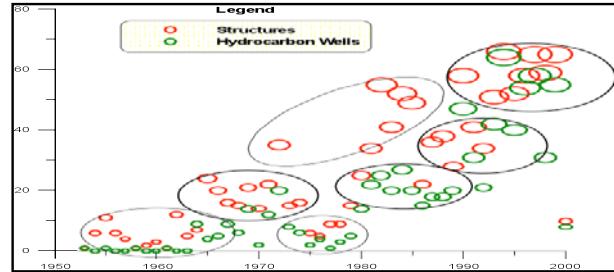


Fig.7c: Data views extracted from warehouse, showing clusters of bubbles attributed to several geological and hydrocarbon wells (Perth basin, WA)

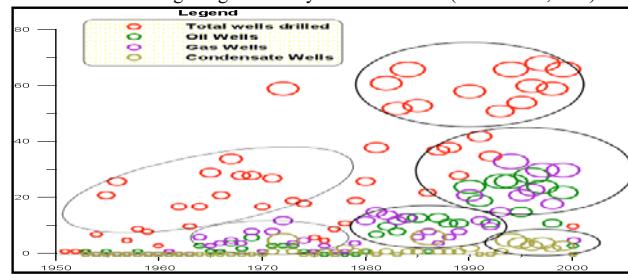


Fig.7d: Data views extracted from warehouse, showing clusters of bubbles attributed to several drilled wells (Perth basin, WA)

As stated earlier, exploration has multiple dimensions narrated in several hierarchies (Figs.1, 3 and 4). Multiple dimensions of exploration are typically, basin, reservoir, structure, drilled well, surveys conducted and oil/gas produced in each field and basin. Clustering algorithms attempt to find natural groups from exploration and production data instances based on certain similarity. These algorithms find the centroid of a group of data sets. To determine cluster membership, most algorithms evaluate the distances between a point and the cluster centroids. The output from clustering algorithm is a statistical description of the cluster centroids with number of data in each cluster. The distance between two points is taken as a common metric to assess the similarity among datasets of volume of exploration data.

In the true representation of clusters for interpretation, authors considered several data instances from the warehoused metadata, populated in several spreadsheets. These data instances are plotted in the form of bubbles, representing clusters in 2D plots. Bubbles of clusters display two variables on a scatter-type plot. In the present study, we used bubbles for narrating and investigating clusters, for varying data dimensions. In a 2D bubble plot, the diameter of each bubble can vary in size, providing a way to represent an additional dimension of data. For example, consider a traditional scatter plot that shows the number horizons/reservoirs producing in Perth, Carnarvon, Cambay basins or Middle-eastern basins over a period of time. Using a 2D bubble plot, you could also display a third dimension of data that shows average production or reservoir type over the same time span or entire basin. Different worksheets are used for changing the data to plot different bubbles. X, Y, Z fields are changed to get appearance of different bubbles with differing sizes and locations. X, Y, Z and size data columns of worksheets and their corresponding attributes are plotted in

the same scale and axes so explore similarity or dissimilarity properties. Change of axes is also performed on different plots, to check the data validity and integrity. Point dimensions attributed in the data trends are also counted, so that strength of attribute property is explored.

Clustering of multiple onshore oil and gas field data from Arabian Gulf Basins: There are multiple oil and gas fields within a producing basin (Fig.8). Each field has hundreds of producing wells, several survey lines and geological information. Though many fields in general have good structural bearing, there are reservoir uncertainties, resulting in the abandonment of some wells. Warehouse modeling has been carried out to understand the data trends and patterns of attributes among these fields.

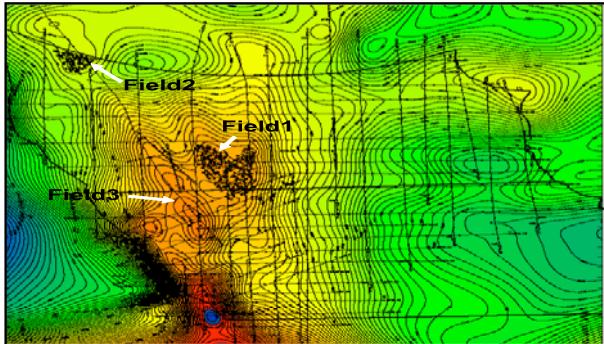


Fig.8: Multiple oil and gas fields in an onshore Arabian Gulf basin

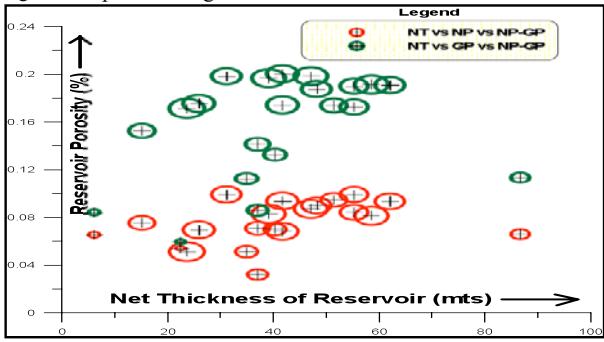


Fig.9a: showing reservoir thicknesses and their porosity distributions

Data attributes associated with multiple fields possess definite data shapes and trends. For example, bubbles represented among plots of reservoir thicknesses and porosities have different clusters by virtue of their unique properties as shown in Figs. 9a-9e.

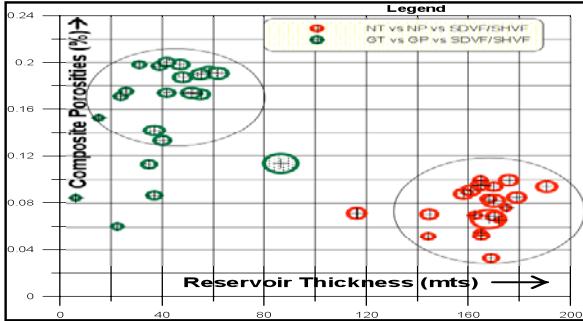


Fig.9b: Clusters of sand volume fraction and their attributed composite porosities from multiple fields

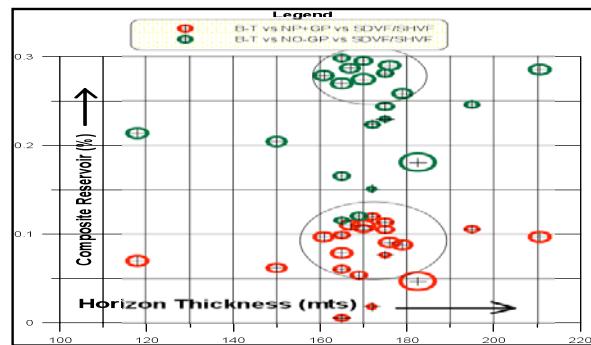


Fig.9c: Clusters of composite porosities of middle-eastern oil and gas reservoirs from multiple fields

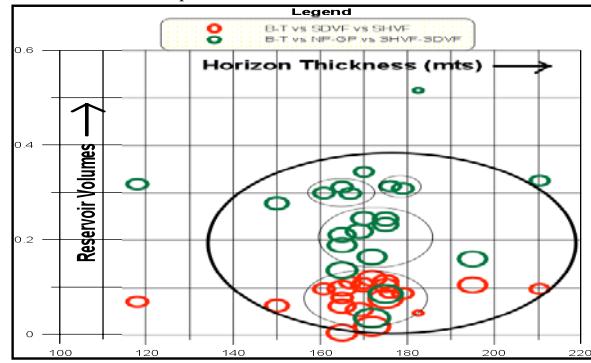


Fig.9d: Clusters of sand volume fraction and their attributed composite porosities from multiple fields

As shown in Figs.10-11, green and red colored bubbles that are clustered with in an encircled cluster are interpreted to be in a single cluster firming up the following equation:

$$(NP - GP) / (NP + GP) = (SHVF - SDVF) / (SHVF + SDVF)$$

The derived equation from this graph is useful for reservoir engineers for computing better reservoir quality areas of the formation. Sand volumes and porosities appear to be forming a cluster within a net reservoir thickness ranging 35-65 mts. This demonstrates that once net porosities (NP), gross porosities (GP) and shale volume fraction (SHVF) are known, sand volume fractions can be computed from the above equation. Porosities clustered in red color are corroborating with sand volume fractions appear to have better correlation within the interpreted net reservoir thickness range.

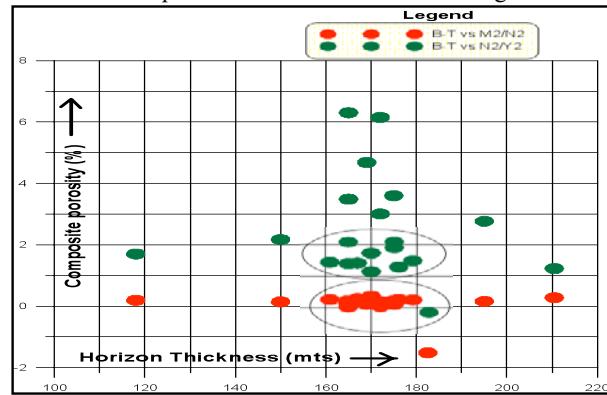
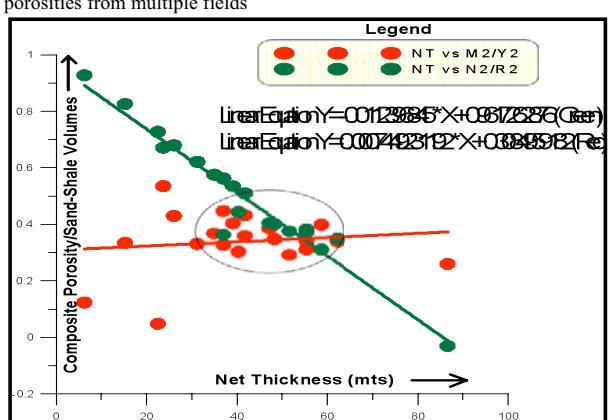
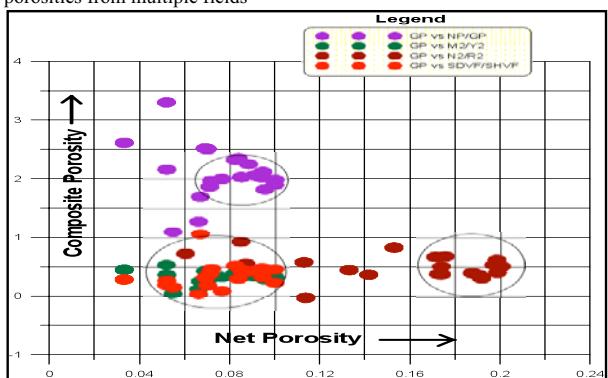
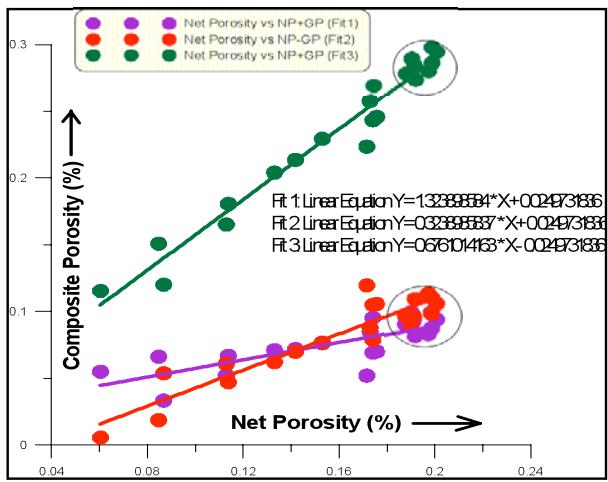


Fig.9e: Clusters of sand volume fraction and their attributed composite porosities from multiple fields



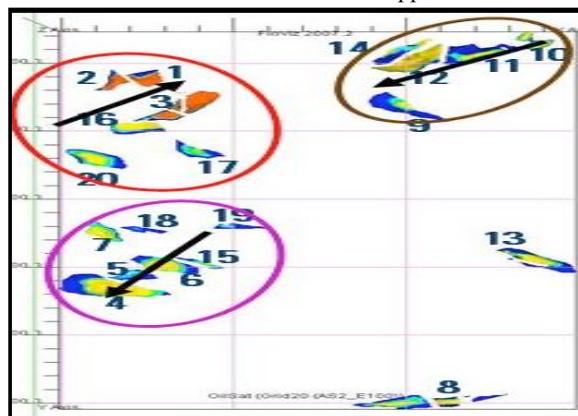
VI. SIGNIFICANCE OF CLUSTERING PETROLEUM DATA MINING

Pattern recognition, artificial intelligence and cybernetics are using the concepts of cluster analysis in the Engineering Sciences [14, 15]. Typical examples include handwritten characters, samples of speech, fingerprints and pictures. In the life sciences, especially, biology, botany, zoology, entomology, cytology, microbiology, objects of analysis are life forms such as plants, animals and insects. The cluster

analysis ranges from developing complete taxonomies to classification of species into subspecies. Cluster analysis is also used in information analysis, policy and decision sciences. Authors attempt to use these ideas in the fields of petroleum engineering and petroleum systems. Non-hierarchical data structures are also suitable for representing various clusters. Centroids and minimum distances are key parameters and values, and thus criteria for exploring and exploiting clusters for maximum knowledge. Various hierarchical data types are used for doing cluster analysis. In the present context, several data views (taken from warehouse) represented as bubbles of clusters, are explored for finding similarity of patterns among petroleum systems data that will further explore the prospectivity of a basin.

The basic idea of this approach is to build conceptual models [16] on a finer scale (in space-temporal domains) and store them intelligently in warehouse environment for investigating fine-grained clusters of petroleum data and accessing cluster views (e.g. clustering of similar reservoir and structural characteristics from several fields and basins, cluster of similar surveys attributed to drillable exploratory well) for interpretation purposes. The motivation is to improve precision and/or recall as well as reduce the overall amount of time spent searching for interpretable cluster views [17, 18], for geological knowledge extraction. Supporting technologies include agents for searching the data and grouping by similar property data clusters, and data delivery agents using meta-data languages (e.g., XML and HTML), and other knowledge representation tools. Clustering algorithm attempts to find natural groupings among exploration and production data based on some similarity. This algorithm finds the centroid of a group of data sets. To determine cluster membership, most algorithms evaluate the distances between a point and the cluster centroids. The output from clustering algorithm is a statistical description of the cluster centroids with number of data in each cluster.

Another data view, with clustered anomalies plotted in Fig. 12, interprets arrow attributes, indicating within encircled data clusters, suggesting complex structure trends and the petroleum occurrences within these entrapped structures.



VII. SCOPE AND FUTURE WORK

Bubble clusters and their distributions, the size and radius of bubble are also contributing to understanding and building relationships among spatial-temporal high-dimensional exploration and production datasets warehoused from different oil and gas fields. Minimum and maximum radius size of the bubble and where they are occurring between minimum and maximum data values are significant criteria. Interpolation is done proportionately among minimum and maximum values to build data trends. Radius formula that computes bubble radius is:

$$Br = ((Z - Z_{min})/(Z_{max} - Z_{min})) * (R_{max} - R_{min}) + R_{min}$$

where Z is the bubble data value and R is the radius of the bubble and B_r = bubble radius. Data ranges are selected to use data minimum and maximum values for determining bubble size. In this method, the smallest value in the Z column (2D) or size column (3D, XYZ) is displayed as the minimum radius value (R_{min}). The largest value in the Z column (2D) or size column (3D, XYZ) is displayed as bubble's maximum radius value (R_{max}).

Bubble Centroids: The centroid location for each bubble is displayed as a symbol on 2D bubble plots. As the bubble symbol is not scaled it is easier to distinguish overlapping bubbles when the centroid is visible. For example, X, Y location of the data point, attributed to any geological structure or producing reservoir is displayed with a symbol with symbol properties. In our present study, circles are plotted and for attribute visualization whilst bubbles are filled with different colors and shades. Bubbles are basically scales based on data values representing the strengths of attribute properties. Gradients of bubbles and symbols are displayed based on the gradients in the data values. At places, gradients are coarse and other places they are fine. These bubble clusters can also be rotated in X, Y directions to explore for meaningful data gradients.

The current study is in progress on “impacts of bubble radius and bubble centroids on different attribute property strengths of different petroleum fields”.

VIII. CONCLUSION

Different domain ontologies that can be integrated in a warehouse environment are briefly discussed. Heterogeneous data are represented in multiple dimensions. Petroleum ontology is appropriately described for representing space and temporal data in multiple dimensions. This approach facilitates to identify and examine several data classifications in the data mining process among petroleum exploration and production data dimensions either geographically located servers or on a local server. Dimensions, attributed to structure, reservoir and other reservoir properties graphically (data visualization) representing in different clusters suggest particular data alignments or patterns. Clustering is a useful technique for discovery of data distributions and trends in the petroleum exploration and production data.

REFERENCES

- [1] A. K. Pujari. *Data Mining Techniques*, University Press (India) Pty Ltd, 2002.
- [2] C. C. Aggarwal, C. Faloutsos, and A. Swami, Efficient similarity search in sequence databases, Proceedings of the 4th International Conferences on Foundations of Data Organization and Algorithms, 1993.
- [3] E. A. Beaumont, and N. H. Foster, Exploring for Oil & Gas Traps, AAPG Publications of Millennium Edition, 1999.
- [4] W. M. Telford, L. P. Geldart, and R. E. Sheriff, *Applied Geophysics*, Second Edition, p.100-350 and p. 600-750, 1998.
- [5] S. L. Nimmagadda, and H. Dreher, Ontology-Based Data warehousing and Mining Approaches in Petroleum Industries: chapter XI in Negro, H.O., Cisaro, S.E.G., and Xodo, D.H, (Eds.), *Data Mining with Ontologies: Implementation, Findings and Frameworks*, pp 211-236. Information Science Reference, IGI Global, Hershey, PA, USA, 2007. <http://www.igi-pub.com/reference/details.asp?ID=6844>
- [6] M. Ester, H.P. Kriegel, J. Sander, and X. Xu, A density based algorithm for discovering clusters in large spatial databases with noise, *Proceedings of the 2nd ACM SIGKDD*, pp. 226-231, 1996.
- [7] E. Forgy, Cluster analysis of multivariate data: efficiency vs. interpretability of classification, *Biometrics*, 21, pp. 768-780, 1965.
- [8] J. A. Hoffer, M. B. Presscot, and F. R. McFadden, *Modern Database Management*, Sixth Edition, Prentice Hall, 2005.
- [9] J. Becher, P. Berkhin, and E. Freeman, E, Automating exploratory data analysis for efficient data mining, *Proceedings of the 6th ACM SIGKDD*, pp. 424-429, 2000.
- [10] E. Keogh, K. Chakraborty, M. Pazzani, and S. Mehrotra, Dimensionality reduction for fast similarity search in large time series databases, *Journal of Knowledge and Information Systems*, 3 (3), 2001.
- [11] A. Rudra, and S.L. Nimmagadda, Roles of multidimensionality and granularity in data mining of warehoused Australian resources data, a paper published in the proceedings of 38th Hawaii International Conference on Information Sciences, Jan 2005.
- [12] A.K. Jain, N.M. Murthy, P.J. Flynn, Data clustering – a review, *ACM Computing Surveys*, Vol 31(3), pp. 264-323, 1999.
- [13] E. Kandogan, Visualizing multi-dimensional clusters, trends and outliers using star co-ordinates, *Proceedings of the 7th ACM SIGKDD*, pp.107-116, 2001.
- [14] C. Fraley, and A. Raftery, Which cluster method? Answers via model-based cluster analysis, *The computer journal*, 41 (8), pp. 578-588, 1998.
- [15] Han, J. Kimer, M. and A. K. H. Tung, Spatial clustering methods in data mining: a survey, in Miller, H. and Han, J. (eds.) *Geographic Data Mining and Knowledge Discovery*, 2001.
- [16] V. Khatri, R. Sudha, and R. T. Snodgrass, Augmenting a conceptual model with geo-spatiotemporal Annotations, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16 (11), 2004.
- [17] P. Berkin, Survey of Clustering Data Mining Techniques, Accrue, Software Inc, 2000 www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf
- [18] P. Andritsos, Data clustering techniques, a qualifying oral examining paper, University of Toronto, 2002, www.cs.toronto.edu/~periklis/pubs/depth.pdf