The North West Shelf (NWS), a Digital Petroleum Ecosystem (PDE) in a Big Data Scale

Shastri L Nimmagadda
School of Information Systems, CBS
Curtin University, Perth, WA.
shastri.nimmagadda@curtin.edu.au

Amit Rudra
School of Information Systems
CBS, Curtin University, Perth, WA.
amit.rudra@cbs.curtin.edu.au

Torsten Reiners
School of Information Systems
CBS, Curtin University, Perth, WA.
T.Reiners@cbs.curtin.edu.au

**presenting author asterisked*

SUMMARY

The North West Shelf (NWS) and its associated petroleum systems have varied geographies, geomorphologies and complex geological environments. In spite of the ongoing exploration activities in many sedimentary basins, the appraisal and field development campaigns are challenging. Besides, interpreting the connectivity between petroleum systems is challenging. The heterogeneity and multidimensionality of multi-stacked reservoirs associated with multiple oil and gas fields complicate the data integration process. Volumes and varieties of data existing in these basins are in different scales, sizes and formats, demanding new storage and retrieval methods, emphasizing both data integration and data structuring. Since the data are in terabyte size; the multiple dimensions and domains need to be brought in a single repository, we take advantage of Big Data tools and technologies. In this context, we aim at articulating the digital petroleum ecosystems and petroleum database management systems, with new data modelling, data warehousing and mining, visualization and interpretation artefacts. This approach facilitates the data management not only for individual basins but groups of basins of the NWS. Warehoused cuboid metadata can explore the connections providing new insights in the data interpretation and knowledge of new prospective areas. The multidimensional warehousing repository that supported by cloud computing, data analytics and virtualization features, provide new opportunities for delivering quality and just-in-time online ecosystem services. Other goals are deducing an integrated unified metadata model and characterizing the connectivity among the basins of the NWS and associated oil & gas fields. The study supports the features of PDE and its knowledge management.

Key words: Big Data, Data Integration, Systems' Connectivity, Digital Petroleum Ecosystems (PDE) and Sustainability.

INTRODUCTION

In Western Australian Coast, the North West Shelf (NWS) comprises of a group of sedimentary basins that produce commercial quantities of natural gas, condensate, and LNG to export to world markets. The NWS is a part of an ongoing investigation of the Westralian super basin, in which several associated sub-type basins are described (Longley et al. 2003). The Carnarvon, Canning, Browse and Bonaparte basins share multiple petroleum systems' elements and processes on a regional scale in WA coast. In recent years, the exploration & production (E & P) of new oil and gas deposits in these basins have become challenging, because of complex geological environments and interpretation of new knowledge on prospects and identification. These basins hold volumes and variety of exploration, appraisal and production data in large quantities. Keeping in view the type, size, volumes and varieties of data, their storage and retrieval demand new guidance in modelling and mapping of multiple data sources existing in multiple domains. The Big Data hype has brought a new direction in the form of an integrated business environment, especially when the entities and dimensions in multiple domains have volumes and varieties of data and they are needed to bring them together in the integrated business. Besides, the heterogeneity and multidimensionality pose data integration challenges complicating the geological and geophysical data processing and interpretation, further affecting the exploration, field development, and production tasks. Adopting Big Data tools and technologies in exploring and establishing the connectivity between elements and processes of multiple petroleum systems is the current focus of research. We investigate the role of Big Data in the integrated E & P business environment. The current study is based on an empirical research, an analytical approach using the existing volumes and varieties of exploration and production data in the NWS.

METHOD AND RESULTS

Analogous to implementation of information system (IS) solutions in healthcare, environment and financial applications (Nimmagadda, 2015b), we propose similar IS solutions with domain ontology descriptions and their integration in the upstream business data management. Integration of shared ontologies is expected to simplify the Western Australian unstructured petroleum resources data and build new knowledge in the unknown geological regimes. For example, a limited E & P success is reported in the onshore and offshore Canning Basin that falls within the vicinity of commercial oil and gas fields with proven discoveries in the NWS. Our shared-ontology approach explores the reasoning of unproductive sedimentary basins, by connecting the petroleum systems of successful basins in various class attributes, sub-classes and their modelling. Considering the existing data and information of E & P of unsuccessful basins is part of our multidimensional data modelling process. The categorizations of data of various elements and processes of petroleum systems can classify and prioritize the areas of E & P for optimizing the investments in the NWS.

A. NWS Ontology Framework

Among several viable alternatives, an appropriate problem solution is analyzed that can work with more flexibility, high-throughput and maintainable data warehouse repositories. Ontologies represented in a real world of petroleum industries, the concepts and contexts must be well understood, described and reflected in the dimensional models. How detailed and fine-grained the data structures are, depend on the descriptions of various conceptual and logical models for building multidimensional warehouse repositories and accommodating them in the petroleum exploration industries. The reason for adopting the ontology descriptions in the petroleum exploration is to explore connections from volumes of data sources and make connectivity between varieties of attribute dimensions. The petroleum ontologies (Chandrasekaran et al. 1999 and Nimmagadda, 2015b) are initial constructs of petroleum domain artefact designs. The ontologies must continue to be revised with updates of petroleum data and information for achieving the life cycle of IS designs (Damiani, 2008) as per life cycle of the oil & gas field. A system development (Damiani, 2008 and Venable, 2006) in the case of NWS is a purposeful artefact, inventing various constructs, models and methods in an integrated framework that suit to NWS geological settings. The petroleum data sources in multiple domains and their attribute dimensions are considered in modelling E & P ventures. The data modelling, warehousing, and mining, visualization and interpretation artefacts are articulated for adding values for the integrated interpretation project. The framework is a formulation of an ontology-based data warehousing and mining (Nimmagadda, 2015a and 2015b and Rudra and Nimmagadda, 2005) approach. It predictably resolves the issues of scaling and formatting the heterogeneous and multidimensional datasets that may have inherited (Nimmagadda and Dreher, 2008 and Nimmagadda et al. 2012) from a large number of domains, basins and their types in the NWS.

B. Big Data Role in the NWS Geo-informatics

Nimmagadda (2015b) describes specific geological events such as the shelf, slope and deep marine in the continental basin areas. These events continue to sustain as long as the connectivity among their data exists. Getting rid of the ambiguity of interpretation by a single dataset, which otherwise cannot interconnect the events from multiple geological regimes calls for new tools and technologies. These events interpreted as dimensions merge in the conceptualization and contextualization of sedimentary basins' attributes and their integration for enhanced understanding of the data connectivity, with the result the basin history and prospectivity of the basin margin areas can become more explicit (Nimmagadda, 2015b). There are multiple domains and attributes in the NWS that needed integration, for which ontologies allow merging of attributes. Welty (2002) defines an ontology as a common vocabulary to explore domains and attributes from which domain experts share information and apply in different contexts. Data integration is a key issue in developing a shared ontology of the NWS. It includes machine-interpretable definitions of basic classes (concepts) in different domains and their respective data relationships to arrive at a meaningful metadata. To analyze the domain knowledge from warehoused NWS metadata, data mining, visualization and interpretation artefacts (Nimmagadda and Dreher, 2009a and Nimmagadda, 2015b) are articulated.

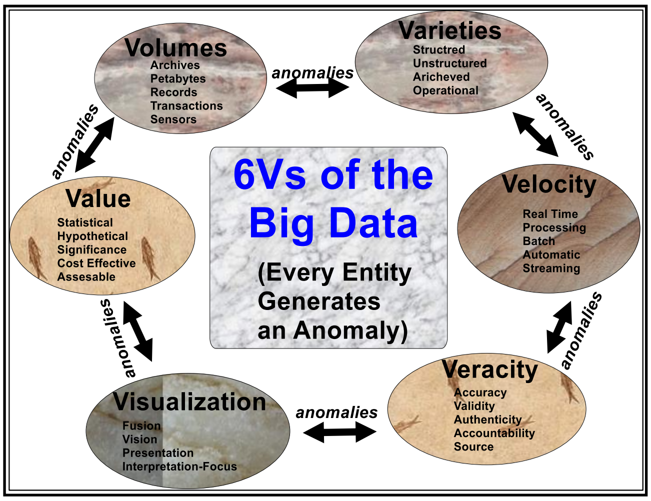

Fig. 1: Big Data characteristics and their modelling framework

The Big Data characteristics described in Fig. 1 are used in the dimensional modelling and integrated with the PDE events as interpreted in Longley et al. (2003) and Nimmagadda et al. (2012). Each Big Data characteristic is influenced by the other (↔). Anomalies generated by these characteristics affect the integrated framework and its components. The data acquisition, dimension categorization, ontology descriptions for every field are described as substantiated in Chandrasekaran et al. (1999). Mapping the dimensions in the multidimensional database management system (MDBMS) and getting the integrated data instances in a warehouse repository are other tasks described in the workflow. The data mining, visualization, and interpretation are the other artefacts from which the knowledge interpreted is documented for identifying the prospective areas. Chen et al. (2012) describe Big Data, the *volumes, variability, velocity, visualization, veracity* and *value* as various attribute characteristics as shown in Fig. 1. We corroborate these attribute dimensions in the artefact design and integrated framework, as detailed in the following sections.

The systems dealing with the Big Data in particular in the context of NWS geo-informatics resolve the fundamental geological problems, owing to the exponential explosion of sequence and structural information with spatial-temporal dimensions (Khatri et al. 2004). The exploration data from different investigation vintages exist, for example, multiple seismic vintages, drilled-wells, permits, and volumes of their linked navigational data in several gigabyte sizes and scales. In spite of major challenges of the data management and knowledge discovery, a large quantity of the production data is underutilized in the integration of the exploration data.

C. Simulation of Digital Petroleum Ecosystems (PDE)

Various petroleum systems and their cognitively connected E&P environments are investigated. As an example, petroleum prospects in the NWS of the Western Australian sedimentary basins are analyzed critically (Nimmagadda, 2015b) using metadata cubes for prospecting. The metadata computed in different geological environments and time periods are analyzed. At specialization data representation level, several issues may have encountered during integrated interpretation of the reservoir, structural and strati-structural plays in different geological settings. The ontology-based data warehousing and mining approach is again an underpinning concept in various application scenarios (Nimmagadda and Dreher, 2009), bringing together various elements and processes of petroleum systems through PDE simulations in the contexts of NWS (Longley et al. 2003 and Nimmagadda and Dreher, 2009b). An integrated framework simulating the PDE in the contexts of NWS is described in Fig. 2, demonstrating the design and development of data acquisition, data modelling, data mining, and visualization and interpretation artefacts. We further intend to evaluate the data models of NWS in the onshore-offshore oil and gas fields of the Western Australian productive sedimentary basins.

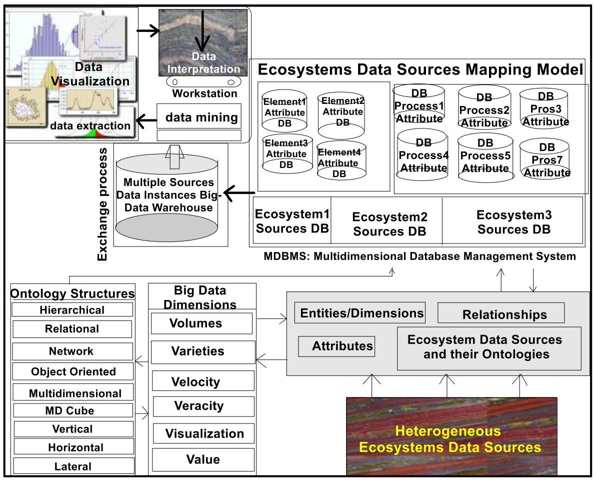

Fig. 2: An integrated framework, connecting data dimensions of structures and reservoirs

D. Big Bata Implementation

For implementing the Big Data, the storage, data processing, and computing requirements are considered. The structured data are commonly stored in the high density and performance warehouse repositories (Nimmagadda, 2015) as a part of implementing the PDE in the Western Australian basins as demonstrated in Fig. 3. The data views are extracted using structured query languages (Chen et al. 2012) or NoSQL (in the case of Big Data with Hadoop technology) and visualization tools of the golden software. The data structures evolve through varieties, multiple data types, and their relationships. We consider the scalability, extensibility and compatibility criteria for the types of structuring, data stores, and retrieval methods, irrespective of architecture, data type, and its structure. The data warehouse includes onsite data processing and analysis, capturing users' data views for visualization and interpretation. Many onsite Workstations (Nimmagadda and Dreher, 2008 and Nimmagadda, 2015) use various computing and processing capabilities. We compute the grids for map views from ecosystems' metadata cubes that may consist of both productive conventional and unconventional reservoirs including information on geological structures. We look at the scope of the same ecosystem that has the potential opportunity of carbon capture and storage (CCS) projects (Nimmagadda et al. 2012). In this context, we capture the data from cloud clusters from multiple oil & gas fields of the NWS that covers geographically spread sedimentary basins in Western Australia.

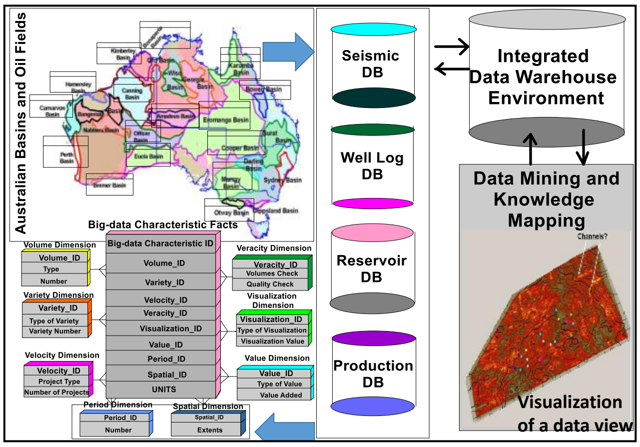

Fig. 3: Digital Petroleum Ecosystem (PDE) framework for implementation in the upstream

E. Integrated E&P Business and Knowledge Management

We use the ontology-based data warehousing for mining the multidimensional warehouse repository (Pujari, 2002) as a part of designing the digital petroleum ecosystems (PDE) for elements and processes of conventional and unconventional petroleum systems in the continental basin margin areas. We examine several data relationships among multiple dimensions, attributes and their instances including units and measures, for modelling various data structures. At places, attribute dimensions are conceptualized and contextualized, thus establishing the connectivity through ontology structuring (Nimmagadda and Dreher, 2009b and Zhong et al. 1996) among various oil & gas fields and basins' data sources. We interpret terabyte size of the Big Data instances for petroleum ecosystem of a basin or groups of basins. At each navigational point dimension, the structure, reservoir, source, seal data attribute dimensions and their instances are interpreted, as demonstrated in the implementation framework in Fig. 3. Extracting useful knowledge on favourable geological structures that trap the productive reservoirs of petroleum ecosystems has significance in the application of the Big Data technology, in particular its implementation for data and business analytics. We use the OLAP tools (Chen et al. 2012 and Rudra and Nimmagadda, 2005) for viewing the data from metadata cubes for visualization. Several data mining operations performed on metadata cubes are shown in Fig. 4.

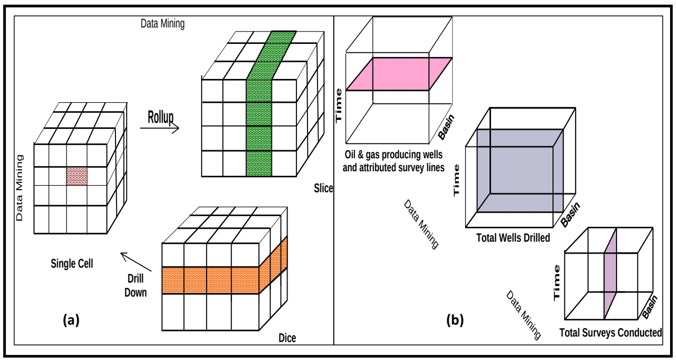

Fig. 4: Multidimensional data mining operations

The data views extracted for exploring the connections are from porosity metadata cubes in an onshore-offshore ecosystem (of NWS) for visualization and interpretation of porosity connections among multiple reservoir systems, as interpreted in a bubble plot view in Fig. 5. We interpret the potential reservoir and production data trends in the bubble plot view in a cluster, drawn from multiple dimensions of the petroleum systems that represent the Big Data. The attributes such as the geological structure, reservoir, production and the number of months produced are used in the interpretation of the bubble plot view. Thicker reservoirs are interpreted where better geological structures and production qualities are interpreted. The orientation and density of the bubble clusters interpreted in large size data of warehouse repositories provide interesting reservoir distributions as described in Fig. 5. The connectivity interpreted between different attributes is useful for NWS joint ventures and exploration investments.

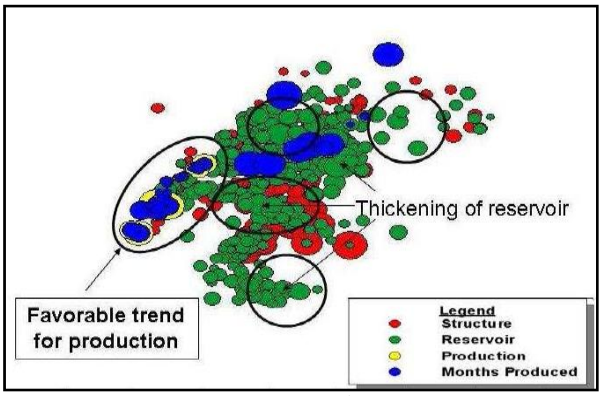

Fig. 5: A data view extracted from metadata cube

CONCLUSIONS

The exploration data are categorized as the Big Data since they are associated with large volumes of navigational and periodic data. The methodology is holistic and effective in integrating and connecting Big Data dimensions and their attributes with data instances of the elements and processes of various petroleum systems of the NWS, WA. The methodology connects both conventional and unconventional reservoir systems such as clastic and limestone reservoirs, fractured-shale networks and tight-gas reservoir systems in a single ecosystem. The framework articulated with the Big Data tools and integration of unstructured data of petroleum digital ecosystems' is acceptable in extracting useful information and adding values to the domain knowledge of the E & P, in particular, the matured fields of the onshore and offshore sedimentary basins. The fine-grained data structuring, mining, visualization and interpretation artefacts, deduced for warehoused E&P metadata analysis demonstrate the implementation of IS solutions in the upstream business and prospecting. Several metadata models are designed with the context of NWS onshore, and offshore basins with new knowledge on the connectivity of petroleum systems through *PMIS* and *PDE* approaches. The data and business analytics are rapidly evolving tools in support of Big Data technology in the integrated upstream business. The size of the Big Data concerning volumes and varieties unquestionably plays a key role in the data analytics, thus impacting the upstream business analytics as well.

REFERENCES

Chandrasekaran, B., Johnson, R. and Benjamins, R., 1999, Ontologies: what are they? why do we need, them?. IEEE Intelligent Systems and Their Applications, 14(1), Special Issue on Ontologies, pp. 20– 26.

Chen, H., Chiang, R. H., and Storey, V. C., 2012, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS quarterly (36:4), pp. 1165-1188.

Damiani, E., 2008, Key note address on 'Digital Ecosystems: the next Generation of Service Oriented Internet", IEEE-DEST, Phitsanulok, Thailand, Feb 2008.

Khatri, V. and Ram, S. and Snodgrass, R.T., 2004, Augmenting a conceptual model with geo-spatiotemporal annotations, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16(11), pp. 1324-1338, doi:10.1109/TKDE.2004.66.

Longley, I.M. C. Buessenschuett, L. Clydsdale, C.J. Cubitt, C.J., Davis, R.C., Johnson, M.K., Marshall, N, M., Murray, A.P., Somerville, R., Spry, T.B and Thompson, N.B., 2003, The North West Shelf of Australia - A Woodside Perspective, M. Keep and S.J. Moss, eds., as Proceedings of the Petroleum Exploration Society of Australia Symposium, Perth WA. Search and Discovery article #10041.

Nimmagadda, S.L and Dreher, H., 2008, Ontology Based Data Warehouse Modelling – a Methodology for Managing Petroleum Field Ecosystems, a paper presented in the *International conference of IEEE-DEST*, held in Bangkok, Thailand.

Nimmagadda, S.L, and Dreher, H., 2009a, Technologies for adaptability in turbulent resources business environments, a book chapter published under a title: Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains, http://www.igi-global.com/, 2009, USA.

Nimmagadda, S.L, and Dreher, H., 2009b, Ontology based data warehouse modelling for managing carbon emissions in safe and secure geological storages, a paper presented in the international *SEG symposium* – Imaging and Interpretation, in a forum "science and technology for sustainable development", held in Sapparo, Japan, Oct 2009.

Nimmagadda, S.L, Dreher, H, Noventianto. A, Mustofa. A and Fiume. G., 2012, On new emerging concepts of Tarakan Sedimentary Basin – a Petroleum Digital Ecosystem (PDE), a paper published in the proceedings of an *International Geological Congress* (*IGC*) held in Brisbane, Australia.

Nimmagadda, S.L., 2015a, Managing the Sustainable Digital Ecosystems using Big Data Paradigm, STAWA, Perth, WA, Australia.

Nimmagadda, S. L., 2015b, Data Warehousing for Mining of Heterogeneous and Multidimensional Data Sources, Verlag Publisher, Scholar Press, OmniScriptum GMBH & CO. KG. p. 1-657, Germany.

Pujari, A.K., 2002, "Data mining techniques", University Press (India) Pty Limited, Hyderabad, India.

Rudra, A. and Nimmagadda, S.L., 2005, Roles of multidimensionality and granularity in data mining of warehoused Australian resources data, *Proceedings of the 38th Hawaii International Conference on Information System Sciences*, Hawaii, USA.

Venable, J. R., 2006, The Role of Theory and Theorising in Design Science Research: In Hevner, A., Chatterjee, S. (eds.) Proceedings of the First International Conference on Design Science Research in Information Systems and Technology, USA.

Welty, C., 2002, Ontology-Driven Conceptual Modelling, invited talk at the Fourteenth International Conference on Advanced Information Systems Engineering (CAiSE), Toronto, Canada.

Zhong, T. Raghu, R. and Livny, M., 1996, An efficient data clustering method for very large databases, *Proceedings of ACM SIGMOD* International conference on management of data, ACM, NY, USA.