Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 1881–1890

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

An Upstream Business Data Science in a Big Data Perspective

Shastri L Nimmagadda*, Torsten Reiners and Amit Rudra

School of Information Systems, Curtin Business School, Curtin University, Perth, WA, Australia

Abstract

The rugged *geographies*, *geomorphologies* and complex *geological* environments make the explorers more challenging exploration and production (E & P). Despite challenges, many sedimentary basins, associated oil & gas fields and E & P Ventures are productive and commercially viable. The difficulty in understanding the connectivity among multiple reservoirs is due to lack of knowledge of multidisciplinary data of petroleum systems, complicating the data integration and interpretation process. The geological and geophysical data of an upstream business are vital assets of any oil & gas industry, in particular in E & P perspective. The data are often unstructured with a variety of anomalous attributes, mingling with volumes of spatial-temporal dimension attributes and instances. In recent years, the concepts of Big Data have taken different hype in petroleum industries, because of involvement of big sized data in the data integration process. Because of the unstructured data sources, a new direction in the database organization is needed. Investigating the science behind the Big Data and their integrated interpretation of the upstream project is a principal objective of the research. In this context, various constructs and models are articulated with different artefacts. Opportunities of Big Data are explored with exploration data and business analytics, supporting sustainable E & P systems. Petroleum management information systems (PMIS) and digital petroleum ecosystems (PDE) are developed to establish a connectivity among various data sources in multiple domains and systems. The implementation of robust methodologies ascertains the significance of the integrated upstream business in the oil and gas industries that comply with the characteristics of the Big Data.

Peer-review under responsibility of KES International

Keywords: Upstream Business; Big Data; Data Science; Digital Ecosystem; Exploration & Production;

1. Introduction

A large amount of heterogeneous and multidimensional E & P data sources of petroleum bearing sedimentary basins is accumulated in many upstream companies and as shown in Fig. 1, they can be represented in digital form[10]. Presently, there is no comprehensive and robust data management and modelling methodologies for managing the E & P data sources of multiple petroleum systems.

* Corresponding author. *E-mail address:* shastri.nimmagadda@curtin.edu.au

1877-0509
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.236

The E & P datasets need a framework and workflow for integrating and risk minimizing the oil and gas businesses, in particular, the Australian contexts. The North West Shelf designated as a total petroleum system (TPS) is termed as the Super Westralian Basin, in which *shelf, slope* and *deep-basin*[10] geological data occurrences seem responsible for the connectedness through a phenomenon, "digital ecosystem"[7, 10]. Also, an assembly of sub-basins, which is a larger part of the super-basin has many petroleum systems, and each system has either known or unknown or incomplete spatial ranges. Each petroleum system may have various oil and gas fields, with hierarchical structuring of attribute data dimensions and their fact instances. Such Big Data[6, 10, 12] sources associated with the western part of Western Australia (Fig. 1) have a choice of analysing their attribute dimensions and instances in multidimensional warehouse repositories through different constructs and models.

For the purpose of data integration, various onshore and offshore data sources[10, 12] are described. The *shelf, slope* and *deep-basin* geological data occurrences of the continental basin margin are characteristic in any geological settings, where large volumes and varieties of data exist. As demonstrated in Fig. 1a, the shelf, slope and deep data events of the continental basin margin are typical in any geological setting and petroleum system development, where volumes and varieties of data exist. A scope of connecting and integrating the Big Data through domain ontologies is explored. Figs. 1b and 1c represent the digital data of the Western Australian upstream business assets. The data modelling aspects, including the description of domain ontologies of the multiple dimensions, are depicted.

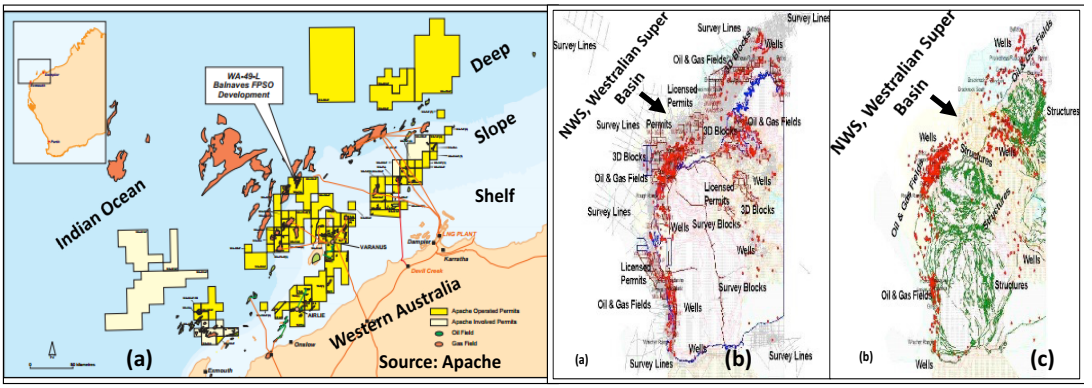

Fig.1: (a) *Shelf*, *slope* and *deep* (geological) data events of the NWS (b) Western Australian assets (c) Oil & gas fields

The description of NWS and the significance of its upstream business data science are discussed in Section 2. Description of ontologies and modelling methodologies are given in Section 3. In the context of NWS in a Bid Data scale, the data mining and visualization models are discussed in Section 4 with certain conclusions and recommendations in Section 5.

2. North West Shelf (NWS) and the upstream business data science

The collective strength of the *structure, reservoir, source maturity, seal,* and *migration* of hydrocarbons, the *timing* of deposition including their *accumulations,* constitutes the existence of a petroleum system in a sedimentary basin. They are characteristically the elements and processes of the petroleum system. The total package of the petroleum system consists of various elements and processes with a variety of their data instances and strengths. The existence of petroleum systems is ensured, and if any of the elements and or processes is missing, it is impossible to expect the oil/gas in that basin. The explorers invest in the upstream business, ensuring the presence of petroleum system and a comprehensive knowledge of data of all elements including the processes of the petroleum system. We attempt to envisage the petroleum system in a different perspective using ontology-based data warehousing and data mining technologies[10], for exploring the capability of the system. In the present study, all the elements and processes are considered as classes of a super class *petroleum system*. In the "exploration" super class, the *surveys*, *wells* and *permits* are the other representative classes. In the E & P class representation, the *production* is a key sub-class. In prospect analysis viewpoint, conceptual models are built, creating data relationships through ontological descriptions. Integrating domain ontologies of representative dimensions of multiple domains is a requirement, as demonstrated