A cloud-based Well Log Database Prototype

Chitra Viswanathan*
CSIRO
Chitra.Viswanathan@CSIRO.AU

Irina Emelyanova
CSIRO
Irina.Emelyanova@CSIRO.AU

Ben Clennell
CSIRO
Ben.Clennell@CSIRO.AU

Stacy Maslin
CSIRO
Stacy.Maslin@CSIRO.AU

SUMMARY

Geoscience data including seismic, well log, sensor and core measurements are fundamental for Petroleum exploration. Due to recent advancements in sensor and computer technologies, the volume of this data is constantly increasing. Having a unified repository of this data of various types, structure and complexity (Big Data) is crucial for maintaining data integrity.

This study addresses petroleum exploration data integrity issues. Current trends in data management technologies and current data practices in Petroleum Geoscience are explored and a practical data management solution to facilitate data access, storage and sharing is recommended. A prototype of a well log database was developed to demonstrate the advantage of having a unified repository of downloaded and sanitized data for multiple users. In comparison with current practices, such a database will prevent duplicate downloads from public websites by petro physicists and make data use more efficient within a particular organisation.

The prototype was developed using cloud-based technology and the PAWSEY supercomputing facility (a joint venture of CSIRO with Western Australian universities) for storing both the raw (.las and .DLIS files.) and the sanitized well log datasets from Bonaparte Basin. PostgreSQL database was used to store the sanitized well log data, metadata and links to raw data. PostgreSQL architecture was selected for its ability to support advanced data types (arrays, JSON etc.), plug in to languages like Python, and link to PostGIS, a spatial database extender. A web-based graphical user interface was developed to view, upload and download well log data. In addition meaningful metadata standards were established in collaboration with expert petro physicists.

Key words: Geoscience data, Data management, Metadata, Well log, Big Data, PostgreSQL

INTRODUCTION

Most industries currently face the problem of storing, sharing and harvesting useful information from the large volumes of data that they are ever increasingly collecting. These large volumes of data, generally in different formats like documents, images, videos and traditional record sets, are collectively termed “Big Data”. Big Data is characterised by 3Vs, volume, variety and velocity. In the scientific domain, several disciplinary areas are facing Big Data challenges as part of an innovative approach to science, with Geoscience as being one of them.

Geoscience data are based on observations and measurements coming from in situ and remote-sensing data with ever-growing spatial, temporal, and radiometric resolution, requiring handling of big volumes (Peter Baumann, 2016). They encompass both structured and unstructured data in the form of maps, images, spatial data, array based data (variety) and they are often real time data demanding fast processing (velocity). Data management is a crucial aspect in the field as huge volumes of data are generated from geoscientific observations and model simulations. Use of a centralised repository equipped with uniform and meaningful metadata make it easier to access and share data for research. The goal of the data management system is to provide a platform for collecting, storing, and sharing monitoring data within a larger network of data providers and end users (Iwanaga, El Sawah, & Jakeman, 2013) One of the major design features of sound data management is having a master set of data stored centrally and managed with adequate security permissions.

In this paper we addressed the Big Data challenge faced by a Petroleum Geoscience research team within CSIRO (a target team, hereafter) by building a well log database (WLD) prototype using the cloud technology. The WLD would dramatically cut duplicate data downloads, provide easy access to the downloaded data and promote data sharing.

Modern Geoscience data management approaches were reviewed and compared with the data storage practises in the target team. Currently, the team downloads well logs from public databases such as WAPIMS (Western Australian Petroleum and Geothermal Information Management System), and the analogous National Offshore Petroleum Information Management System (NOPIMS) administered by Geoscience Australia. Extensive data quality control (QC) is required before importing to a custom software for processing. Our studies showed that lack of communication between the researchers often leads to multiple downloads of same log data which increases total computing time and data processing cost. The team raised a number of concerns including limited data sharing, hard to access to data, limited storage and lack of meaningful metadata.

We performed an audit of the available data storage facilities within CSIRO and affiliated bodies to assess whether existing data infrastructure can be utilized for building of a unified Petroleum data repository. Also, several database architectures such as HBase (A. B. M. Moniruzzaman and S. A. Hossain, 2013) and MongoDB (A. B. M. Moniruzzaman and S. A. Hossain, 2013) which are NoSQL (Not Only SQL) databases were investigated as they are specifically designed to perform data analytics on Big Data. The PostgreSQL platform was also examined as it is an open source database platform and provides the benefits of both traditional relational databases and the newer non-relational (NoSQL) databases.

We built the WLD prototype, a cloud-based database system to store and access sanitised well log data and vast amounts of raw data from various well logs that are accessible over the internet using a user friendly web based GUI.

DESIGN AND IMPLEMENTATION

The primary design strategy has been to involve the end users of the system, petro physicists and geoscientists in the development of the WLD prototype. The requirements were to provide:

- User friendly GUI to upload and share the sanitised well log data,
- Storage for the bulky raw data from various logs and associated reports for every well,
- An easy access to raw data,
- Ability to query the database as to what sort of data is available for each well.

The graphical user interface (GUI) for the WLD prototype is a web user interface created using ASP.Net platform. We chose PAWSEY supercomputing facility (a joint venture of CSIRO with Western Australian universities) for WLD raw data for the following reasons:

- Cost-free service,
- Big volume data storage,
- Upload and download of raw data made simple through windows explorer like interface,
- Use of metadata with files,
- Help from PAWSEY data scientists' team.

We also applied for PAWSEY's cloud services (NIMBUS) that are available to any researcher in Australia and have acquired these services. These services provide faster access to PAWSEY storage facility and also toolset for performing data analytics.

Based on the database architecture review, we chose PostgreSQL as the database platform for developing the WLD prototype. The PostgreSQL architecture was chosen for its ability to support advanced data types (arrays, JSON etc.), plug in to languages like Python, and its link to PostGIS, a spatial database extender.

PostgreSQL 9.6 server was set up and a well log database was created. We defined the well log parameter names in consultation with the Petro physicists. Table 1 gives the parameter (also called as curves) names that are built into the database as an initial, simplified set of commonly available petro physical logs. The original tool mnemonic is captured in the description field for future reference.

Name	unit	Description
Depth	m	Depth (drillers depth in meters below rig floor)
GR	API	Natural gamma ray activity
K	%	Formation potassium content from spectral gamma ray
Th	ppm	Formation thorium content from spectral gamma ray
U	ppm	Formation uranium content from spectral gamma ray
BS	in	Borehole size (smooth hole diameter produced by drill bit)
CAL	in	Caliper (hole diameter)
SP	mV	Spontaneous potential between ground and electrode at depth
Rxo	ohm.m	Resistivity of zone invaded by mud filtrate
Rs	ohm.m	Shallow Resistivity of formation close to borehole
Rm	ohm.m	Medium Resistivity for formation beyond shallow zone
Rt	ohm.m	Deep or true resistivity of formation
Rmf	ohm.m	Resistivity of mud filtrate
RHOB	g/cc	Formation density from gamma-gamma absorption lithodensity tool
ΔRHO	g/cc	Density measurement correction
PEF	B/e	Photoelectric factor (barnes per electron) from lithodensity tool
NPHI	v/v	Neutron porosity, limestone basis standard
DTC	us/ft	Delta-T Compressional (sonic travel time of compressional wave)
DTS	us/ft	Delta-T Shear Slowness (sonic travel time of shear wave)
DTST	us/ft	Sonic travel time of Stoneley wave

Table 1 – List of basic well log fields

Standard well logs have measurements six inches but there can be non-standard curves (either higher resolution, or more widely spaced readings) with different depth intervals. For many such logs, measurements can be computed at the standard six inch spacing and added to the well log basic data. For this reason, in addition to the fields in Table 1, extra columns were built into the table to store these additional curves. An example of such an additional field would be Nuclear Magnetic Resonance (NMR) porosity or dielectric constant.

We developed a mapping table to link the various nomenclature used by different logging companies for every log parameter with a single field in the database. For example, Gamma Ray from basic log and API standardized total gamma from spectral gamma ray are mapped to the same field "GR", but the original mnemonic was always recorded in the description field for future reference.

The raw data from complex downhole measurement tools that produce large data arrays, such as sonic waveforms, image logs, and NMR are stored in PAWSEY facility and the pointers to the files are stored in the database. Information about what types of data are available for a particular well are captured at the well level so that queries can be made. Other types of well data that are not depth-based log records include vertical seismic profiles (VSPs), pressure data, and written reports in pdf format. Again these raw data are stored at PAWSEY and the pointers are stored in the database. Well header data using the standard CWLS .LAS 2 or LAS 3 headings are stored as metadata.

The workflow of the prototype is as follows (See Figure 1):

1. Raw log data files are downloaded from public websites such as WAPIMS or GA and
2. Data QC is performed on them using third party software tools,
3. These sanitised files and associated metadata are then uploaded to the PostgreSQL database using the web user interface,
4. The raw data files along with the associated metadata from various logs are stored in PAWSEY facility,
5. The web GUI is used to download, view, append and share the sanitised well log data.

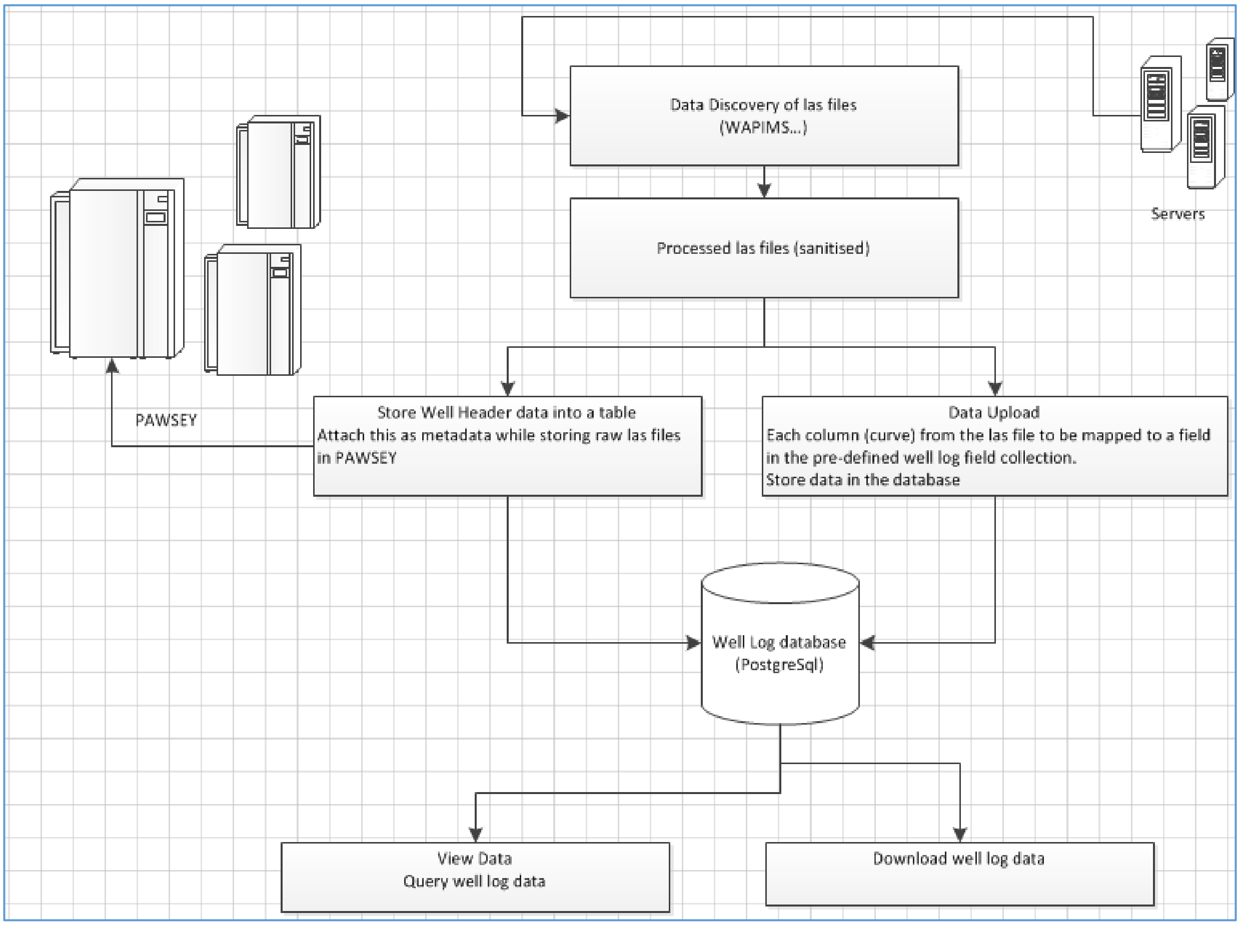

Figure 1 – Schematic of the Well Log Prototype Implementation

CONCLUSIONS

Development of the well log database (WLD) prototype using the latest data management technologies has addressed the data integrity issues in the target Petroleum Geoscience research team within CSIRO. It has curtailed duplicate data downloads from public websites. Selection of the design features was heavily influenced by the available hardware / software infrastructure within CSIRO and affiliated facilities such as PAWSEY supercomputing facility.

It is anticipated that this will be an ongoing project expanding to include other types of data such as seismic in the future. In this phase, the PAWSEY supercomputing facility was used mainly as a back end database for bulky unprocessed well log files. PAWSEY supercomputing facility also offers cloud services (NIMBUS) that is specifically designed for research applications. To take this data management solution to the next phase, we applied for and secured NIMBUS services. This would facilitate performing data analytics and visualisation. The data management technologies used in this WLD prototype to address the data integrity issues in a research organisation are easily transferable to manage the needs of industry as well.

REFERENCES

A. B. M. Moniruzzaman and S. A. Hossain, "NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison," ArXiv13070191 Cs, Jun. 2013.

Hsu, L., Martin, R. L., McElroy, B., Litwin-Miller, K., & Kim, W. (2015). Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities. *Geomorphology, 244*, 180-189, doi:https://doi.org/10.1016/j.geomorph.2015.03.039.

Iwanaga, T., El Sawah, S., & Jakeman, A. (2013). Design and implementation of a web-based groundwater data management system. *Mathematics and Computers in Simulation, 93*, 164-174, doi:http://doi.org/10.1016/j.matcom.2012.11.009.

Kingdon, A., Nayembil, M. L., Richardson, A. E., & Smith, A. G. (2016). A geodata warehouse: Using denormalisation techniques as a tool for delivering spatially enabled integrated geological information to geologists. *Computers & Geosciences, 96*, 87-97, doi:https://doi.org/10.1016/j.cageo.2016.07.016.

Li, Z., Yang, C., Jin, B., Yu, M., Liu, K., Sun, M., & Zhan, M. (2015). Enabling Big Geoscience Data Analytics with a Cloud-Based, MapReduce-Enabled and Service-Oriented Workflow Framework. *PLOS ONE, 10*(3), e0116781, doi:10.1371/journal.pone.0116781.

Peter Baumann, P. M., Joachim Ungar, Roberto Barbera, Damiano Barboni, Alan Beccati, Lorenzo Bigagli, Enrico Boldrini, Riccardo Bruno, Antonio Calanducci, Piero Campalani, Oliver Clements, Alex Dumitru, Mike Grant, Pasquale Herzig, George Kakaletris, John Laxton, Panagiota Koltsida, Kinga Lipskoch, Alireza Rezaei Mahdiraji, Simone Mantovani, Vlad Merticariu, Antonio Messina, Dimitar Misev, Stefano Natali, Stefano Nativi, Jelmer Oosthoek, Marco Pappalardo, James Passmore, Angelo Pio Rossi, Francesco Rundo, Marcus Sen, Vittorio Sorbera, Don Sullivan, Mario Torrisi, Leonardo Trovato, Maria Grazia Veratelli, and Sebastian Wagner. (2016). Big Data Analytics for Earth Sciences: the EarthServer approach. *International Journal of Digital Earth, 9*(1), 27.