AAPG Datapages/Search and Discovery Article #90319 ©2018 GEO 2018 13th Middle East Geosciences Conference and Exhibition, Manama, Bahrain, March 5-8, 2018

Introducing a Big Data System for Maintaining and Controlling SEG-Y Data Quality and Integrity in a Continuously Growing Digital Environment

Abdulelah Bin Mahfoodh[1], Mohamad A. Ibrahim[1], Maan Hawi[1], and Nasir Seadi[1]

[1]EASD, Saudi Aramco, Dhahran, Eastren Province, Saudi Arabia.

ABSTRACT

In an era of digital fields, oil and gas industry is constantly requiring seismic survey data in quest to reveal underlying geophysical and geological structures. There exists many data formats to store and exchange seismic surveys data but the most popular one is SEG-Y format which was developed in 1975. Typically in an oil and gas industry, there are hunders of thousands SEG-Y files of different types residing on live disks and ranging between few megabytes of size to hundreds of gigabytes for each file. The presence of minimum set of information in correct format within each SEG-Y file is essential for further processing seismic data. Unfortunately, SEG-Y files that are said to be recorded in accordance to the standard, erroneously sometimes, deviate from it due to various reasons such as acquisitions software, transmission, human factors, etc. This deviation from standard format makes it difficult for SEG-Y files to be processed further. In order to address the before mentioned impasse, we present a novel system that aims to verify and correct SEG-Y files based on automatically generated quality reports using big data techniques. In recent days, big data solutions have advanced to process large volumes of data in order to improve its quality at almost near real time. The proposed solution will process a given SEG-Y file, validates its header, trace headers and data blocks by iterating over the seismic survey file trace by trace and perform sets of mathematical and logical operations. This way one can only assure the target SEG-Y file is following the standard format and maintaining evergreen quality status. The introduced system is written in Python and utilizes a big data framework called “DASK” for parallel and distributed processing of traces over a number of computers to gain better performance. Through a set of predefined business rules, the system will read data from SEG-Y EBCIDC header making sure the minimum set of information required exists. Moreover, the solution will process, in parallel, the SEG-Y file traces calculating its actual inline, xline, x and y values and comparing against data provided in the trace header. The final output of the system is a data quality statistics report along with a plot showing each trace location as well as actual and header survey corner points for comparison purposes. Finally, the showcase will conclude by an overall assessment on functionality and performance aspects of the introduced system.