

Final Report - P2

XBRL

Shangwen Yan | N17091204 | sy2160

Introduction & Background

XBRL is the open international standard for digital business reporting, managed by a global not for profit consortium, XBRL International. We are committed to improving reporting in the public interest. XBRL is used around the world, in more than 50 countries. Millions of XBRL documents are created every year, replacing older, paper-based reports with more useful, more effective and more accurate digital versions.

In a nutshell, XBRL provides a language in which reporting terms can be authoritatively defined. Those terms can then be used to uniquely represent the contents of financial statements or other kinds of compliance, performance and business reports. XBRL lets reporting information move between organisations rapidly, accurately and digitally.

WHY

Describe the problem that exists

With the fast development and wide use of electronical equipment, the magnitude of dataset is becoming so voluminous and complex that traditional data-processing application software are inadequate to deal with them. Also, people need a standard rule that the industry would follow to make the exchange of data easier and faster.

WHAT

Identify through literature search what solutions have been proposed

Typically, XBRL could simplify the way that people can use, share, analyse and add value to the data.

The change from paper, PDF and HTML based reports to XBRL ones is a little bit like the change from film photography to digital photography, or from paper maps to digital maps. The new format allows you to do all the things that used to be possible, but also opens up a range of new capabilities because the information is clearly defined, platform-independent, testable and digital. Just like digital maps, digital business reports, in XBRL

format, simplify the way that people can use, share, analyse and add value to the data.

HOW

how to apply advanced technologies such as Spark to process XBRL

A large proportion of enterprise data is locked away in complex and verbose industry data standards (ACORD, FpML, HL7, ISO 20022, XBRL etc.) or other proprietary formats based on XML/JSON. Standard ETL tools do not work well to unlock this data for data analytics. They do not scale, don't perform well, and they don't handle change gracefully. Besides, traditional ETL development takes forever to create the required data pipelines. Flexter (a distributed big data solution from Sonra) has solved this problem with Apache Spark and completely automated the process of converting complex XML/JSON into text, a relational database, or Hadoop. In this talk, we will describe how we solved the problem of processing complex XML files (modeled after ACORD Insurance Industry standard) at Cincinnati Insurance Company. Cincinnati Insurance Company is a subsidiary of Cincinnati Financial Corporation, a FORTUNE 500® company and included in the 2017 FORTUNE 500 list of the largest U.S. companies. We will walk you through the architecture and go into the technical details of Cincinnati Insurance Company's data warehouse solution, which runs on Hortonworks HDF/HDP platform enabled by Flexter which runs on Spark and is written in Scala. Along the way we will describe the problems we have solved with our existing data pipeline and how we made our platform more flexible in terms of ingesting different types of XMLs originating from our operational systems. We will also provide an architectural inner workings of the Flexter tool for XML/JSON processing.

To implement parallization, data need to be stored in distributed system, FIOS is one of them. As shown in the picture below, FIOS' architecture is separated into two types of components, the offline ETL components and online analysis components, described in the following.

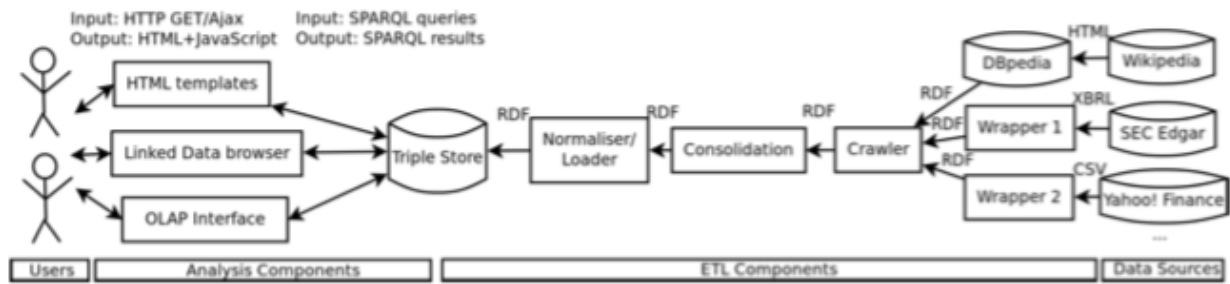


Fig. 1. Flow diagram illustrating architecture of Financial Information Observation System (FIOS)

FIOS uses the Linked Data principles to identify and retrieve relevant information spread across different web servers.