Candidate: Nicholas Talian
Position: IMF Economic Data Architect
Date: October 28, 2015

Essay 1: Big Data and Advanced Search Analysis at the International Monetary Fund

Data sources have grown increasingly large and complex at many international institutions over the past decade, and the IMF is no exception. By applying advanced search algorithms, pattern recognition, and analytical software methods to such datasets, IMF economists can continue to glean valuable insights about the current state of the global economy.

Big Data Analytics using Alteryx

Historically, one of the major impediments to creating robust software tools for end-users has been scheduling delays during the software development lifecycle. Information technology projects often suffer from scope creep, spiraling budgets, and waning interest as timelines slip and requirements change. To mitigate these risks, rapid prototyping and agile software methodologies may be of particular interest to the IMF.

One specific area that the IMF may wish to explore is the use of Alteryx. Alteryx is an affordable software package based on R, a programming language and software environment for statistical computing and graphics. One of the key benefits of Alteryx is that it provides users the ability to rapidly create custom workflows and search parameters for merged datasets. Unlike many custom developed software systems, which can take weeks or months to go live, and require expensive or high-maintenance computer clusters, Alteryx users can produce useful output in a matter of hours, using basic desktop systems. In an era of scarce resources the benefits of highly efficient and productive software tools such as Alteryx cannot be overstated.

During a recent audit client engagement, I used Alteryx to perform data mining and analysis of datasets larger than 10 Gigabytes (GB). Working independently, I rapidly developed a custom workflow that extracted salient data fields and searched for anomalies and irregularities in the data that would likely trigger audit findings. I used the data collected from Alteryx to provide substantive evidence and support when reporting my findings and remediation recommendations to the client. Perhaps a similar approach could be used at the IMF.

Data Mining and Search using IBM Watson Analytics

Another software product worthy of consideration is IBM's Watson Data Analytics engine. Based on the software that successfully competed in the quiz game "Jeopardy!" in 2011, IBM's Watson Analytics software allows users to import their own databases and provides an intuitive user interface for exploring, querying, and visualizing complex datasets. A smart data discovery

service available in the cloud, it guides data exploration, automates predictive analytics, and enables effortless dashboard and infographic creation. I have personally experimented with the IBM Watson service and believe it to be a promising new technology. Using tools such as Watson would quickly place the IMF in the vanguard of modern econometric analysis and could enable IMF economists to rapidly obtain publishable results. With appropriate training and technical support, these tools can allow the IMF to build a plethora of new reporting capabilities.

Human Intelligence, Technical Expertise, and Language Abilities

Any research and development project should take into consideration several key factors (Fig. 1). Ultimately, what is needed is a cadre of experienced professionals – technically savvy interdisciplinary experts who can effortlessly and seamlessly conduct business in both the technical realm and the world of international economics. Although great strides have been made by software firms in the domain of automated data analysis, there is still no substitute for human expertise. My dual background in financial accounting and software development makes me uniquely qualified to translate between the languages of computer science and economics. I am capable of fundamentally understanding the business needs of the organization, evaluating new technologies, and providing original research and recommendations to upper management. I have a proven track record of managing large projects and delivering technical solutions to complex challenges in governmental organizations. Finally, my foreign language skills (Spanish and Italian) serve to enhance my value as human capital working within an international organization, and my dual ethnic heritage (half Peruvian / half American) gives me a unique perspective on matters of international consequence.

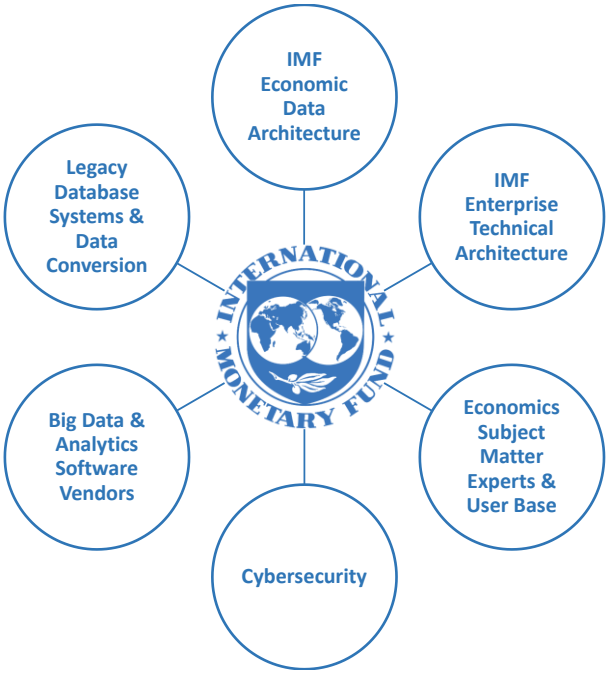

Figure 1. – Key influencing factors and considerations for Big Data Analytics projects at the IMF

Essay 2: Data Management Standards at the International Monetary Fund

Standardizing data management in an enterprise-wide environment such as the IMF has many potential benefits, such as improving collaboration, enhancing efficiency, and reducing costs. IMF's recommended data management standards should account for the following:

1. Structured Data:

- **Database element naming conventions.** Within a traditional relational database, data elements are logically organized into related tables and columns, thereby reducing redundancy and increasing coherence. IMF's database standards should include recommendations for naming commonly used database fields. For example, multiple IMF databases might contain statistical data measuring global economic performance. An appropriate data management standard would publish standardized definitions for key elements that are likely to be reused throughout the organization. The following table demonstrates a possible data management standard, also known as a Data Dictionary:

Standardized Data Element Name	Definition	Datatype	Field Width	Nullable
ECON_GRP	Standard Economic Groupings, such as G7, G20, Emerging and Middle-Income economies.	Alphanumeric	50	Null
COUNTRY	Name of the country	Character	100	Not Null
GDP	Gross domestic product is the most commonly used single measure of a country's overall economic activity. Expressed in current U.S. dollars.	Numeric	30	Not Null

Table 1. – Prototypical data management standards

- **Awareness and compliance with pre-existing international data standards for economic data.** Any new data standardization initiative should take into consideration pre-existing standards for the dissemination and sharing of financial data within and among IMF member nations. For example, the following standards have already been published and should be taken into account when developing new standards:

    - Special Data Dissemination Standard (SDDS / SDDS Plus)
    - Enhanced General Data Dissemination System (e-GDDS)
    - IMF Data Quality Reference Site (DQRS)
    - IMF Data Quality Assessment Framework (DQAF)
    - eXtensible Business Reporting Language (XBRL)

- **Well-defined interfaces and data interchange formats.** For systems which must load external data sources, or provide periodic outputs for external consumption, I recommend interfaces between systems be well-defined and documented via Interface Control Documents (ICDs). I also recommend the IMF reduce overreliance on customized Extract-Transform-Load (ETL) software by standardizing on non-proprietary file formats. Data management standards should include recommended and permitted file formats, which may include the following:

    - **CSV** (Comma Separated Value)
    - **XML** (Extensible Markup Language)
    - **XLS** (Microsoft Excel Spreadsheet)

2. Semi-Structured and Unstructured Data:

Text documents and other files that contain unstructured data may be sorted and catalogued by means of electronic document repository, enterprise content management, and collaboration software suites such as:

- Microsoft SharePoint
- OpenText
- EMC Documentum

The IMF should evaluate such commercial off-the-shelf (COTS) software for suitability within its overall technical architecture, on the basis of the following criteria: search, retrieval, version control, security, configurability, customization, cost, and stability. During the course of my career, I have evaluated and implemented similar software solutions using SharePoint and OpenText.

Other relevant data standards would also include: document approval workflows, taxonomies for document classification, and basic metadata such as content author and departmental affiliation. Finally, identification and archival of documentation that is outdated will prevent the document repository from becoming cluttered with irrelevant or obsolete information.

3. Anticipated Challenges

It may prove difficult to convince IMF's departments to adopt new data management standards given that they tend to operate relatively autonomously of one another. Individuals working within one subject matter domain may not wish to burden themselves by voluntarily complying with these standards.

To address these challenges a proactive managerial approach is suggested. Any standardization initiatives should be sponsored from the outset by a senior executive within the

IMF and should be accompanied by a strongly worded policy directive articulating the logic behind the initiative and the benefits that will accrue to the entire organization. This will guarantee that compliance with these standards shall be given the necessary level of consideration and respect by all department heads.

Furthermore, I would recommend that an executive steering committee be established, with hand-picked representatives from each department. This will ensure there is adequate participation with input solicited from various stakeholder groups, senior leadership, and subject matter experts, while also providing a suitable forum for conflict resolution during the implementation process. Voluntary compliance with IMF data standards should be commended and appropriately recognized.

In conclusion, data management standards are an essential aspect to ensuring data are compatible, searchable, and usable across disparate datasets and user groups within the International Monetary Fund. These standards, when coupled with new software tools and processes, will enable IMF economists to aggregate and digest an increasingly wide range of macroeconomic factors, thus creating new trend analyses and global economic forecasts (Fig. 2). The concepts outlined here provide a general framework for implementing a set of lightweight and portable standards across the enterprise without creating excessive overhead.

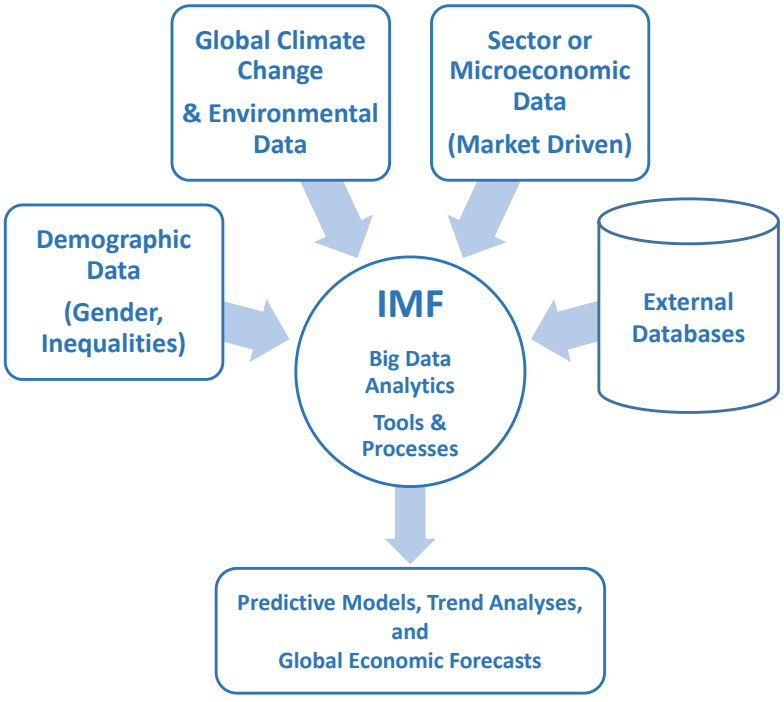

Figure 2. – Solution concept diagram for Big Data Analytics at the IMF

References

- Alteryx.com, (2015). ***Data Blending and Advanced Analytics Software - Alteryx Analytics***. [online] Available at: http://www.alteryx.com/ [Retrieved on Oct. 28, 2015].

- IBM Watson Analytics, (2015). ***Guided and automated analytics from the cloud.*** [online] Available at: http://www.ibm.com/analytics/watson-analytics/ [Retrieved on Oct. 28, 2015].

- Xbrl.sec.gov, (2015). ***eXtensible Business Reporting Language***. [online] Available at: http://xbrl.sec.gov/ [Accessed Oct. 28, 2015].

- Dsbb.imf.org, (2015). ***IMF – Dissemination Standards Bulletin Board – Home Page*****.** [online] Available at: http://dsbb.imf.org/ [Retrieved on Oct. 28, 2015].

- Imf.org, (2015). ***IMF Data Standards & Codes***. [online] Available at: http://www.imf.org/external/data.htm [Retrieved on Oct. 28, 2015].