SOLVING THE BIG DATA CHALLENGE

Optimizing Advanced Analytics for Today’s Digital Government Organizations

By Dr. Rod Fontecilla

A Digital Government Series by Unisys Corporation
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>2.0</td>
<td>Data Analytics</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>Challenges</td>
<td>3</td>
</tr>
<tr>
<td>2.2</td>
<td>Concepts and Capabilities</td>
<td>4</td>
</tr>
<tr>
<td>2.3</td>
<td>Evolution/Road Map</td>
<td>5</td>
</tr>
<tr>
<td>3.0</td>
<td>Implementation</td>
<td>6</td>
</tr>
<tr>
<td>3.1</td>
<td>Methodology and Framework</td>
<td>6</td>
</tr>
<tr>
<td>3.2</td>
<td>Reference Architecture</td>
<td>8</td>
</tr>
<tr>
<td>3.3</td>
<td>Platform Offerings</td>
<td>8</td>
</tr>
<tr>
<td>4.0</td>
<td>Use Cases</td>
<td>9</td>
</tr>
<tr>
<td>4.1</td>
<td>General Services Administration</td>
<td>9</td>
</tr>
<tr>
<td>4.2</td>
<td>United States Department of Agriculture</td>
<td>10</td>
</tr>
<tr>
<td>5.0</td>
<td>For More Information</td>
<td>11</td>
</tr>
</tbody>
</table>
1.0 INTRODUCTION
Along with tremendous opportunity, today’s rapid growth of data also presents agencies with a daunting challenge that is pushing the limits of existing Information Technology (IT) infrastructure. Agencies hoping to reap the proven value of big data through effective use of their IT investments must contend with a wide range of new and rapidly-shifting technologies. In fact, the majority of modern data – up to 80-percent, by some industry estimates – is now unstructured, file-based and resistant to traditional analytic processes.

To face what we call this “big data challenge,” agencies must have access to highly skilled data scientists, advanced tooling and the sophisticated domain expertise needed to unlock value from complex and massive floods of data. Further, agencies need to deploy these resources without the burdensome expense of building massive IT infrastructure from scratch and recruiting an entire team of analytics employees.

This white paper examines that big data challenge and shares the Unisys perspective on optimizing Advanced Data Analytics to solve many of the problems this challenge creates. We blend public sector context and some real world case studies with a discussion of our own Advanced Data Analytics solutions, which combine comprehensive data analytics tooling with disciplined, enterprise-scale engineering approaches.

In the following pages, we’ll discuss Unisys Advanced Data Analytics solutions in considerable detail, given that our own approach is derived from more than four decades of experience across vertical markets – including the government, financial services and commercial sector. Over that period, we’ve developed best practices and time-tested methodologies, many of which have been adopted industry-wide. We hope a thorough discussion of our own successful approach to some of the most critical and complex data challenges can embolden a wide range of government agencies to embrace analytics for a data-driven and prosperous future.

Unisys experts leverage proven, real-world strategies to augment and enhance an agency’s existing IT capabilities with analytics over multiple platforms and devices. Our own successful answer to the big data challenge is an end-to-end data analytics solution that enables our government customers to make vital insights and timely, well-founded decisions efficiently and effectively to better serve the American public. We tailor solutions to agencies’ specific needs, combining a proven analytics platform with skilled data scientists who partner closely with agency teams to solve their most challenging analytics problems.

Unisys has refined a unique consumption-based model that allows agencies to right-size their analytics projects and fill performance gaps involving skills, technology or IT infrastructure. Our government customers pay for what they use, not what they may use; this allows for customized and affordable analytics efforts that follow a natural and agile course of iterative knowledge discovery.

2.0 DATA ANALYTICS
Data analytics involve a wide range of techniques to derive actionable information from available data. The traditional Business Intelligence approach focuses primarily on retrospective analysis – descriptive information on prior activities and trends. Modern analytics techniques, by contrast, leverage huge volumes and varieties of structured and unstructured data at very high speeds for predictive analysis to accurately forecast trends, identify significant events and discover underlying patterns. This all contributes to a culture of data-driven decision-making that can vault agency operations to new levels of performance and mission success.

2.1 CHALLENGES
Creating a data driven culture, however, is easier said than done. Agencies face many challenges as they work to leverage the data available to them. These include:

• Too much data to process – Advances in IT and consolidation of systems and architectures create vast amounts of data. Agencies typically encounter problems trying to process all that information.
• Not enough trained staff to analyze data – As data grows in volume and complexity, value can be lost without the capacity and expertise to examine that information for insights and patterns. Typical business analysts focus on specific types and scopes of data. They don’t have the skills and experience to understand the analytic potential within the volumes and varieties of available data.

• Inability to securely access and process data – Enterprise infrastructure often lacks connectivity and capacity to provide broad, yet secure data access. Among other things, modern architectures must provide authorized, secure access via mobile devices, so the right people can access data when and where they need to.

• Lack of analytics tools – Traditional analysis methods and tools can’t effectively ingest and correlate data as it comes in multiple formats and levels of structure. Investigating, selecting, acquiring and integrating analytics tools requires specific expertise and the ability to correlate disparate data sets such as email communications, asset management, events, etc.

• Budget limitations – Agencies work under constant budget pressures. Analytics can bring value to an agency, but not before a period of investment and capacity-building that must be done in the most strategic and cost-effective way possible.

Unisys identified these challenges as the most commonly-cited among clients who seek us out for big data analytics support. To overcome these challenges, we’ve prioritized certain concepts and capabilities – proven techniques for agencies to exploit the potential of data analytics.

2.2 CONCEPTS AND CAPABILITIES

To satisfy client demands and achieve maximum benefit from advanced analytics, Unisys has embraced a holistic view of Enterprise Data Management that encompasses the entire value-generating lifecycle around data. This includes:

• Data Integration – Smooth operations and high performance rely on effective management of data through a number of processes:
  - Data Migration
  - Data Consolidation
  - Data Profiling
  - Data Cleansing
  - Data Provenance
  - Master Data Management

• Data Visualization – Visualization is a critically important capability that makes data and insights more accessible and actionable across an agency. Multiple users and analysts share visualizations to quickly and accurately identify valuable patterns, relationships and characteristics of data. The Unisys analytics platform supports advanced visualization capabilities in numerous contexts, including:
  - Drill-down Reports
  - Dashboards
  - Geographical Reports
  - Social Network Analysis

• Business Intelligence – Even in advanced analytics implementations, traditional BI techniques remain valuable when applied to structured data such as:
  - Data Warehouses
  - Data Marts
  - RDBMS
  - SQL

• Data Intelligence – Part of data’s value comes from what analytics can derive from metadata and data interrelationships. Key techniques include:
  - Link Analysis
  - Entity Resolution
  - Entity Extraction
  - Business Rules

• Data Governance – Unisys recommends a comprehensive data governance approach to assure integrity and accuracy of analytical results and maintain compliance with any legislative or regulatory policies on usage controls and constraints. Successful data governance involves:
  - Data Stewardship
  - Data Auditing
  - Data Access
  - Data Ownership
  - Data Quality
  - Data Security

Analytics solutions that satisfy these concepts and deliver the needed capabilities with the greatest success invariably must take a comprehensive and holistic view of data. This has certainly been the Unisys approach in providing its own advanced data analytics capabilities and platforms.
2.3 EVOLUTION/ROAD MAP

Leveraging advanced data analytics on the road to becoming a data-driven agency involves an evolution of both data complexity and process maturity, as illustrated by the following:

Data Complexity – As volume, velocity and complexity increase, traditional structured data repositories become less effective. Advanced systems need to accommodate big data and NoSQL repositories, as well as legacy systems.

Process Maturity – As demands from data analytics evolve from backward looking (descriptive) analysis to forward-looking (predictive) and ultimately to prescriptive analysis, advanced analytics and data science must be applied to recognize and extract non-obvious relationships and previously hidden patterns.

Taken together, data and process necessitate advanced data analytics methods and tools that can execute large-scale analysis and data mining. Unisys has built its own systems to accommodate growing data complexity and process maturity, with a range of services and service levels appropriate to an individual agency’s needs. Our Advanced Data Analytics experts assess available agency data, then develop and utilize custom Hadoop-based infrastructures to address both technical and logistical challenges. Common steps include consideration of technical, security, policy and other requirements; data ingestion from multiple sources and formats; data cleaning, organization, processing and transformation; and predictive modeling and algorithm development.
3.0 IMPLEMENTATION

3.1 METHODOLOGY AND FRAMEWORK
As illustrated in the Figure below, the Unisys Analytics platform contains a full suite of tools for implementing advanced analytics capabilities. We describe in the paragraphs below several key activities that happen within the platform.

Data Input – Our platform includes raw data inputs for open source and internal sources using Flume and Pentaho for ETL operations. In addition, Scoop provides Enterprise Data Warehouse (EDW) ETL capability for structured data input. Data are loaded into Hadoop for storage within a massively scalable data lake that can incorporate structured, unstructured and semi-structured data – including documents, emails, presentations and images. BDAP also utilizes a RESTful API web service for direct data ETL connections and automation.

Our platform will accommodate partitioning structures and access-controlled repositories for secure compartmentalization of particular data sets. We can also implement fine-grained access controls and statistical metrics, with an automated source catalog utilizing the capabilities provided in Storm, Hbase, Accumulo, Hive and Cascalog.
**Data Discovery** – Data Discovery allows files to be tagged by features – including owner and type. Customized tags can also be created to align with specific objectives. Our platform incorporates an algorithm layer of Data Science and analytics tools that analyze and extract important details and insights from the underlying data. Throughout, the Unisys platform leverages advanced “machine learning” packages for several important tasks:

- Entity Extraction
- Sentiment Analysis
- Key Word Extraction
- Concept Tagging
- Taxonomy Classification
- Data Preparation Routines

Beyond these particular packages, Unisys exposes its complete machine learning toolkit to constantly develop and enhance new algorithms. This broader machine learning toolkit includes Naive Bayes, Network Graphing, and Support Vector Machines. Such algorithms are the foundation for predictive advanced analytics. The system also generates prescriptive reporting – analytics that not just forecast future conditions, but actually recommend future actions – for maximum decision support and improved service delivery.

Our interface layer allows extraction of results either via Web Deployed Applications or programmatically through exposed APIs. It also provides a visualization layer that integrates D3 libraries, as well as standardized analytic reports. Unisys can also develop specific APIs for users and applications in the Secure Cloud, providing access both through the Hadoop file system and SQL interfaces.

**Data Security** – The Unisys platform incorporates enhanced auditing functions and detailed reports. This includes enriched data tags and elements that enable data access controls and governance for specific user roles and data policies.

**Agile Engagement Model** – Agile Engagement Model – As the adjacent Figure illustrates, Unisys employs Agile development principles to understand client data and work collaboratively to develop and implement analytics capabilities that are effective and targeted to agencies’ needs.

*The system also generates prescriptive reporting – analytics that not just forecast future conditions, but actually recommend future actions – for maximum decision support and improved service delivery.*
During the proof of concept phase, we work closely with agency staff to identify available data and get an overall understanding of specific mission objectives. This collaboration leads to a proof of concept model limited to one or two questions or concepts representing specific business needs. From this model we work with agencies to refine and validate both the model and the analytic results.

The next step is the production phase, where we continue the agile engagement process to expand and formalize the analytical framework – incorporating additional data and predictive models while continuing to monitor and refine the overall accuracy and integrity of results. We find that this becomes an ongoing process of agencies discovering new insights through advanced data analytics, which in turn prompts them to ask new questions and make new predictions from their data. Unisys remains a partner throughout this evolutionary process of agencies becoming fully data-driven organizations.

3.2 REFERENCE ARCHITECTURE

Unisys Advanced Data Analytics reference architecture is depicted in the illustration below. It is divided between IT-focused infrastructure, and analyst-focused services and functions. This layered structure enables Unisys data scientists and architects to flexibly tailor analytics platform elements to specific agency needs while preserving cost-effective options for future expansion, capacity increases and adoption of new technologies.

3.3 PLATFORM OFFERINGS

The following exhibit illustrates our Advanced Data Analytics as a Service environment, which delivers a full range of analytics capabilities. Proven configurations and best of breed technologies allow systems to be easily and quickly deployed and hosted on industry leading infrastructure providers, including Amazon Web Services, Microsoft Azure and our own Forward! environment.

Our platform combines the flexibility of a configurable architecture and the agility of cloud-based infrastructure with the power of an integrated set of industry-leading tools and the stability and reliability of proven, time-tested systems.

Since many agencies want the option of consulting services in support of our analytics platform, Unisys Data Scientist as a Service provides access to our experienced staff for help on multiple fronts, including:

- Understanding the hardware and software toolsets that support data analytics
4.0 USE CASES

4.1 GENERAL SERVICES ADMINISTRATION

GSA Product Data Library – For the General Services Administration (GSA), Unisys maintains a product data library that utilizes several analytic models:

- Churn Analysis and Opportunity Growth – Unisys examines transactional sales data for the Product and Service Codes (PSC) that the Federal Acquisition Service (FAS) manages in order to identify and prioritize customers who are more likely to decrease their GSA spend. We also developed a “Customer Flight” dashboard to improve customer retention. Customer survey results (i.e., customer loyalty & detractor feedback) data are matched up to each level as needed.
• Customized Segmentation Validation – To better enable FAS target-marketing and sales efforts, we help the GSA Office of Customer Accounts and Research (CAR) group customers, with important shared characteristics, into clusters. We apply exploratory data mining techniques (clustering) to identify similar buying patterns and then overlay those patterns onto existing categories and segmentation defined by the CAR team.

• Opportunity Chatter – Unisys helps GSA perform text analysis to understand keywords coming through different listings from FBO.gov. Rankings are based on observations of the keywords within the FBO data and are coalesced with the market category data and other PSC and North American Industry Classification System (NAICS) information. This data allows GSA to refer back to actual opportunity listings from FBO.gov.

• Customer Relationship Management (CRM) and Integration of Data – To understand the relationship between buyer purchasing and overall customer satisfaction, Unisys developed concepts and uses for CRM data to augment what’s known from transactional data.

4.2 UNITED STATES DEPARTMENT OF AGRICULTURE
United States Department of Agriculture (USDA) Loan Delinquency Mitigation – To identify and minimize default rates for mortgage loans across portfolios, Unisys developed a two stage model to assess the probability that a loan would be delinquent and estimate the outstanding portfolio balance of delinquent loans. This gave the USDA the ability to proactively manage and mitigate loan loss on the individual and portfolio levels. As illustrated in the figure below, the accounts and outstanding balance that make up a transition state of interest can be identified and then grouped by their probability of delinquency. These data-driven insights help focus the client’s mitigation and remediation activities.

The 30 Day Loans that will go 60 days are of interest

<table>
<thead>
<tr>
<th>Account Number</th>
<th>Probability of Delinquency</th>
<th>$ Outstanding Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXXX209</td>
<td>.92</td>
<td>$80,891</td>
</tr>
<tr>
<td>XXXX165</td>
<td>.5</td>
<td>$96,830</td>
</tr>
<tr>
<td>XXXX483</td>
<td>.92</td>
<td>$4,431</td>
</tr>
</tbody>
</table>

Layer in updated demographic and credit score information
Focused Collection Attention
The transition to digital government is not an all-or-nothing proposition. Agencies will continue to optimize and leverage their legacy environments while they identify opportunities to implement new hosting models; strengthen data collection, integration, and analytics; and expand their digital capabilities throughout the enterprise. The hallmark of a digital organization is its ability to leverage free flowing information through an interconnected world.

The next step for government agencies is to build on their experimentation and initial deployments to create a digital government roadmap. The lessons learned from earlier programs will help agency leaders establish priorities and identify the benefits they want to achieve. This planning will also help them create the governance mechanisms to manage their digital programs and keep them moving forward.

Digital government organizations will enjoy many benefits. Agencies will realize cost savings and operational efficiencies to help them meet expanding mission requirements even as budgets tighten. And the ability to collect and analyze the enormous amounts of data will generate insights for improving the mission capabilities of warfighters, civilian employees and government systems. Overall, digital government will empower employees to bring forward the most advanced and innovative solutions for spending taxpayer dollars wisely, serving citizens, and performing governments’ many missions.

For more information on Digital Government visit: www.unisys.com/digital-government

5.0 FOR MORE INFORMATION
Contact Us at AdvancedAnalytics@unisys.com to connect with our experts. We’ll help you implement affordable big data technologies for data-driven insights that deliver tangible mission impact.

To find out more about how our solutions can work for you, please visit: http://www.unisys.com/offerings/application-services/big-data-analytics

ABOUT UNISYS
Unisys is a global information technology company that specializes in providing industry-focused solutions integrated with leading-edge security protocols to clients in the government, financial services and commercial markets. Unisys offerings include security solutions, advanced data analytics, cloud and infrastructure services, application services and application and server software. For more information, visit: www.unisys.com

ABOUT THE AUTHOR
Dr. Rod Fontecilla is Vice President, Advanced Data Analytics for Unisys Federal. In this role, Rod leads all aspects of software development, system integration, mobile development and data analytics focused on the federal government. He also leads advanced data analytics for Unisys globally. In this capacity, he is responsible for creating data analytic products for multiple industries creating predictive models using machine learning algorithms. He leads a cadre of data science professionals providing business insights to our customers. Rod is responsible for providing leadership, coordination and oversight on all IT solutions, data analytics, emerging technologies and IT services delivery.

Rod has more than 25 years of professional experience in the implementation of large and complex mission-critical IT solutions for both the federal government and commercial clients. Rod also brings an extensive background and expertise in data analytics, cloud computing, mobile development, enterprise architecture, and IT governance and strategy.