Introduction

Along with tremendous opportunity, today's rapid growth of data also presents businesses with a daunting challenge that is pushing the limits of existing IT infrastructure. While big data is a proven driver of value – a fact borne out by countless academic studies, use cases, and corporate earnings reports – companies hoping to reap that value 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 analytics processes.

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

This white paper examines this big data challenge and shares Unisys’ perspective on optimizing Advanced Data Analytics to solve many of the business problems this challenge presents. In the pages that follow, we blend industry context and 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.

We’ll discuss Unisys’ Advanced Data Analytics solutions in considerable detail, since our own approach is derived from more than four decades of experience across vertical markets – including the commercial sector, financial services, and government. 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 industries to embrace analytics for a data-driven and prosperous future.

Unisys’ experts leverage proven, real-world strategies to augment clients’ existing IT capabilities with analytics that span over multiple platforms and devices. Our own successful answer to the big data challenge is an end-to-end data analytics solution that helps our customers to draw vital insights and timely, well-founded decisions “at the speed of business” in order to accelerate growth. We tailor solutions to our clients’ specific needs, combining a proven analytics platform with skilled data scientists who partner closely with client teams to solve their most challenging analytics problems.

Challenges
Creating a data-driven culture, however, is easier said than done. Organizations face many challenges as they work to leverage the data that is available to them. These challenges include:

- **Too much data to process** – Advances in Information Technology (IT) and consolidation of systems and architectures create vast amounts of data. Organizations 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 analytics potential within the volumes and varieties of available data.

- **Inability to access and process data securely** – Enterprise infrastructure often lacks the 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 social media, asset management, events, etc.

- **Budget limitations** – Organizations work under constant budget pressures. Analytics can bring value to a company, but not before a period of investment and capacity-building that must be carried out in the most strategic and affordable way possible.

Unisys singles out these challenges as the most commonly-cited by the clients who seek us out for big data analytics support. To overcome these challenges, we’ve prioritized certain concepts and capabilities – that is, proven techniques for organizations to exploit the potential of data analytics.
Concepts and Capabilities

To satisfy client demands and achieve maximum benefit from advanced analytics, Unisys has embraced a holistic view of Enterprise Data Management, one that encompasses the entire value-generating life-cycle 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 critical capability that makes data and insights more accessible and actionable across the organization. Multiple users and analysts share visualizations to quickly and accurately identify valuable data patterns, relationships, and characteristics. 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 ensure the integrity and accuracy of analytical results and maintain compliance with any internal or regulatory 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 approach Unisys has taken with its own advanced data analytics capabilities and platforms.
**Evolution/Road Map**

Leveraging advanced data analytics on the road to becoming a data-driven organization 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 (i.e., descriptive) analysis to forward-looking (i.e., 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 client’s needs.

Our Advanced Data Analytics experts assess available client data, then develop and utilize customized Hadoop-based infrastructures to address both technical and logistical challenges. Common steps include the consideration of technical, legal, security, and other requirements; data ingestion from multiple sources and formats; data cleaning, organization, processing, and transformation; as well as predictive modeling and algorithm development.
Implementation Methodology and Framework

As illustrated in the figure below, Unisys’ Analytics platform contains a full suite of tools for implementing advanced analytics capabilities. We describe in the paragraphs below several key activities that take place within the platform.

Data Input – Our platform includes raw data inputs for open source as well as internal sources, using Flume and Pentaho for ETL operations. In addition, Scoop provides the Enterprise Data Warehouse (EDW) ETL capability for structured data input. Data is loaded into Hadoop for storage within a massively scalable, data lake that can incorporate structured, unstructured, and semi-structured data – including documents, emails, blogs, presentations, and images. BDAP also utilizes a RESTful API web service for direct-data ETL connections and automation.
Our platform accommodates partitioning structures and access-controlled repositories for secure compartmentalization of particular data sets. We 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, Unisys’ platform leverages advanced “machine learning” packages for several important tasks, among which are:

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

Beyond these particular packages, Unisys shares its complete machine learning toolkit to continuously 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 only forecast future conditions, but actually recommend future actions – for maximum decision-making support.

Our interface layer allows extraction of results either via Web Deployed Applications or programmatically, via exposed APIs. It also provides a visualization layer that integrates D3 libraries as well as standardized analytics reports. Unisys also develops 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. It includes enriched data tags and elements to enable data access controls and governance for specific user roles and data policies.

**Agile Engagement Model** – As the figure below 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 customers’ needs.

---

**The system also generates prescriptive reporting – analytics that not just forecast future conditions, but actually recommend future actions – for maximum decision support.**

---

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

The next step is the production phase, where the agile engagement process is extended 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 the results. Clients find this to be an ongoing process where they discover new insights through advanced data analytics, which in turn prompt them to ask new questions and make new predictions from their data. Unisys remains a partner throughout this evolutionary process where clients become fully data-driven organizations.

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 client needs while preserving cost-effective options for future expansion, capacity increases, and adoption of new technologies.

Platform Offerings

The exhibit below also illustrates our Advanced Data Analytics as a Service environment, which makes available a full range of 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, as well as 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 clients 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 assistance on multiple fronts, including:

- Understanding the hardware and software toolsets that support data analytics
- Making the business justification for follow-on work, including use cases, to demonstrate value and ROI
- Understanding IP issues that help differentiate Unisys/client collaborations within the Big Data and Data Analytics marketplace

Unisys delivers expert advice on data strategy, clarifying priorities needed to put data analytics into action and establishing success measures and goals for analytics outcomes in the organization. We provide a roadmap to lead clients from raw data to business insight.

A typical 3 to 6-month engagement includes:

- Assessment of current information capabilities
- Identification of desired information capabilities
- A high-level roadmap
- Clarification of target outcomes
- Executive briefing of results

### Commercial Use Cases

**Predictive Sentiment Analysis** – The measurement of user and customer sentiment is as challenging as it is vital to business success. Traditional measures, such as Likert scale surveys, provide easy to understand objective data, but they frequently miss subtleties in attitudes and topics outside of the survey ratings.

Analytics can bring greater clarity in the form of predictive sentiment analysis on unstructured data. Written comments from users and customers, for instance, can be analyzed by automated processes to determine overall attitude without the need to manually read and tabulate comments. We have applied this technique to service desk response surveys, including the case illustrated in the graph that follows.
**TOP 10 PRODUCTS BASED ON NUMBER OF COMMENTS AND THEIR AVERAGE SENTIMENT**

<table>
<thead>
<tr>
<th>Product</th>
<th>Number of Comments</th>
<th>Sum of Sentiment Score</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise Portal Platform EEP (RXPSecurity)</td>
<td>633</td>
<td>4039</td>
<td>2.65</td>
</tr>
<tr>
<td>Microsoft Office Outlook 2010 (SP1)</td>
<td>1335</td>
<td>2625</td>
<td>3.19</td>
</tr>
<tr>
<td>Microsoft Office Outlook 2013 (RTM)</td>
<td>824</td>
<td>3289</td>
<td>2.9</td>
</tr>
<tr>
<td>Microsoft Windows 8 Enterprise</td>
<td>1220</td>
<td>1527</td>
<td>2.11</td>
</tr>
<tr>
<td>Password Reset</td>
<td>739</td>
<td>1555</td>
<td>2.1</td>
</tr>
<tr>
<td>Physical SmartCard</td>
<td>644</td>
<td>1795</td>
<td>2.79</td>
</tr>
<tr>
<td>SmartCard</td>
<td>856</td>
<td>2360</td>
<td>2.76</td>
</tr>
<tr>
<td>TSPT - Remote Conn - VPN App</td>
<td>987</td>
<td>2781</td>
<td>2.82</td>
</tr>
<tr>
<td>User Account</td>
<td>1224</td>
<td>2938</td>
<td>2.4</td>
</tr>
<tr>
<td>Virtual SmartCards</td>
<td>742</td>
<td>2034</td>
<td>2.74</td>
</tr>
</tbody>
</table>

Number of comments | Sum of Sentiment Score

**Pharma customers were able to increase their volume forecasting from 20-percent to 90-percent, which translated into a 30-percent increase in efficiency**

From these data, we see how products like Microsoft Outlook may have the highest sentiment scores, even though they are not a significant source of tickets. This kind of analytics insight helps us concentrate our efforts on reducing tickets and expediting resolution of the ones that still flow in. Other surveys have shown dramatic differences between the sentiments expressed in written comments and objective scale ratings. Analytics can clarify the factors at work in such situations.

**ITSM support for Pharma** – Unisys provides IT Service management for Amgen, one of the world’s largest biotechnology companies. Unisys coordinates service desk and desk-side support with Amgen’s business operations to serve a network of 34,000 desktop/laptop/tablet systems and other services – including telework and service desk support – at approximately 45 locations in 40 countries worldwide.

Unisys delivers centralized desktop/laptop/tablet IT services, including application delivery, modernization, and compliance with Amgen security policies. In the process, we’ve helped Amgen achieve a cost savings of $8.8 million.

To support continuous improvement in our service delivery, we analyzed multiple documents as well as categorized and linked various profiles to optimize end-user support roles. The goal was to automate the process of accurately identifying topics from text documents and building a knowledge graph that would allow support team members to discern issues without spending excessive diagnostic time. We fed the required unstructured and structured information into our analytics platform to serve as our data lake. This allowed us to apply advanced algorithms to numerous tasks, including:

- Text Extraction
- Concept Tagging
- Sentiment Analysis
- Knowledge Base and Inference Engine
Unisys worked directly with Amgen to identify a range of hypotheses, requirements, business values, and results. A flexible and iterative approach allowed us to update and adjust methods as needed. In the process, we encountered and overcame a number of complex challenges, among which were:

- Analyzing text and documents was proving to be difficult, and a number of ingestion algorithms had to be applied to extract data as required.
- Accuracy of the learning algorithm improved, but in many cases additional data was needed to explain variances. Unisys was able to strategize where to devote additional time and data resources for this purpose.
- Developing appropriate roles and personae is a business matter that many clients don’t fully comprehend. Unisys leveraged its deep domain expertise to help clarify needs and achieve positive outcomes.

Ultimately, Unisys was able to boost overall operational efficiency of the Amgen support team and improve communication and sentiment with end-users.

Telco Data Products Library — Unisys delivered analytics services on a data products library for a major telecommunications firm. For this engagement, we developed several models of client operations, including:

- Customer Segmentation and 360° Views of Customers — In this model, we used multiple internal databases to create a 360-degree view of each customer. This made it possible for our client to have the ability to categorize and group customers based on past behavior and historical trends.
- CPU Utilization and Batch Scheduling — This model permitted analysis of log files from transactional databases in order to identify historical usage and predict spikes in demand. It made it possible for our client to examine usage — by specific customers on a minute-by-minute basis — to define and classify behavior and anticipate future activity. The model also allowed for exhaustive stress-testing of the online systems to predict the likelihood and impacts of high utilization.
- Routing Optimization — With this model, we examined specific routing patterns to predict underutilization of certain routing options. Underutilization leads to decreased revenue, so this model helped the client to identify trends and patterns on routing information in order to maximize efficiency and offer new features to customers in the future.

Government Use Cases

GSA Product Data Library — For the General Services Administration, Unisys maintains a product data library that utilizes several analytics models:

- Churn Analysis and Opportunity Growth — Unisys examines transactional sales data for the PSC codes that the Federal Acquisition Service (FAS) manages in order to identify and prioritize customers who are more likely to decrease their GSA spending. We also developed a “Customer Flight” Dashboard to improve customer retention. Customer Survey Results (i.e., customer loyalty and detractor feedback) data are matched up to each level as needed.
- Customized Segmentation Validation — To better equip FAS with the ability to target marketing and sales efforts, we help GSA/ CAR to 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 flowing in from multiple FBO listings via FBO.gov. Rankings are based on keyword observations within the FBO data and coalesced with the market category data and other PSC and NAICS information. This data helps reference back to actual opportunity listings from FBO.gov.
- 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 transactional data knowledge.

A Telco customer was able to offer new online products specifically targeted to groups of customers with 90 percent effectiveness.
USDA Loan Delinquency Mitigation – To identify and minimize default rates for mortgage loans across portfolios, Unisys developed a two-stage model to assess the delinquency probability of loans and estimate the potential outstanding portfolio balance. This equipped the USDA with the ability to proactively manage and mitigate loan loss at 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 delinquency probability. These data-driven insights help focus the client’s mitigation and remediation activities.

<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>

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 business 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

Dr. Rod Fontecilla

Dr. Rod Fontecilla is Vice President of 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 analytics products for multiple industries and for creating predictive models using machine learning algorithms. Rod leads a cadre of data science professionals providing business insights to our customers. He is responsible for providing leadership, coordination, and oversight on all IT solutions, data analytics, emerging technologies, and IT service delivery.

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