Analytic Technology™ Industry Roundtable

Analytic Technology Industry Roundtable Study: Industry Collaboration and Review

Mr. Bryant Choung, Palantir

Mr. Matthew Chandler, Palantir

November 2016

This page intentionally left blank.

Architecture Working Group Study Series

Industry Architecture Survey and Review

Abstract

The Analytic Technology Industry Roundtable recently assessed product vendor perspectives on trends in Government acquisition of analytic tools and data management architectures. Based on a survey and in-depth study, the Roundtable found that in the current environment of shrinking budgets, Government clients are increasingly opting for commercial technologies over custom software development projects. The Roundtable also found that commercially available analytic tools and data management platforms are increasing in number and technological sophistication. Both Government and product vendors see open architectures, APIs, and standards as important in the development and implementation of enterprise technology.

Table of Contents

Background and Executive Summary

MITRE hosts the Analytic Technology Industry Roundtable, an association comprising leading vendors of analytic tools and data management platforms. While the products offered by the vendors are all different, all provide analytic solutions to customers in the commercial and Government sectors. MITRE regularly hosts Roundtable events and working groups for industry participants and oversees the publication of studies and publications. The goal of the Roundtable is to provide informed guidance on data and technology issues to interested parties across the Government.

As part of the Roundtable, the Architecture Working Group fielded a survey to industry to capture the experience and best practices associated with delivering analytic tools and data management architectures to Government customers. The questions included in the survey covered topics on high-level data architectures, trends in customer preferences, and similarities and differences among clients across intelligence, defense, civilian and commercial organizations.

A review of the answers collected from the survey yielded several high-level themes that we describe in this paper from two distinct perspectives:

What factors drive Government technology acquisition?

- With decreasing budgets, cost is emerging as the strongest influencer
- Customers are adopting commercial solutions that can be customized instead of engaging in custom software development
- Openness and extensibility of solutions is important

These themes emerged from the observations and perspective of industry participants who have worked with government clients in intelligence, defense, and civilian agencies

What factors drive vendor offerings?

- Vendors are finding overlap across customer needs and have developed common product offerings
- Vendors find success using time-tested platforms over trendy technology
- Openness and extensibility is a strong driver in the development of Commercial solutions

These themes are informed by vendor decisions in the development and deployment of commercial solutions.

What factors drive Government customers in choosing a technology solution?

This section explains three influential factors in the Government acquisition of analytic tools from the perspective of vendors. While Roundtable members are not government officials making technology acquisition decisions, their industry insights are based on experience working across the whole of government.

With decreasing budgets, cost is emerging as the strongest influencer.

In recent years, a common challenge across US government agencies has been the continued pressure to reduce costs and spending. Consistent with this macro trend, industry has observed Government clients re-evaluating technology investments and moving towards streamlining architectures and solutions. There is decreasing interest in large-scale custom software development projects; instead, Government is increasingly adopting technologies available through hosted cloud models and prefers solutions with a smaller number of discrete software architecture components.

Industry participants noted that agencies have been forced to refactor and re-evaluate legacy data architectures in light of constrained budgets. Although previously customers had the resources to invest in complex custom software development, today there seems to be a reduced appetite for the high development cost and risk associated with such projects. Instead, customers are now opting for readily available commercial solutions and Commercial Off the Shelf (COTS) technologies.

In contrast to Government off-the-shelf (GOTS) and Custom IT solutions (CITS), COTS based commercial software solutions, are ready-made technologies immediately available for sale. According to the International Council on Systems Engineering, COTS products carry:

- **Lower costs**: COTS products cost less than custom software development because the development costs of COTS products are distributed between many users with private industry paying for continuous improvement and innovation.
- **Higher quality**: COTS products have been developed and improved over time, with many stakeholders participating in finding bugs and limitations in the product.

- **Fewer customer start-up costs**: Productized solutions are designed for quick implementation, with full documentation accompanying the start-up process.
- **Less development time**: Customer organizations can spend time learning about and customizing the product, rather than developing it from scratch,[1]

Developing new GOTS solutions carries significant risk because most software development projects fail.[2] This failure only becomes more likely as the project becomes more complex. Organizations that choose to develop software are accepting the associated cost and time-related risks.

It is true that custom software development projects promise to meet 100% of the requirements of a given Government acquisition, while commercial solutions may only meet 80-90% of the same requirements. Yet customers are still increasingly choosing commercial solutions that deploy more capabilities sooner instead of waiting for potential solutions (that are likely to fail).

With GOTS solutions, the Government has the apparent advantage of owning the software source code. However, this ownership also comes with the costs of maintaining and updating the software baselines. Although there is a common perception that owning the source code eliminates the risk of "vendor lock", in fact government agencies often remain beholden to the original contractor that developed the source code for them. Instead, the use of open API's and open data interchange formats is more important in eliminating "vendor lock", regardless of software code ownership.

Innovation in the data and analytic technology sector is fast-paced, and lengthy custom software development cycles are often unable to keep up with new commercial software offerings. Because it can take years for software to be developed after it was designed, custom software projects are often functionally obsolete by the time of release.

[1] Numerous International Council on Systems Engineering (INCOSE) publications detail the advantages of COTS (e.g., http://www.incose.org/northstar/2001Slides/McKinney%20Charts.pdf).
[2] The Standish Group, Chaos Manifesto 2013: Think Big, Act Small (Version 1). The Standish Group International, 2013. pp. 4.

Customers are adopting commercial solutions that can be customized instead of engaging in custom software development.

With the rapid growth of data creation, consumption, and analytics across government and commercial sectors, analytics tools and data management architectures have transitioned from highly bespoke solutions developed for a few select agencies and organizations into mature commercial platforms offered by experienced vendors. Vendors have developed commercial analytic software tools and platforms to meet the increasing volume, velocity, and variety of data requirements for customers across many different mission use cases.

As described above, commercial and COTS solutions often meet the vast majority of data requirements of a modern organization. Customers can further customize or configure commercial software to meet the entire range of data requirements. Most commercial software solutions are fundamentally built to be customizable and extensible to support a variety of use cases and configurations.

Examples of these customizations may include:

- Configuration of the data model to support customer needs
- Integration with internal and external data sources owned by an organization
- Individualized analytics and visualizations on custom data sets
- Interoperability with other existing systems and integration into workflows
- Adoption of organizational security, authentication, and authorization protocols

When deciding between COTS software and GOTS software solutions, the government often incorrectly assumes that COTS solutions cannot be both customized and still acquired as a commercial item. However, under the applicable provisions of the Federal Acquisition Regulation (FAR), commercial items can be bought and customized under a commercial item contract as long as the modifications to the software are of a type that is customarily provided in the commercial sector. In fact, the acquisition regulations and guidelines have a preference for commercial items over developmental items and even encourage agencies to consider modifying their requirements to a reasonable degree to acquire commercial solutions.

This preference highlighted by acquisition regulation is supported by the reality that it often makes sense to deploy a working solution quickly rather than embark on the time intensive exercise of developing new technology from scratch.

Technology is evolving at a pace that is measured month-to-month. Adopting commercially available solutions and customizing them for the particular needs of an agency is much cheaper and expeditious than waiting for custom software development that can take years.

Openness and extensibility of solutions is important.

Government customers value openness and extensibility in technology solutions. Customers seek to maximize the value of their overall technology investment by ensuring that any new software platform can be easily integrated, customized, and improved within the existing customer environment.

It is often incorrectly assumed that COTS solutions are less open and less extensible than GOTS or open source solutions. In fact, the only technical difference between COTS and other solutions is in the licensing structure, which has no bearing on the technical extensibility or openness of a software package.

When properly evaluating software packages for openness and extensibility, customers evaluate software based on the following criteria:

- **Open, non-proprietary data and file formats**: Once data is integrated into the system, it should be easy to export or otherwise be accessible to other software systems
- **Cross-platform programming language**: The system should be developed in a common computer language for performance on any platform
- **Public and open application programming interfaces (APIs)**: The system should possess open APIs at every service level (e.g., data integration, authentication, search) so that the product can be extended through the addition of applications, plugins, and servers on top of the existing capability
- **Plug-in architecture**: The system should support the addition of plugins, from small helpers applicable to single use cases to full-fledged applications
- **Data agnostic**: The system should be unencumbered by source data formats, with the ability to 1) ingest data from any accessible source and 2) export all data in a common data format
- **Highly configurable**: The system should be highly configurable and customizable to enhance user interactions and optimize workflows
- **Wide Adoption**: The system should be fielded in multiple contexts across government and the commercial industry

- **Interoperable**: The system should have demonstrated integration with external systems from a variety of vendors across government and the commercial industry
- **Platform agnostic**: The system should be deployable on bare metal hardware, private cloud, and commercial cloud environments

What factors drive vendor offerings?

Vendors part of the Roundtable regularly meet with new and existing customers across the government and commercial sectors. Based upon observations and feedback from customers, vendors either improve existing solutions or provide new products to customers. The survey identified several factors that influenced how vendors shaped their product offerings.

Vendors are finding overlap across customer needs and have developed common product offerings.

Although Government customers often perceive their data and analytic use cases and requirements to be unique, vendors regularly find common workflows and product solutions that fit multiple customers and industries.

Challenges related to analytic tools and data management architectures are not confined to select intelligence agencies or defense customers within the government. In fact, effectively understanding and analyzing data is a common need of agencies across the whole of government.

Sometimes the similar needs of different agencies are obvious. For example, sometimes the same data sources are shared across diverse customers in military, law enforcement, and financial organizations. Instead of each customer developing their own unique data integration, vendors can do the integration once and offer it to multiple customers across industries.

Similarly, multiple customers often confront the same broad technology challenges even if their core mission areas may differ A good example of this is cyber defense. Regardless of the core mission area, customers across all sectors are cyber defense analytics and solutions.

At other times, a less obvious overlap between the needs of different organizations may exist. For example, two distinct industries may have very different workflows and draw on non-overlapping datasets and yet an exploration of the data

in each case may yield similar patterns. For example, the same techniques to identify outliers in credit card transaction data may be also be applicable to identifying potential malicious cyber data or relevant SIGINT targets.

In all of these cases, commercial vendors can develop products that address similar needs and challenges across a broad swathe of organizations and industries. Productizing these solutions yields benefits due to shared costs of research, development, sustainment, and evolution of software.

Commercially developed and productized solutions results in software that is cheaper for government agencies to acquire and maintain while being of higher quality and greater utility than the alternatives. Rather than developing and improving software with input from a single agency or workflow, these broadly deployed solutions capture enhancements and improvements that are collected and shared across a community of users that spans multiple agencies and industries. This results in better solutions than if custom solutions were developed for individual customers based on narrower use cases.

Vendors find success using time-tested platforms vs. trendy technology.

Recent years have witnessed explosive data growth due to developments in Internet and sensor technologies. A wide range of data management, analytic, and visualization technologies have been developed to manage unprecedented data scales. There are many new emerging technology trends perceived to represent the “bleeding edge” of available analytic and data management. However, these new technologies often push the limits of current processing and storage technology at the expense of reliability and robustness.

The leading commercial vendors of data analytic and management solutions continuously evaluate how incorporating these emerging technologies will impact the *reliability* and *integrity* of their systems. Although these new technologies often provide isolated gains in data processing or analysis, they are often unsuited for deployment into mission critical customer environments.

A common theme shared by the Roundtable vendors is their focus on developing and delivering robust product-based solutions. While all the vendors are competing for best-in-class performance and capability, the integrity of customer data is non-negotiable. Additionally, product companies optimize on the ability of their technology to be easily deployed and configured for repeatable success at multiple client sites with minimal effort and cost. The inclusion of unproven or untested technologies puts these goals for productized solutions at risk.

An example of a trend that has matured and gained strong adoption across vendors and customers is cloud technology. Initially, many emerging technologies were part of the cloud trend, and many customers were uncomfortable with entrusting mission critical workflows and data to these unproven technologies. As the components supporting cloud technologies have matured, we have seen increased adoption and usage of cloud infrastructure, especially when there is scope for flexible and scalable technology infrastructure. Today, customers are either willing to adopt solutions that incorporate cloud technologies or have already adopted and invested in cloud infrastructure. As a result, solutions from vendors now increasingly incorporate these cloud technologies as part of their solutions.

Openness and extensibility is a strong driver in the development of Commercial solutions.

This survey revealed that commercial software vendors also view openness and extensibility as critical to the development of commercial solutions.

This may come as a surprise to those who equate commercial software solutions with “vendor lock” and “proprietary” data formats. However, commercial industry solutions stand to benefit from the adoption, inclusion, and promotion of open platforms, extensible interfaces, and non-proprietary data formats.

As commercial analytic tools and data management platforms need to easily integrate with existing customer environments, commercial vendors have built their solutions to be make data ingress and egress as easy and efficient as possible. Most commercial software products support data import and export using open APIs and formats including REST, JSON, and XML.

Beyond the free movement of data, vendors are also interested in creating robust ecosystems for their products. This means that the core analytic and data platforms can be extended with functionality developed by in-house and third party developers. This is enabled by the availability of open and well-documented APIs to the developer community.

In sum, open and extensible software integrates more easily with legacy customer environments ensures that vendor products are applicable to a broad range of organizations and industries, and enables adaptation to new and emerging use cases over time. The Roundtable vendors implement many of the principles and practices listed.

Conclusions

The Roundtable's Architecture Working Group assessed study inputs and survey responses from vendors that provide analytic tools and solutions.

Vendors highlighted the many similarities in data and analytic challenges faced by customers across government and commercial sectors. As a result, vendors have prioritized the development of productized solutions that address these common challenges, and that can be easily tailored to a wide range of use cases and environments.

Vendors observed that the current budgetary environment has imposed constraints on Government customers. The study showed this results in a Government preference for simpler technology solutions and a bias against new custom software development projects from government customers.

Finally, vendors highlighted that while both Government and industry have a strong preference for open platforms, standards, and extensible solutions, certain misconceptions about how commercial solutions relate to these factors persist.

Study Oversight and Guidance:

MITRE Corporation

Study Participants:

SAP, Recorded Future, ESRI, MarkLogic, MapLarge, Cloudera, MongoDB, Thomson Reuters, Lexis Nexis, NetOwl, IBM, Centrifuge, Tableau, SAS, FICO

Study Authors:

Bryant Choung (Palantir), Matt Chandler (Palantir)