INSIGHT PAPER

THE CORE COMPONENTS OF AN ENTERPRISE SELF-SERVICE DATA STRATEGY

For companies to compete in a rapidly changing business environment, they need to empower their workers across the enterprise with self-service access to data to explore and uncover insights for informed decision making. An enterprise strategy for self-service data access, exploration and curation encourages every person to solve business questions in a simple, intuitive and engaging way. Business users empowered with this autonomy can test hypotheses, explore "what if" scenarios, evaluate new data, better understand existing business processes and make their company more agile and responsive to opportunities or risks – ultimately positioning the company to beat the competition.

While imperative for any business, the reality is that enabling every business user in an enterprise can be challenging when there are multiple tools to integrate and different products to manage – each with a different interface and nomenclature for users to learn. When self-service tools and capabilities work together transparently they become a unified environment for the user. This unified platform approach centralizes data and processes within the environment and improves manageability and stability from an IT perspective. It also lends to a positive user experience, which greatly impacts pervasive use and adoption and, ultimately, the value delivered from a self-service strategy.

THE SIX PILLARS OF ENTERPRISE SELF-SERVICE DATA

To implement an enterprise self-service data strategy, the following six pillars are the foundation required to fully empower users with data. When used in concert with each other, these components balance freedom with governance, individual work with community collaboration and insight, and individual domain knowledge with the assistance of machine learning and artificial intelligence – all while improving speed and agility in solving business needs.

The following six pillars have been distilled from many critical success factors found in self-service programs that have enterprise-wide expectations for business value through agility and speed, balanced with consistency and accuracy:

1. **DATA GOVERNANCE AND SECURITY** are essential to ensure that people with access to data are authorized and understand its context and proper usage. Beyond access, an important component of data governance is the process that facilitates the review, consensus and certification of new data discoveries to be shared, analyzed and built-upon by others in the enterprise. In this area are capabilities for role-based security, semantic layers, certification, publishing and usage monitoring or audits. Maintaining data lineage is also important for supporting governance, security and masking of all integrated and derived data sets to ensure complete authorized access. From a strategic perspective, data governance should be ingrained from the beginning of the self-service program, as it is risky and difficult to retrofit once data flows have been created.

2. **DATA CATALOG AND DISCOVERY CAPABILITIES** allow business analysts to quickly find and verify quality data that will be useful for their projects. Discovery enables people to find available sources of data, upload new internal or external data sources they may need, inspect and understand what the data represents, and decide whether it is applicable and feasible for the task at hand. The ability to tag and annotate data sets for further discovery aids collaboration and helps co-workers understand the data faster. Discovery capabilities include visually interacting with data to see information such as data profiles, metadata and data lineage. Data discovery is a highly iterative process that needs to execute at the speed of the end user's thought process and, therefore, requires a highly engaging, intuitive user experience.

3. **DATA PREPARATION** is at the heart of self-service data and empowers the businessperson to build data manipulations and integrations while reducing dependence on IT data and programming professionals. With data prep capabilities, business analysts don't require technical database skills to quickly create or validate data transformations when producing the result data sets needed in reporting, exporting or analytics. An important aspect of data prep

is dealing with the ever-increasing volume of unstructured data, such as weblogs, clickstream or Internet of things data sources. Adding structure to this unstructured data through normalization is a key element of data prep. The ability to join multiple disparate data sources together using a common key is a requirement that must be simple enough so that business users with no programming skills can achieve such functions and find the insights they need without IT intervention.

4. **COLLABORATION AND COMMUNITY** become critical within self-service platforms for users to accelerate communication and validation of their work. As in any environment, users must be able to share their insights with co-workers, thereby increasing the cumulative data knowledge for the organization. This community support increases user confidence in how to create appropriate workflows and work with the data. Crowdsourcing clarification from other data users or data stewards can assist users when data catalogs and search don't provide all the information or clarification needed. People can publish and share their results for sources of inspiration and reuse derived data sets to be applied in similar business scenarios elsewhere in the organization.

5. **CLOUD OPTIMIZATION** focuses primarily on stability by eliminating concerns and disruptions from an underlying platform while providing scalability, elasticity and a high-performance user experience. Cloud-based platforms deliver the benefit of lower total cost of ownership, making self-service platforms cost-effective for the demands of many enterprise users and with global reach. Additionally, cloud platforms are better equipped to handle the unpredictable nature of self-service, ad hoc data work by leveraging elastic services that can scale up and down with a large amount of user activity without disruption in user experience. To reduce complexity, the platform should enable an in-place architecture whereby data is sampled into the cloud processing environment and change data capture functions allow automated workflow to easily update derived data sets as source data changes. A cost-based optimizer should be employed to allow the platform to select the most appropriate processing engine for the task at hand in order to maintain a cost-effective processing environment – rather than committing every task to expensive memory-based processing such as Spark. This can save substantial processing costs as concurrency increases.

6. **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** have recently become a necessary component of enterprise self-service data environments for assisting people in understanding vast amounts of data and revealing possibilities in the data that are challenging for the analyst to see. Built-in AI and ML routines are a part of the data discovery, data prep, governance and collaboration pillars for everyone to leverage, specifically targeting data intensive areas such as data quality and formatting, semantic layer and contextual inferences, and identifying potential PII data risks. Many of these benefits are simply conveyed as recommendations with a one-click fix for the user working with the data.

SELF-SERVICE DATA IN A UNIFIED PLATFORM

These six pillars are foundational to a comprehensive self-service endeavor. As a self-service strategy matures, it becomes even more important for all six pillars to work together in concert for companies to become more data-centric and agile.

Companies that do not approach their self-service data strategy from a unified platform perspective often struggle to integrate and maintain an ecosystem of products that facilitates all six pillars. When companies begin with only one pillar, such as data prep, business users still rely on IT to discover and source data sets, which delays time to insight. Subsequent addition of other point solutions to add the functionality of other pillars increases complexity on multiple fronts, for both business users and IT. Business users typically face a steep learning curve, as they must be skilled with each vendor's interface to tackle each of the stages of self-service data and to pass data from one interface to another. IT must now maintain multiple systems and their underlying service level agreements (SLAs), connect to multiple

data sources redundantly and establish a means for the multiple services to interact with each other.

Further, as the number of vendors in the ecosystem increases, the associated overhead for vendor management, tool administration and communication is compounded, impacting total cost of ownership.

WHERE SELF-SERVICE DATA IS MAKING AN IMPACT

Leading business use cases are customer-centric programs for better understanding and targeting while compliance-centric programs such as governance and Global Data Protection Regulation (GDPR), serve as examples where enterprise self-service is having an impact and changing the way large initiatives are undertaken.

CUSTOMER ANALYTICS

Customer 360 initiatives provide a more complete understanding of customers with data from many new and existing data channels and systems. This includes mobile and web interactions that can be leveraged by cross-functional business departments such as customer support, finance, sales and marketing. With a self-service data platform encompassing all six pillars, business users can deploy AI/ML routines to help understand the data, assess and resolve data quality and infer customer data relationships across data sources. Business users can also easily ingest social media data sets and conduct net positive and negative audience sentiment analysis. Most importantly, business users across different business functions can integrate data for their own needs with help from data catalogs, collaboration and community resources to focus on predictive analytics and discovering the next best actions to take for reducing churn and increasing customer lifetime value. Additional customer insight initiatives include audience analytics, micro-segmentation and micro-targeting for sales.

GLOBAL DATA PROTECTION REGULATION (GDPR) AND DATA COMPLIANCE

Most companies have customer data that is protected by industry or government regulations and are subject to large penalties and fines for noncompliance. Companies that have customers in the European Union or process data on behalf of European companies will soon be required to comply with the upcoming Global Data Protection Regulation (GDPR), which goes into effect May 2018. While the benefits of democratizing data across the enterprise are compelling, organizations must also ensure that business users working with customer data only have access to customer attributes as appropriate, and their derived data sets recognize and maintain this lineage and protection. Beyond ensuring authorized access to protected customer data, GDPR compliance will require the audit of access notifications and use of data. A reactionary approach to these guidelines and audits cannot keep up with the speed and volume of enterprise-scale user-data interactions. The key is to establish data governance, masking and compliance from the onset of a self-service data strategy and build the tools and processes directly into the self-service data environment so users can confidently work with data without risk. A proactive approach will leverage AI and ML routines to reveal new customer data that should be marked sensitive, provide user notification and metadata when preventing access to designated customer data and obscure data correctly at row and column levels, thus preventing unauthorized information access while allowing analytics.

A self-service platform must also meet the requirements of additional data privacy regulations, such as Personally Identifiable Data (PII) Privacy Act, Protected Health Information (PHI)/Health Insurance Portability and Accountability Act of 1996 (HIPAA) and Payment Card Industry Data Security Standard (PCIDSS).

CONCLUSION

The goal with enterprise self-service data is to develop a culture of self-sufficiency to maximize the impact of people working with data. Understanding how the six pillars of enterprise self-service data benefit ongoing adoption and maturity from the onset of the initiative will facilitate progress without taking backward steps or operating with limitations. A unified self-service data strategy enables both business users and IT to immerse themselves in working with data that results in a fast, agile and frictionless working environment.