WINTERCORP REPORT

BY RICHARD WINTER

SPONSORED RESEARCH PROGRAM

WINTER CORPORATION

teradata. AppCenter

Easing Development and Deployment of Analytics in the Modern Data Ecosystem

EXPERTS IN ANALYTIC DATA MANAGEMENT AT SCALE

teradata. AppCenter

Easing Development and Deployment of Analytics in the Modern Data Ecosystem

RICHARD WINTER

WinterCorp
www.wintercorp.com
42 DERBY LANE, TYNGSBORO, MA
617-695-1800

Summary

EXECUTIVES AROUND THE WORLD are seeking a faster and more effective way to put analytics into action. An improved model, for example, only pays off when that churn predictor is used in the business as part of a customer retention system, actually increasing lifetime customer value and profit.

But in practice the data scientist or analyst who creates the new model is often far removed from the person in customer service or marketing who can take the action. Because the data scientist and business user may be in different locations or organizations, it may take months or even years for the new analytic to have a business impact.

Then, to maximize its value, the company will want to use that churn predictor in our example repeatedly throughout its enterprise operations.

To truly put analytics into action, companies need a way to create, deploy, consume and reuse analytics with speed and agility, often on a national or global scale.

Teradata AppCenter takes aim at this need. Inspired in part by smartphone "app stores," AppCenter provides one place for a business user to leverage pre-built analytic functions or efficiently assemble analytic components, and quickly deploy those applications within the business.

PROBLEM/REQUIREMENT	APPCENTER CAPABILITY
Self-Service for Analytic Apps	End User Graphical Interface: Click An Icon To Run An App
Finding What Apps are Available in the Enterprise, Encouraging Reuse Across Organizational Units & Geography	Central Repository, Tagging, Search
Guided Development	Built-In Solutions for Data Access, Security, Logging to Speed Development and Standardize Methods
Access to Multiple Data Sources	Teradata Vantage™ and Teradata Querygrid™ Provide Transparent Access from Multiple Data Storage and Streaming Platforms
Use of Externally Developed Analytic and Machine Learning Functions	Any Function in a Docker Container can be Deployed as an AppCenter App
End-to-End Analytic Function Development, Deployment and Consumption is Slow and Frustrating	Entire Process is Accelerated, Reducing Redundant Efforts and Increasing Productivity

How Teradata AppCenter Adds Value

Use of Teradata AppCenter has a major positive impact on an enterprise. Most importantly, it enables a modern analytics ecosystem. AppCenter apps can be set up to access data in processing engines on a wide range of data platforms or combine data from multiple data platforms, and provides a consistent framework for deployment, with standards for how key interfaces are implemented. Deployment is a guided experience, in which the developer is assisted with such issues as security, connection to data sources, logging and other common problems, that results in a new level of interchangeability for analytic components and reduces friction throughout the chain of analytic development and use. AppCenter is designed for a variety of users (i.e. data scientists, data engineers,

data analysts, developers and database administrators) and reduces the time to design and deploy analytics because all functions and components are housed in one place. Users shop for components in a self-service graphical interface, and then deliver solutions to his or her target consumer. The user doesn't need to know or care where the function came from; what language or tool it may have been created with; what data it needs to access; or, even where that data is stored. The least technical user can readily find, assess, and apply the apps he or she needs. Beyond that, the Teradata AppCenter enables the cross-organizational sharing of analytic solutions and components that have been created. That empowerment harnesses the creative skill and energy of all those who can contribute across the enterprise and reduces redundant work.

This report introduces Teradata AppCenter and describes examples of its business impact.

On the basis of its independent research, WinterCorp recommends that companies investing in analytics look closely at Teradata AppCenter, with an eye to near term implementation. ●

1 Issues in Deploying Analytics

Virtually every major business process today requires analytics: running a marketing promotion; streamlining a supply chain; or, altering an employee benefits package. The same holds true for any major business decision: locating a warehouse; entering a new market; or, deciding where to cut so as to rein in costs. Business success requires that these tasks be accomplished effectively and without delay.

That need for speed means there is tremendous pressure to implement these analytics faster and more often, but the process of creating a single new analytic solution and applying it for business impact can be both long and frustrating.

To those tracking the technology of data analytics — data science, machine learning, big data, the cloud — this may seem surprising. After all, nearly every week brings news of some major advance in analytic tools and platforms. However, while these new developments provide more options for solving an analytic problem (for example, increasing the power and intelligence of algorithms for analyzing the data), **they don't shorten the end-to-end path from business problem to business impact, which remains painfully slow**.

Why can't we build and deploy analytics faster?

First, the analytic data environment is now a lot more complicated than it was even five years ago. There are more data platforms: the typical large company has at least one data warehouse platform, at least one on-premises data lake and at least one cloud data platform. Frequently, there are several of each, as well as data marts. In addition, there may be separate, specialized platforms for analytics, as well as external platforms from which machine learning algorithms and public data may be sourced. This means there are a lot of different places where data and analytic functions may be stored.

Second, implementing analytics is a multi-step process requiring different skills — and, usually, different personnel — at each step. Imagine the typical process of analytic problem conception to the delivery of value, in four major steps (*Figure 1*). In the initial step, a new or improved analytic function is usually created or adapted for use in the enterprise by a technical specialist such as a data scientist, a model developer or a machine learning expert. In the validation step, other specialists — typically data engineers or data analysts — prepare the analytic function for production use and

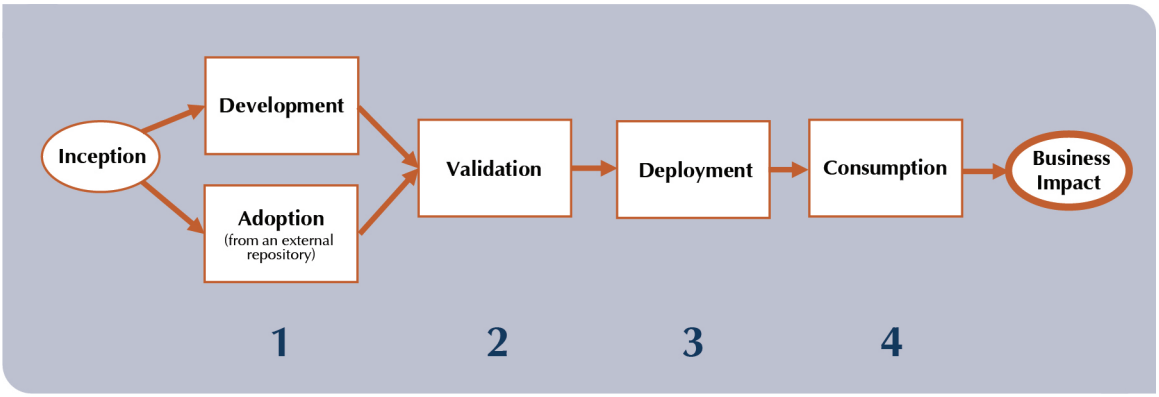

Figure 1: From Analytic Problem Conception to Delivery of Business Value

confirm that it functions correctly in the company environment. In the deployment step, yet other specialists make the analytic function available to business users. In the final consumption step, the function is applied to live data in an effort to produce a business result.

Unfortunately, streamlining this process is only the beginning. Few complex analytics produce the desired result the first time: glitches in the function, the implementation or the data often generate unexpected results. This means that one or more of the above four steps is often repeated until the desired result is obtained, with data quality and data integration problems providing a notorious source of delay and iteration.

While iteration may slow things down, there is a larger issue: lack of reuse. 80% of the time, a similar problem — if not exactly the same problem — has already been solved before, either entirely or in part.

This leads to the final source of delay: searching for existing apps or components that solve the problem in whole or in part. As noted above, the proliferation of data platforms on which an analytic function can be developed and saved means more places to search, and organizations rarely take the time to develop a company-wide catalog of such analytic functions and procedures.

Often a similar need arises elsewhere in the company some weeks or months after the initial development of an analytic application. In these subsequent attempts at deployment and consumption, the search delay occurs again. People ask: wasn't something like this done a few months ago? If so, where did they put it?

In practice, this causes the process to look more like *Figure 2*, in which search delay and iteration interact to result in a large extension of the time and effort required for the end-to-end process.

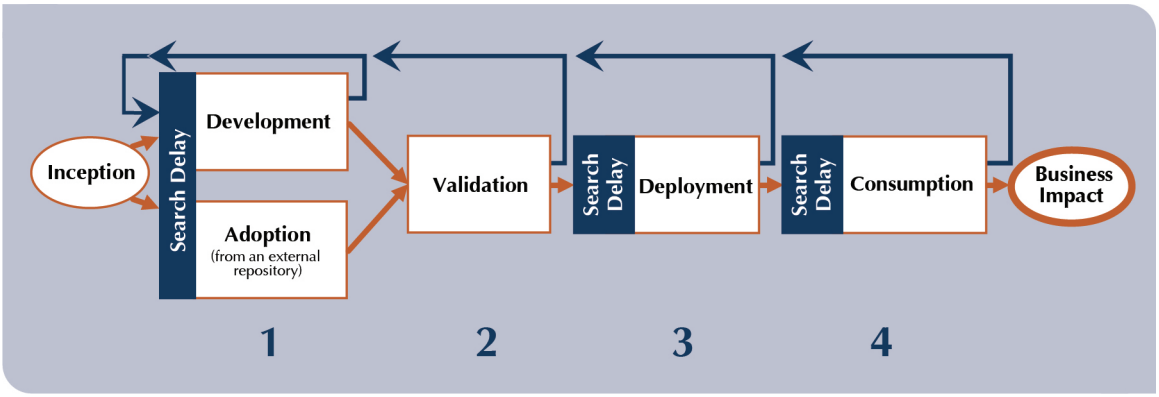

Figure 2: From Analytic Problem Conception to Delivery of Business Value, with Iteration & Delays

Unfortunately, even the slow and meandering path to business impact in *Figure 2* is not actually the worst part of the problem. In a misguided effort to save time, the search for existing solutions is often abandoned. This results in a waste of effort and delay, when a previously deployed and proven solution could be reused. Instead, because the search fails frequently, step 1 is repeated in its entirety. This replication of effort is almost guaranteed when the business user is organizationally or geographically removed.

Today, with the ever-increasing diversity and power in the tools and platforms available for data analytics, we might imagine that analytic progress is getting easier to come by. Unfortunately, the opposite is true in many companies. More tools and solutions are available, **but it takes more time and effort than ever to find and apply the right tool or solution to a given problem and actually deploy it to business advantage**. ●

It's clear that a complex multi-step process fraught with search delays and other problems, is a major source of friction in moving from inception to impact in the analytic development process. This, together with the complexity of the modern data environment, with its many platforms and locations, can really slow down the implementation of analytics for business impact.

The next section will examine a response to these and other issues in the deployment, consumption and reuse of analytic apps and microservices: Teradata AppCenter.

2 Introducing Teradata AppCenter

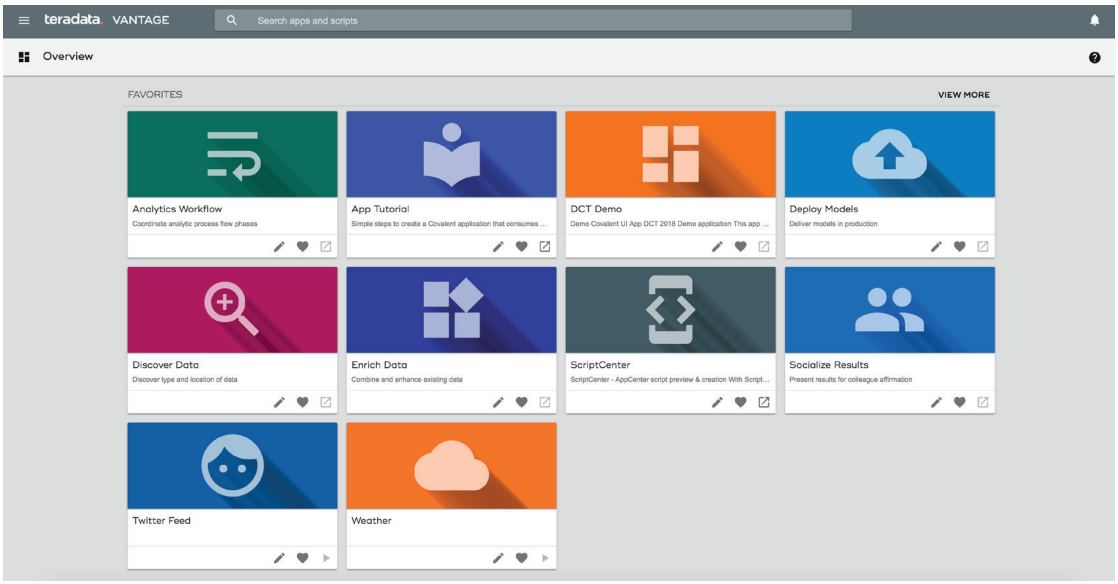

Figure 3: A Sample Teradata AppCenter Screen

Teradata has introduced Teradata AppCenter, a new self-service web-based platform that puts analytics into action. Available either standalone or as a bundled component of Teradata Vantage, AppCenter provides a front-end environment for users to rapidly access analytics.

Teradata AppCenter empowers developers, data scientists and end users to easily develop, deploy and reuse analytic apps across the enterprise, without sacrificing power and flexibility. Teradata AppCenter dramatically reduces delays and replicated effort in the end-to-end process of creating and consuming analytics, which in turn reduces time-to-value and increases return on investment. In addition, Teradata AppCenter makes it much easier to adopt externally developed analytics, including leveraging open source.

The effect is similar to that achieved with smartphone app stores, which have encouraged the widespread development and use of "apps" via the combination of a common framework, a central exchange and a self-service interface.

The fundamental concepts of Teradata AppCenter are:

1. Teradata AppCenter provides a single place where all authorized users can look for, find, and launch *analytic* apps. As used here, an *analytic* app is an intuitive, portable, reusable software component used to derive insights from data.
2. AppCenter apps can access data from a wide variety of platforms including Teradata Vantage and Hadoop Clusters, as well as Spark, Oracle, Kafka and others via Teradata QueryGrid — on-premises and in the Teradata Cloud. AppCenter apps can also access a variety of data stores, including SQL Server, MySQL, PostgreSQL, Cassandra, Kafka and Oracle. In addition, Hive tables can be accessed.
3. All types of analytic data users are supported by the Teradata AppCenter: business users, data analysts, data engineers and developers, and data scientists.

4. All authorized users can deploy analytic functions and solutions via the Teradata AppCenter, making them available across the global enterprise, if desired. Thus, Teradata AppCenter promotes the widespread reuse and sharing of common analytic components and unleashes creativity with data and analytics across the organization.
5. Teradata AppCenter provides an appealing graphical user interface that enables self-service for all users whereby they can readily browse, search and identify functions and solutions.
6. Teradata AppCenter provides tagging to facilitate finding solutions relevant to a particular need or interest.
7. Teradata AppCenter is founded on a framework for application deployment that simplifies the use of analytics on a variety of platforms. A key element of this is containerization. By leveraging the advanced infrastructure of Docker containers and Kubernetes, applications can be created in AppCenter and rapidly deployed using Docker containers, with minimal effort. Additionally, developers can import any external dockerized engine or workbench, such as Jupyter notebooks, Tensorflow or Spark, into the framework.
8. Existing SQL and Batch Teradata Query (BTEQ) scripts can be deployed via AppCenter, making them available via self-service interface. This helps Teradata customers to more fully leverage their existing investment in such scripts.
9. Teradata AppCenter enables data visualization. The creator of the app can use any of the common data visualizations built into AppCenter in addition to third-party tools and externally sourced data visualization functions.

The first principle of Teradata AppCenter — that it provides a single place to look for analytic functions and applications — has a surprising power. A business user with an analytic requirement has one place to look so that he or she can take advantage of any prior work that has been done anywhere in the company, at any time, since AppCenter has been in use.

Other users also benefit from the centralization. Data scientists working on a new problem can look in Teradata AppCenter to see what data science has been done before on that problem. The new problem can leverage what is already available and build on top of it with further improvements.

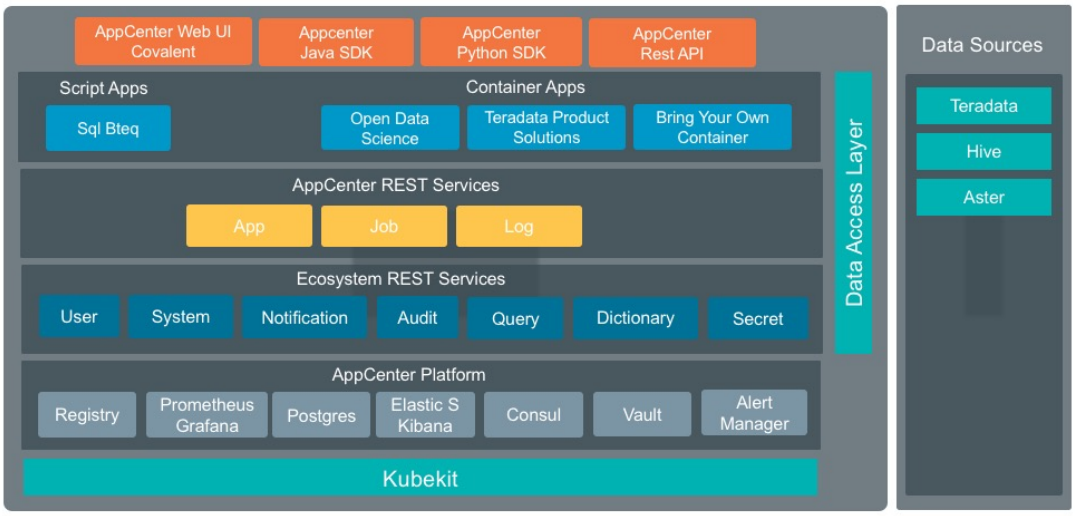

Figure 4: Teradata AppCenter Platform Architecture

As well as a complete solution for a newly stated requirement, the data scientist could also find useful components of a solution. For example, if the data scientist wants to work on customer retention, he or she may benefit from any existing components that score satisfaction or predict lifetime customer value.

The same principle applies for data engineers, developers and database administrators. At every step in the analytic solution development lifecycle, having a central repository of apps adds value and confers efficiency on the process, eliminating the search delays shown in *Figure 2* in *Section 1*.

As depicted in *Figure 4*, Teradata AppCenter provides a framework for deployment of analytic apps created with a variety of development tools, workbenches and languages. A major advantage of AppCenter is that data scientists are able to leverage a scalable analytic data environment with shared services, data access and deployment.

This is vastly different than the typical data science environment in which this process is attempted on a desktop or a small server. First, it is impossible to implement or even effectively test at scale in the typical development environment. Second, such issues as data access and security are much more labor intensive and less standardized across the enterprise when solved independently for each project.

In addition, apps can draw upon a wide range of open analytic frameworks such as Spark Mllib, Tensorflow, H2O.AI and others. As previously noted, AppCenter apps can access data stored in a variety of data platforms or streaming in via Kafka or Spark leveraging Teradata QueryGrid.

Teradata AppCenter also provides built in capabilities for creating apps, running apps, managing the AppCenter environment and incorporating existing apps that have been containerized with Docker.

Purpose and Methodology for this Report

This WinterCorp Report describes the Teradata AppCenter and its significance to customers for development and enterprise deployment of analytics. In developing this report, WinterCorp drew on its own independent research and experience, interviewed Teradata developers and witnessed live AppCenter demonstrations. Teradata was provided an opportunity to comment on the paper with respect to facts, in its capacity as the sponsor of this research. WinterCorp has final editorial control over the content of this publication and is solely responsible for any opinions expressed.

WinterCorp

3 Usage Examples

Online Advertising. A data scientist has come up with a new algorithm in R to improve the performance of an ad targeting engine already in use in the company. The data scientist wants to make the algorithm available to developers for A/B testing with the ad server, using data stored in Teradata Vantage. With a few hours of effort, the data scientist can readily deliver the new algorithm in a Docker image via AppCenter. Developers can then run the required A/B tests and, if positive, deploy the algorithm to production, greatly reducing the overall time and effort to realize the business result. If the A/B tests indicate that the algorithm needs adjustment — or that a different algorithm is called for — the cycle of revision and testing is greatly accelerated via the containerization supported by AppCenter. Prior to AppCenter, the same process would have taken days or weeks for each iteration. Note that the algorithm need not have been developed with Teradata in mind. The containerization of the algorithm and the use of the AppCenter deployment framework enable porting of such independently developed data science and machine learning functions directly into the Teradata ecosystem.

Market Segmentation. A data analyst needs to create an app that can be used to target customers for a new marketing campaign. Variations of the campaign will be run in different countries and for different lines of business. The query will produce two results: a table that the business user can then feed to a campaign management app, and an online data visualization that the user can examine and explore to see if the target set looks right for his or her purpose. The data analyst decides to write a parameterized SQL query to accomplish this and to deploy it via AppCenter, along with instructions on how to use it. The resulting app will actually retrieve data from Teradata Vantage, S3, and from a Hadoop cluster. However, the user of the app need not be aware of this. The user will simply click on the icon in AppCenter to launch the app. The app will then prompt for the needed inputs, such as country and line of business. It will run the encapsulated SQL query, retrieving the data from the multiple platforms, joining it, analyzing it and producing the required tabular output. In addition, the app will feed the result table to the data visualization function built into AppCenter. The data analyst will tag the query to make it easy for business users to find it. Using capabilities built into AppCenter, the data analyst will be able to track usage to get a sense of how often it is being used and how much effort he or she should spend maintaining and/or improving it. If a user has a question about the query, he or she can contact the developer via AppCenter. Note that the facility to deploy existing SQL and BTEQ scripts mean that established Teradata customers will find that much previous work can be deployed via AppCenter and thus become available via its self-service interface. With tagging and the built-in AppCenter search, such existing scripts will be much more easily found by those who could benefit from them.

Risk Management. An application developer in a bank wants to try out a new fraud model to help the company run its risk management processes on a larger scale (that is, use a fraud model that takes in a larger set of input data in an effort to predict fraud earlier and with more accuracy). The new fraud model will use data from Teradata Vantage on the customer; the customer's accounts; and, the customer's history of transactions with the bank. In addition, it will consider the history of the customer's visits to the bank's websites, which is stored in web logs on Hadoop. To create the app, the developer will use the Teradata AppCenter SDK to perform such common tasks as user login, audit trail and data access. Via the SDK, it will be straightforward to access the data on Vantage and Hadoop via the Teradata Analytical Ecosystem. Using the SDK, the developer will actually create multiple versions of the app: one each for fraud modelers, the compliance department and the line of business executive. Using built-in capabilities of AppCenter, the developer can provide the degree of data access appropriate to each class of user.

4 Conclusions and Recommendations

Teradata AppCenter provides unique capabilities to empower all participants in the process of developing and applying analytics for business impact: each participant is made more productive and the end-to-end analytics process is accelerated; reuse and standardization of analytic components and solutions is encouraged; and business users are provided with complete solutions via a self-service interface that incorporate data access, analysis and/or machine learning and data visualization.

For the first time, business users can go to one place in the enterprise; browse for apps relevant to the problem at hand; search for apps via tags; and, when an appropriate app is located, run it with the click of a mouse. Apps that require input at run time will ask the user to supply the missing parameters, and data visualizations, when needed, will be built into the app and automatically displayed to the user. Remarkably, the business user does not need to know where the data has been stored; where the analytic will run; or, where the analytic came from — though all this information will be available in the online documentation of the app in the AppCenter.

As discussed, while the AppCenter is empowering for business users, its effect for developers may prove to be even more dramatic. App developers benefit from the existence of a single repository for analytic apps and their components, by sharing and reusing their work. AppCenter makes available documentation and the means to run the app easily, and AppCenter provides a framework for common tasks such as security, logging, and connecting to data sources. Analytics already available in Docker containers, which may be sourced from a wide variety of libraries and open-source repositories, can be readily added to AppCenter and used on data in Vantage. Additionally, AppCenter apps can access data virtually anywhere in the analytic data ecosystem of the enterprise, whether on Teradata systems, Hadoop, Spark, under Kafka or a variety of other data platforms.

Teradata AppCenter thus streamlines the end-to-end process of creating analytics and applying them to generate business impact. It enables a modern data ecosystem that encourages the adoption of the best analytic solutions across and from outside of the enterprise; delivers a self-service interface to all of the participants in the analytic process empowering them to do more and to get it done more rapidly; and provides a unifying framework for analytic apps and a standard way of doing certain common functions.

Indications to date are that Teradata AppCenter will yield significant dividends in the use of analytics to produce business results, both by accelerating time-to-value and by empowering developers and users via self-service. ●

WinterCorp is an independent consulting firm expert in the architecture and strategy of the modern analytic data ecosystem.

Since our founding in 1992, we have architected and engineered solutions to some of the toughest and most demanding analytic data challenges, worldwide.

We help customers define their data-related business interests; develop their data strategies and architectures; select their data platforms; and, engineer their solutions to optimize business value.

Our customers, with our help, create and implement cloud, multi-cloud and hybrid cloud architectures; they create the data foundation needed for data science, artificial intelligence and machine learning.

Our customers get business results with analytics in which their return is often ten or more times their investment.

When needed, we create and conduct benchmarks, proofs-of-concept, pilot programs and system engineering studies that help our clients manage profound technical risks, control costs and reach business goals.

We're expert with structured data, unstructured data, and semi-structured data — with the products, tools and technologies of data management for data analytics in all its major forms.

With our in-depth knowledge and experience, we deliver unmatched insight into the issues that impede scalability and into the technologies and practices that enable business success.

WinterCorp
www.wintercorp.com
42 DERBY LANE, TYNGSBORO, MA
617-695-1800