

David Chappell

SELLING PROJECTS ON THE MICROSOFT BUSINESS ANALYTICS PLATFORM

A PERSPECTIVE FOR SYSTEMS INTEGRATORS



DavidChappell
& Associates

Sponsored by Microsoft Corporation

Copyright © 2014 Chappell & Associates

Contents

- Business Analytics Today 3**
- Data Warehousing 3**
 - Microsoft Data Warehousing Technologies.....3
 - Example Projects4
 - Creating a Relational Data Warehouse5
 - Deploying a Data Warehouse Appliance.....6
 - Processing Big Data with a Hadoop Cluster7
 - Building a Modern Data Warehouse with Relational Data and Hadoop.....8
- Business Intelligence 9**
 - Microsoft BI Technologies9
 - Example Projects10
 - Implementing Corporate BI.....10
 - Supporting Self-Service BI10
- Data Integration..... 12**
 - Microsoft Data Integration Technologies12
 - Example Projects13
 - Implementing a Traditional ETL Solution13
 - Supporting Self-Service Integration13
- Summary..... 13**
- About the Author 16**

Business Analytics Today

The business analytics market has changed. Five years ago, a systems integrator needed to know how to create relational data warehouses and build corporate business intelligence solutions. Today, you still sell these projects—they might even be the majority of your current revenue—but you also offer a broader range of services.

Rather than using just relational data, for example, your customer might also need to work with large amounts of non-relational data, presenting you with a big data problem. Or they might want you to combine relational and non-relational data, creating a data warehouse that spans both. Instead of a traditional corporate BI project initiated by the CIO, you might be asked to help deploy self-service BI for a single department. And any of these things might need to be in the public cloud, depending on where the data lives and what your customer wants to do with it.

As customer requirements change, the platforms you depend on must also change. Today, for instance, the Microsoft business analytics platform includes support for big data, self-service BI, cloud deployment, and more. What follows gives a high-level view of this platform, then illustrates the kinds of projects your firm can sell on it. And while this survey looks at the big picture, the focus is on the newer scenarios.

As Figure 1 shows, the basics of a business analytics platform haven't changed; it's still possible to group the main technologies into three categories.



Figure 1: Business analytics platforms have three main parts.

Organizations still need to create data warehouses using data integration technologies, then analyze that data with business intelligence tools. What's expanded significantly is what's in those boxes and the kinds of projects you can now sell. The rest of this overview looks at these technologies and projects, grouping them into the three categories shown in the figure.

Data Warehousing

Two big trends are driving data warehousing today. First, the cost of storage continues to drop like a stone, making it feasible to store lots more information. At the same time, the number of possible data sources and the amount of diverse data they generate has increased enormously. Taken together, these two trends have had a big impact on the traditional world of data warehousing.

Microsoft Data Warehousing Technologies

To address these changes, the Microsoft business analytics platform has expanded in several ways. Figure 2 shows what it provides today for data warehousing.

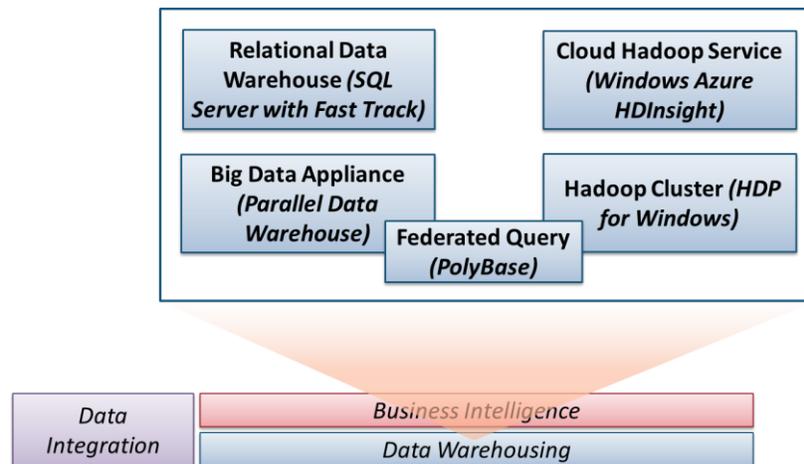


Figure 2: The Microsoft business analytics platform includes data warehousing technologies for relational and non-relational data.

The technologies and the problems they address are:

- For traditional relational data warehouses, the platform includes **SQL Server**. Microsoft also provides a **Fast Track** reference architecture, offering pre-defined hardware and software configurations for creating these warehouses.
- For very large data warehouses that can benefit from parallel processing, Microsoft offers **Parallel Data Warehouse (PDW)**. Sold as a hardware appliance, PDW can support both relational and non-relational data, offering federated query across both with a technology called **PolyBase**.
- For working with large amounts of non-relational data in an on-premises cluster, the Microsoft platform includes **Hortonworks Data Platform (HDP) for Windows**. This technology implements standard Apache Hadoop on Windows Server. Once again, PolyBase provides federated query, letting applications issue a single SQL query that returns results from both relational data in PDW and non-relational data in HDP for Windows.
- For working with large amounts of non-relational data in the public cloud, Microsoft offers **Windows Azure HDInsight**. It provides Hadoop clusters as a managed service, letting your customer pay only for the computing power and storage they use.

Example Projects

What kinds of data warehousing projects can you sell on these technologies today? There are four main categories:

- **Relational data warehouses** with SQL Server
- **Large-scale data warehouses** with PDW
- **Big data projects** with HDP for Windows and HDInsight
- **Modern data warehouses** that combine these technologies.

Creating a Relational Data Warehouse

Just about every SI that provides business analytics solutions sells data warehousing projects today. Those that use SQL Server commonly choose it for a few reasons:

- Microsoft focuses on providing easy-to-use tools, which can make SQL Server solutions more attractive for both the creators of those solutions and the people who use them.
- Unlike many vendors in this market, Microsoft doesn't have a significant consulting business. SIs that build on SQL Server won't find themselves competing for projects against their software provider.
- SQL Server costs less than many other solutions. It also includes BI components—reporting, analysis, and more—for which other vendors charge separately. This lowers the overall cost of the project, which can help an SI win deals against competitors using other technologies.

Project: Creating an on-premises data warehouse. While you're free to use SQL Server to create data warehouses from scratch, using the Fast Track option can make this simpler. Fast Track provides pre-defined configurations for:

- Hardware, including specifications for servers, storage, and networking. Various hardware providers can be used, including HP, Dell, IBM, Cisco, and others.
- Software, including SQL Server and Windows Server.
- Database, defining the architecture, configuration, performance settings, and more.

Because Fast Track makes it faster to create a new data warehouse, SIs can offer their customers projects with shorter completion times.

Project: Creating a data warehouse in the cloud. An even quicker way to make a data warehouse available to your customers is to build it on Windows Azure. This cloud platform provides standard Windows Server images that let you create new virtual machines running SQL Server. These images are configured with the relevant aspects of Fast Track, making them a good fit for supporting a data warehouse in the cloud.

Rather than waiting for hardware, then installing and configuring software, you can have a data warehousing foundation available in just a few minutes. This allows faster completion of some projects, and it can also provide a quicker proof of concept when you're competing for a new deal.

Deploying a Data Warehouse Appliance

Sometimes, creating a data warehouse using just SQL Server isn't enough. If a project requires handling lots of relational data—many terabytes or even a few petabytes—a dedicated hardware appliance is probably a better solution. Parallel Data Warehouse is designed for this situation.

PDW is just what it sounds like: a cluster of machines that store and process data in parallel. Those machines can come from Dell, HP, or other hardware vendors, and the approach PDW uses is called *massively parallel processing (MPP)*. While PDW was originally just for relational data, the current version also allows storing and working with non-relational data.

Project: Making an existing data warehouse more scalable. Many of your customers probably have mature data warehouses in place, especially if you work with large organizations. For those that need more scale—and eventually, this will be most of them—adding PDW can make sense. Even though it's still common to think of SQL Server as a departmental-scale system, PDW can handle up to five petabytes of data. And as with SQL Server, PDW is typically less expensive than its competitors.

PDW's MPP approach brings a different paradigm. Even if your customer is already using SQL Server, they can't just move everything to PDW unchanged. And this migration isn't a simple wizard-based process; it requires specialized skills. This is why Microsoft always sells PDW together with a consulting engagement. SIs are essential for helping customers learn to use the system and move their data to this new environment.

Project: Creating a new large-scale data warehouse, perhaps combining relational and non-relational data. PDW allows creating very large relational data warehouses. But the majority of new data created today isn't relational—it's unstructured. For analyzing large amounts of unstructured data, our industry has settled on Hadoop. Because both kinds of big data are important, PDW allows creating a Hadoop partition directly in the appliance.

Once this is done, applications can use PolyBase to submit ordinary SQL queries against relational data, non-relational data, or both. This lets your customers use Excel and other common tools to work with large amounts of heterogeneous data.

Project: Improving reporting times. Most often, PDW is used with large data warehouses, starting around 100 terabytes. But some customers use it with smaller amounts of data, as little as six or seven terabytes. The reason is that PDW's MPP approach can allow running reports significantly faster. When shorter reporting times have high business value, PDW can help your customers address a business need that can't easily be solved in other ways.

Processing Big Data with a Hadoop Cluster

While relational data is certainly important, the largest share of new data created today is unstructured. The least expensive approach to analyzing large amounts of purely unstructured data is often to create a Hadoop cluster.

Why might your customers do this? There are lots of scenarios, spread across many vertical markets. Examples include the following:

- Retailers analyzing clickstream data from their online stores, sensor data from their warehouses, or customer data from point-of-sale terminals.
- Marketing professionals performing sentiment analysis, i.e., understanding what customers are saying about their products by analyzing comments on social media.
- Managers of large datacenters doing automated analysis of server logs to find security problems or other issues.
- Financial analysts examining large amounts of unstructured data to do better risk management.
- Analysts in telecom companies searching for customer patterns across large numbers of call detail records.

Project: Creating an on-premises Hadoop cluster. To work with large amounts of unstructured data that's stored in an organization's own datacenter, the best option might well be to create an on-premises Hadoop cluster. HDP for Windows supports this scenario, letting your customers use common tools to work with the data: Excel, SQL Server Analysis Services, and others.

As you might expect, HDP for Windows is also integrated with the rest of the Windows environment. It allows using Active Directory authentication, for instance, and can be managed with System Center. Especially for Windows-oriented organizations, creating a Hadoop cluster on Windows Server can be a better fit than building one that's based on Linux.

Project: Creating a Hadoop cluster in the cloud. To work with large amounts of unstructured data that's stored on Windows Azure, the best option is likely to be HDInsight. Unlike an on-premises cluster, which requires buying and managing a fleet of physical servers, HDInsight offers a managed cloud service with pay-as-you-go pricing. The data it works with is stored in Windows Azure blobs, a very inexpensive storage option, and so relying on HDInsight is likely to be the cheapest way to use Hadoop in the Microsoft world.

HDInsight also makes setting up and managing a Hadoop cluster much easier—it's done for you—which lets you sell big data solutions to less sophisticated organizations. And because customers can pay for a cluster only when they're using it, this cloud solution also lets you sell projects with a lower price tag. Even smaller organizations can now afford to use Hadoop to process big data in the cloud.

Building a Modern Data Warehouse with Relational Data and Hadoop

Some organizations will need every option: traditional relational warehouses, very large warehouses that use an appliance, and Hadoop clusters. Going forward, the idea of a modern data warehouse will encompass all of these things.

Project: Creating a scalable data warehouse for mixed data. PDW supports petabyte-scale relational data. HDP for Windows supports petabyte-scale unstructured data. An organization working with both needs to be able to use them together. To allow this, Microsoft's PolyBase technology lets applications issue SQL queries that span both PDW and HDP for Windows. (As described earlier, these queries can also be used with relational and non-relational data stored in PDW.) This lets the organization use Excel and other familiar tools to work with big data of all kinds in a consistent way.

Providing a solution that works consistently across all kinds of big data and that fits into existing Microsoft infrastructure gives you an attractive offering to sell. And since PDW and HDP for Windows both require consulting services, projects using that offering can generate significant revenue for you.

Not Just Data—A Data Strategy

The big trends in data warehousing today can be summarized in just one word: *More*. All of your customers have lots of data today, and they'll have lots more tomorrow. Data warehouses can help them use this data more effectively, but what they really need is a data strategy. Whatever data warehouse project you're doing, you should also be helping your customers define an overall approach for dealing with their diverse data in a consistent way.

A key part of this is matching the cost of storage to the value of the information. What data is worth storing in PDW, for example, rather than a cheaper solution such as an on-premises Hadoop cluster? And when should data be moved to Windows Azure blob storage, the cheapest option of all? Most organizations will need most of these solutions—each one makes sense in some situations—and the Microsoft platform supports all of them. Expect your customers to look to you for help in creating a long-term strategy for when to use which technology.

Business Intelligence

Collecting data in a data warehouse isn't an end in itself—it's a means to an end. The real goal is to analyze that data in some useful way. This is why customers buy the business intelligence solutions that your firm sells, and it's also why the Microsoft business analytics platform includes an extensive set of BI technologies.

Microsoft BI Technologies

It's useful to group Microsoft's BI offerings into two categories: corporate BI and self-service BI. Figure 3 shows the main components of each.

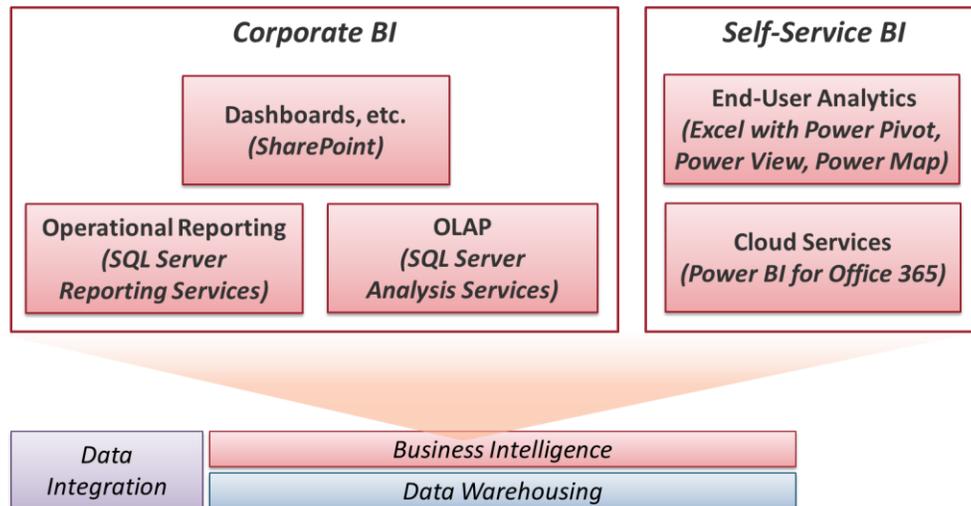


Figure 3: The Microsoft business analytics platform supports corporate BI and self-service BI.

Corporate BI projects are typically done by technically skilled people—yours or your customer's—and an organization's IT department is commonly involved. Microsoft's offerings in this area include these:

- For operational reporting, the platform provides **SQL Server Reporting Services (SSRS)**. This technology is included with a SQL Server license.
- For OLAP, Microsoft provides **SQL Server Analysis Services (SSAS)**. Like SSRS, SSAS is included with a SQL Server license.
- For analytics dashboards and other ways to display information, the Microsoft platform relies on **SharePoint**.

Unlike corporate BI, self-service BI can commonly be set up and used by mainstream business users (although there's still work here for SIs, as described later). While these solutions aren't typically as scalable as corporate BI applications, they're fast and relatively simple to create, which makes them appropriate for many departmental scenarios. To support this style of end-user analytics, the Microsoft platform includes these technologies:

- To work with data in a familiar way, users rely on **Excel**. This desktop tool also provides the environment in which many of the other self-service BI technologies run.

- To perform data analysis, the platform includes **Power Pivot**. An Excel add-in, Power Pivot is the desktop analog to SQL Server Analysis Services.
- To provide data visualization and reporting, Microsoft offers **Power View**. Also an Excel add-in, one way to think of this technology is as the desktop analog to SQL Server Reporting Services.
- To let users visualize geographic data, the Microsoft business analytics platform includes another Excel add-in called **Power Map**.
- To let users share data, submit natural language queries against that data, and more, Microsoft provides a cloud service called **Power BI for Office 365**. (In fact, Microsoft commonly uses the term “Power BI” to refer to all of its self-service BI offerings.)

Example Projects

It’s useful to group the kinds of BI projects you can sell today into the two categories we’ve already seen: corporate BI and self-service BI. This section looks at projects in both areas, focusing more on the newer self-service options.

Implementing Corporate BI

In most companies, the reports, dashboards, and other business analytics used company-wide rely on corporate BI technologies. The projects you sell in this category can be big, and they might well provide the majority of your firm’s revenue today.

Project: Providing an on-premises corporate BI solution. This category includes many things: creating reports with SSRS, analyzing multi-dimensional cubes or doing data mining with SSAS, and others. Even though the boundaries of BI have expanded considerably, this kind of traditional project is likely to remain a mainstay of your BI business, at least for the next few years.

Project: Providing a corporate BI solution in the cloud. Running applications on public cloud platforms surely represents a large part of the future for most organizations. This is just as true for BI applications as for any other category. Maybe your customer has established a cloud-first policy, for example, where new applications run in the public cloud unless there’s a compelling reason not to. Or maybe the project can be completed in less time or for less money if it runs in the cloud. Whatever the reason, the Microsoft platform lets you sell corporate BI projects that rely on SSAS, SSRS, and other technologies running on Windows Azure.

Another option that every SI should at least consider is providing reusable cloud services. For example, your firm might create a Windows Azure application that uses SSAS to analyze retail data. By selling this as a subscription-based cloud offering to multiple customers, you might be able to expand your business into new areas. And since selling applications commonly provides better margins than selling consulting services, you might also be able to improve your firm’s profitability.

Supporting Self-Service BI

The success of self-service BI results to some degree from the failures of corporate BI. Large-scale data warehouses supporting company-wide reporting and analysis are great when they work. Too often, though, these projects take too long, cost too much, and deliver too little. For a brand manager who wants to understand what’s happening

with her product, the manager of a distribution center trying to understand trends in his local environment, and many others, corporate BI is overkill. Something quicker, cheaper, and simpler—self-service BI—is just fine.

For SIs, the rise of self-service BI might be seen as a mixed blessing. While corporate BI projects commonly require lots of effort and plenty of billable hours, what kinds of projects can you sell using a “self-service” technology? The answer is that, while they’re commonly smaller than traditional corporate BI projects, there are still plenty of scenarios where SIs can provide necessary services for organizations using self-service BI.

Project: Creating a shared data model. Just as corporate BI commonly relies on a data warehouse, self-service BI must also rely on stored data. For simple scenarios, that data can be stored in an Excel workbook. For more scalable and more sharable situations, however, that data can be stored in SQL Server Analysis Services. Alongside traditional multidimensional cubes, SSAS also provides a tabular data model that can be used to do this. Among other things, this model lets self-service BI users see data in business terms they can understand rather than as relational tables.

Creating this sharable model in SSAS takes some technical skill, however—it’s not likely to be something that a typical self-service BI user can do. Accordingly, it’s an opportunity for SIs to provide value to their customers.

Project: Configuring self-service BI cloud services. As you’d expect with a self-service technology, your customers can set up Power BI for Office 365 on their own. But there are extra configurations that provide real value while requiring more technical skill. For example, Power BI for Office 365 includes a data catalog that allows searching data shared through this cloud service. It’s possible, however, to connect this catalog to on-premises data sources within an organization. This lets searches cover both data shared through the cloud and data stored on premises. Doing this requires setting up connections between the catalog and the corporate data sources, creating indexes, and other consulting services that an SI can provide.

Project: Optimizing natural language queries. One of the most interesting parts of Power BI for Office 365 is a feature called Q&A. A user can type a natural language question into a question box and get a formatted result. For example, a user might enter “total sales by hour for diapers and shampoo as a line chart”. Q&A will return the result of this search in the format requested.

Without any customization, Q&A can handle a range of natural language queries. Organizations and industries have specialized vocabularies, however, with terms that map to data models in particular ways. An SI can help its customers create, tune, and maintain these vocabularies, making their Q&A searches more effective. An SI might even build a specialized vocabulary for, say, the packaged goods industry, then sell that vocabulary to multiple customers in this vertical.

Complex Event Processing with StreamInsight

Most of business analytics works with data at rest, i.e., information stored on disk. In complex event processing (CEP), however, applications work with data in motion, such as information sent over networks, that might never be stored anywhere.

For example, think about the data generated by a sensor on an oil well. A CEP application can analyze this data as it's sent, looking for trends. This can allow much quicker response to problems than waiting for the data to be stored, then analyzed later. Or suppose somebody is using the same credit card an hour apart in San Francisco and Munich. Discovering this immediately with a CEP application has more value for fraud prevention than learning about it through traditional BI the following week.

To support scenarios like these, the Microsoft business analytics platform provides **StreamInsight**. While CEP isn't (yet) a widely popular approach, it has a lot to offer, and so StreamInsight is an important part of this technology family.

Data Integration

Business analytics depends on business data. With corporate BI, data is typically kept in a data warehouse created through some set of extract, transform, and load (ETL) operations. With self-service BI, the data being analyzed must be copied into a workbook or SSAS. In either case, data integration technologies are commonly required.

Microsoft Data Integration Technologies

As Figure 4 suggests, Microsoft's data integration technologies can be thought of in the same two categories as BI: corporate and self-service.

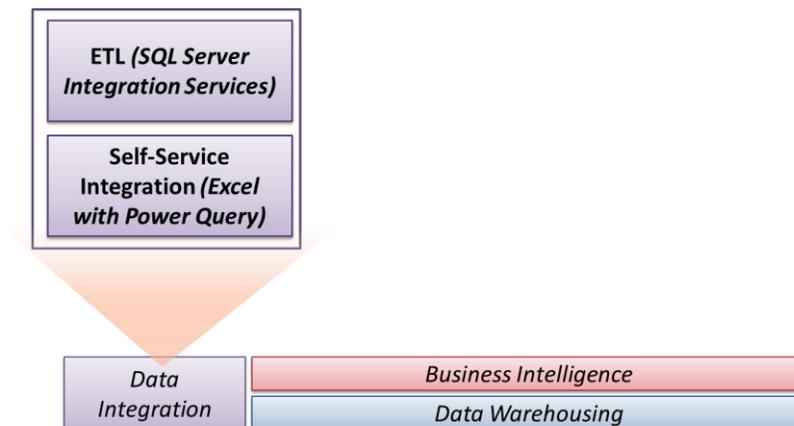


Figure 4: The Microsoft business analytics platform provides tools for both traditional extract/transform/load and self-service data integration.

The technologies are these:

- For traditional ETL, organizations can use **SQL Server Integration Services (SSIS)**. SSIS is a mature, widely used offering, and many SIs have extensive experience with it.
- For data integration with self-service BI, the Microsoft platform provides **Power Query**. Like Power Pivot, Power View, and Power Map, Power Query is an Excel add-in.

Example Projects

Data integration is a hard problem. Organizations commonly have lots of diverse data, with similar (or even the same) information stored in different ways in different places. In the last few years, the problem has gotten even harder. More kinds of data now exist, relational and non-relational, in more formats. At the same time, external data has become more important: Twitter and Facebook data, weather data, geographic data, and more. Getting all of this into a consistent and useful form isn't simple. This is why data integration has been—and will continue to be—an important source of project revenue for SIs.

Implementing a Traditional ETL Solution

Project: Providing data integration for a relational data warehouse. Projects like this are a mainstay for many SIs with a business analytics practice. With SSIS, the project commonly involves creating an automated workflow that regularly extracts data from various operational databases (SQL Server or others), transforms that data into a common format, then loads it into a data warehouse. Doing all of this takes effort, especially for larger organizations with lots of diverse data, and so these are attractive projects for SIs.

Supporting Self-Service Integration

Project: Providing data integration for self-service BI. Even though it's aimed at simpler scenarios, self-service BI needs to address the same set of problems as corporate BI. Just as Power Pivot is analogous to SSAS and Power View is analogous to SSRS, Power Query is analogous to SSIS.

Power Query can be used directly by business users to access many kinds of data: relational data, data in files, data stored in Hadoop, public data in the Windows Azure Marketplace, Facebook data, and more. Yet for more complex scenarios, an SI with broader experience in data integration can provide significant value here. This is especially true if the result of a Power Query integration will be shared among a number of people. While this kind of project might start small, it can be a beachhead for larger projects in the future.

Summary

SIs can sell a wide range of projects on the Microsoft business analytics platform. Figure 5 summarizes the scenarios in each area, along with the technologies each one uses.

This breadth is a strength of Microsoft's platform. You might first sell a new customer a small project that supports their self-service BI efforts, then move from that to larger corporate BI work. Or perhaps an existing customer wants to expand their data warehouse to include Hadoop, or move some business analytics to the cloud, or

increase their use of self-service BI. Because the Microsoft platform supports all of these, it's straightforward for your firm to sell projects that span this range. Building on a broad business analytics platform also helps you use your people's skills more effectively, since the technology foundation is the same for multiple kinds of projects.

Whatever projects you sell, it's a safe bet that the demand for business analytics will continue to grow. New capabilities, such as Hadoop and self-service BI, keep appearing, with each advance addressing new problems and new customers. All of this makes one thing very clear: It's a great time to be an SI in this market.

	Category	Project	Microsoft Technologies	
Data Warehousing	Creating a relational data warehouse	Creating an on-premises data warehouse	SQL Server with Fast Track	
		Creating a data warehouse in the cloud	SQL Server, Windows Azure Virtual Machines	
	Deploying a data warehouse appliance	Making an existing data warehouse more scalable	Parallel Data Warehouse (PDW)	
		Creating a new large-scale data warehouse for diverse data	PDW, PolyBase	
		Improving reporting times	PDW	
	Processing big data with a Hadoop cluster	Creating an on-premises Hadoop cluster	Hortonworks Data Platform (HDP) for Windows	
		Creating a Hadoop cluster in the cloud	Windows Azure HDInsight	
	Building a modern data warehouse	Creating a scalable data warehouse for relational and non-relational data	PDW, HDP for Windows, PolyBase	
	Business Intelligence	Implementing corporate BI	Providing an on-premises corporate BI solution	SQL Server Analysis Services (SSAS), SQL Server Reporting Services (SSRS)
			Providing a corporate BI solution in the cloud	SSAS, SSRS, Windows Azure
Supporting self-service BI		Creating a shared data model	SSAS	
		Configuring self-service BI cloud services	Power BI for Office 365	
		Optimizing natural language queries	Power BI for Office 365 Q&A	
Data Integration	Implementing a traditional ETL solution	Providing data integration for a relational data warehouse	SQL Server Integration Services (SSIS)	
	Supporting self-service integration	Providing data integration for self-service BI	Power Query	

Figure 5: SIs can sell a variety of projects on the Microsoft business analytics platform.

About the Author

David Chappell is Principal of Chappell & Associates (<http://www.davidchappell.com>) in San Francisco, California. Through his speaking, writing, and consulting, he helps people around the world understand, use, and make better decisions about new technologies.