

2nd Quarter 2015

# *Monticello Consulting on Data*

Business Intelligence Challenges in the Financial Services Industry



*Monticello*  
CONSULTING GROUP

**02** INTRODUCTION  
What the executive suite should understand about data

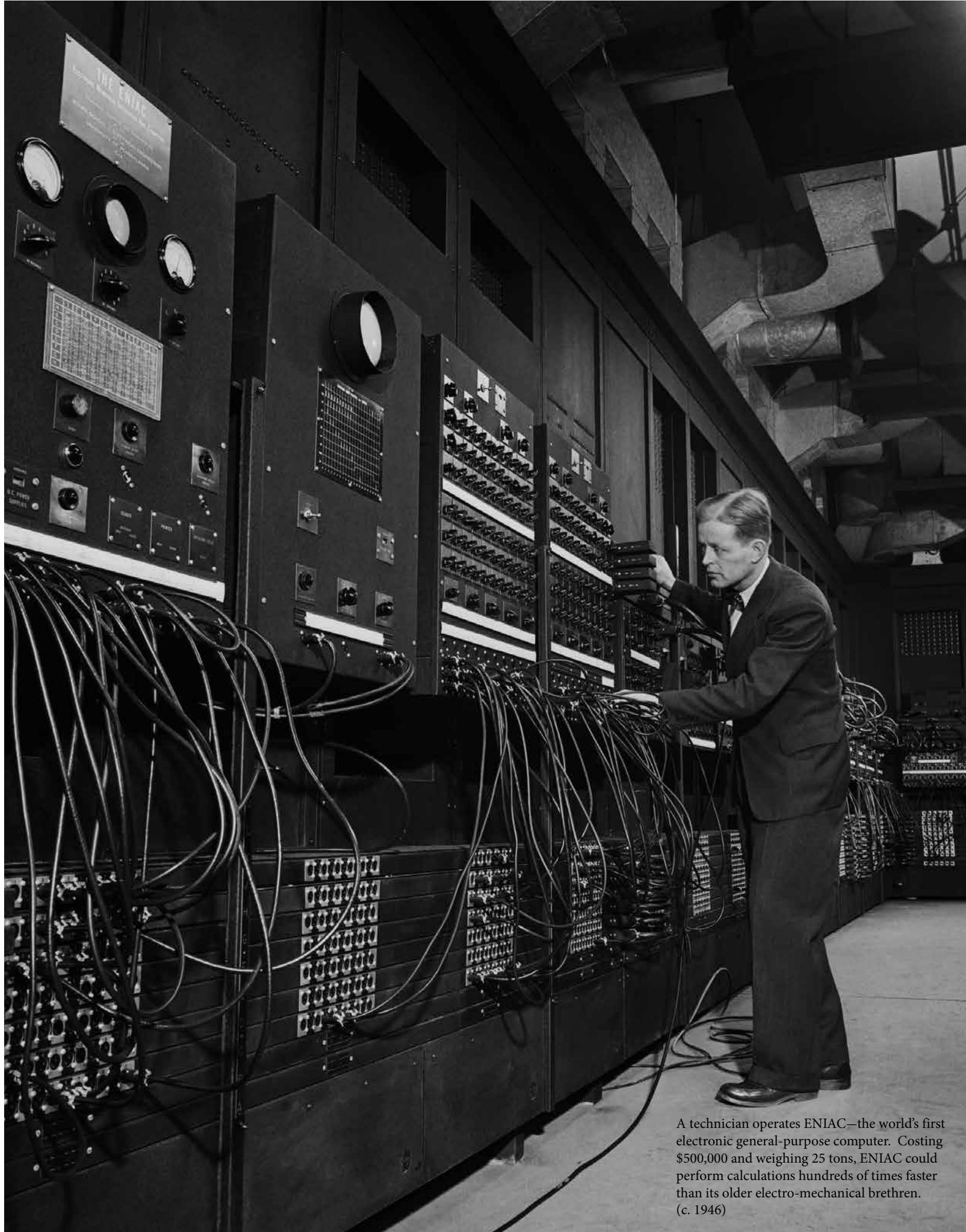
**03** ANALYSIS TOOLS  
Time-tested data analysis tools, on the go

**05** DATABASES  
Selecting the right database for the job

**07** CASE 1:  
DATA QUALITY  
A look into the nuances of preparing big banks for Dodd-Frank

**09** CASE 2:  
BROKERAGE  
Using automation to quell a growing data problem

**11** ANALYSIS TOOLS  
A hands-on look at the Goal Seek function in MS Excel



A technician operates ENIAC—the world's first electronic general-purpose computer. Costing \$500,000 and weighing 25 tons, ENIAC could perform calculations hundreds of times faster than its older electro-mechanical brethren. (c. 1946)

# What the Executive Suite Should Understand About Data

I recently attended my good friend's wedding at the Ritz Carlton in Dove Mountain, Arizona. While enjoying the company of friends I had not seen in a very long time, I ran into Frank<sup>1</sup>—an old friend and a political consultant to some of the well-known names in American politics.

"Frank! Great to see you, buddy," as I extended my hand. "What political big-wigs have you been working with recently?" During a campaign season Frank works incessantly, advising top political contenders who seek election to high office. More importantly—and for the purpose of this narrative—Frank also has a knack for commanding a room by telling highly engaging, even humorous, stories about some of the political personalities he has helped to run for office over the years. On that warm night in the Arizona desert, Frank didn't disappoint and proceeded to share a few of his war stories about working in an advisory capacity for Mitt Romney's 2012 presidential campaign.

It's funny how drinking top-shelf liquor and debating politics seem to go hand in hand, especially when there are plenty of well-heeled East Coasters attending a wedding cocktail hour. That night it wasn't long before Frank had a congregation of *self-proclaimed* political experts gathered around the bar, listening—and in some cases disputing—his blow-by-blow account of the final months of the 2012 presidential election.

In the end, we all know how that presidential race ended, but I give Frank

a lot of credit for fair storytelling and a mostly unbiased account of how his team lost their bid for the White House that year. With that said, I could not help but notice the ear-to-ear grins on many of the faces of our little group of New Yorkers when Frank's story ended with President Obama trouncing former Governor Romney. This is when the party got really interesting—one party-goer went so far as dancing on the grave of the defeated by claiming that the President's superior intellect was the key driver in his victory over the Governor. Thanks to Frank's better nature—and his ample enjoyment of Jim Beam and Ginger—he responded smartly: "Now, I'm not going to debate the breath or depth of each candidate's intellect, but what I *can* say of the Governor and his intelligence is that he's the only politician I've *ever* worked for who personally asked me to run a *regression analysis* on voter turnout!"

Perhaps it was the Wall Street pedigree of that group around the bar, but Frank's response did elicit more than a few laughs from the crowd. It was certainly no Lincoln-Douglas debate, but it left Frank the reigning champion of political storytelling that night.

So how does this all tie back to the importance for today's business executive to understand data and the tools for analyzing it? The point of Frank's story was that a candidate for President was leveraging regression analysis, a common tool for prediction and forecasting, to make more informed decisions on where to deploy his campaign's limited resources—namely dollars, people, and time.

If we look back as far as the eighteenth century, for over 50 years Thomas Jefferson made systematic daily weather observations of temperature and precipitation at his beloved Monticello estate. To support his extensive collection of information, Jefferson employed the services of dozens of family members, fellow politicians, scholars, and citizens—his band of so-called *watchmen*—and made significant contributions to the advancement of the science of meteorology in colonial America. The technologies that exist today for gathering and managing data are creating truly transformational business models—Amazon.com and Facebook to name just two—but the execution challenges are no less daunting than in Jefferson's time, and arguably no less central to the continued success of the enterprise. Today, with the sheer volume of data and the numerous channels for its collection, CIOs and the executive suite are being forced to put enterprise data strategy first in business technology decisions.

In this newsletter, we will highlight some of Monticello's recent engagements where we provide our clients with the enterprise data tools and analytics to add tangible value to their businesses. In addition, we will review some widely available Excel-based analytical functions, with the goal of providing the busy executive with a few tools for gaining powerful insights into what their organization's data indicates about the current state and future of their business. ■



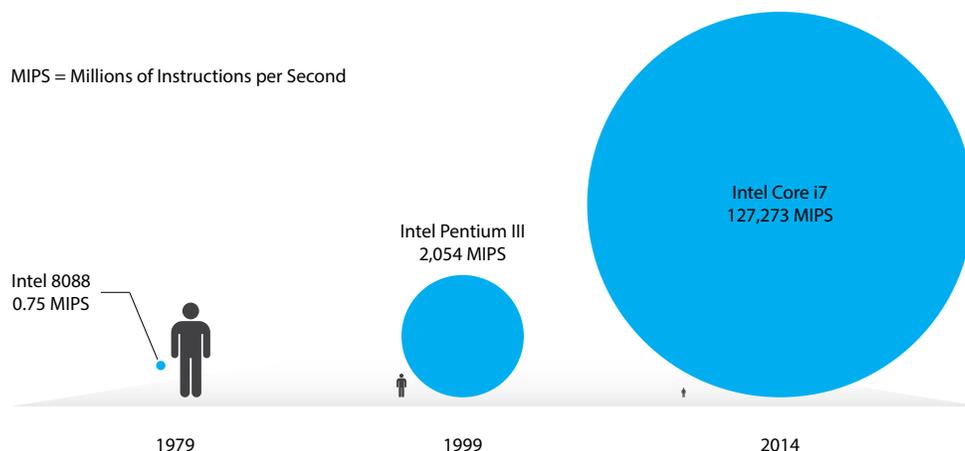
In the 18<sup>th</sup> and the early part of the 19<sup>th</sup> century, Jefferson's estate—Monticello—became his laboratory for many experiments involving the capture and organization of large data sets such as daily weather readings.



Will Morgan, the founder and president of Monticello Consulting Group, is an experienced business and implementation manager with over 18 years of consulting expertise across several industries.

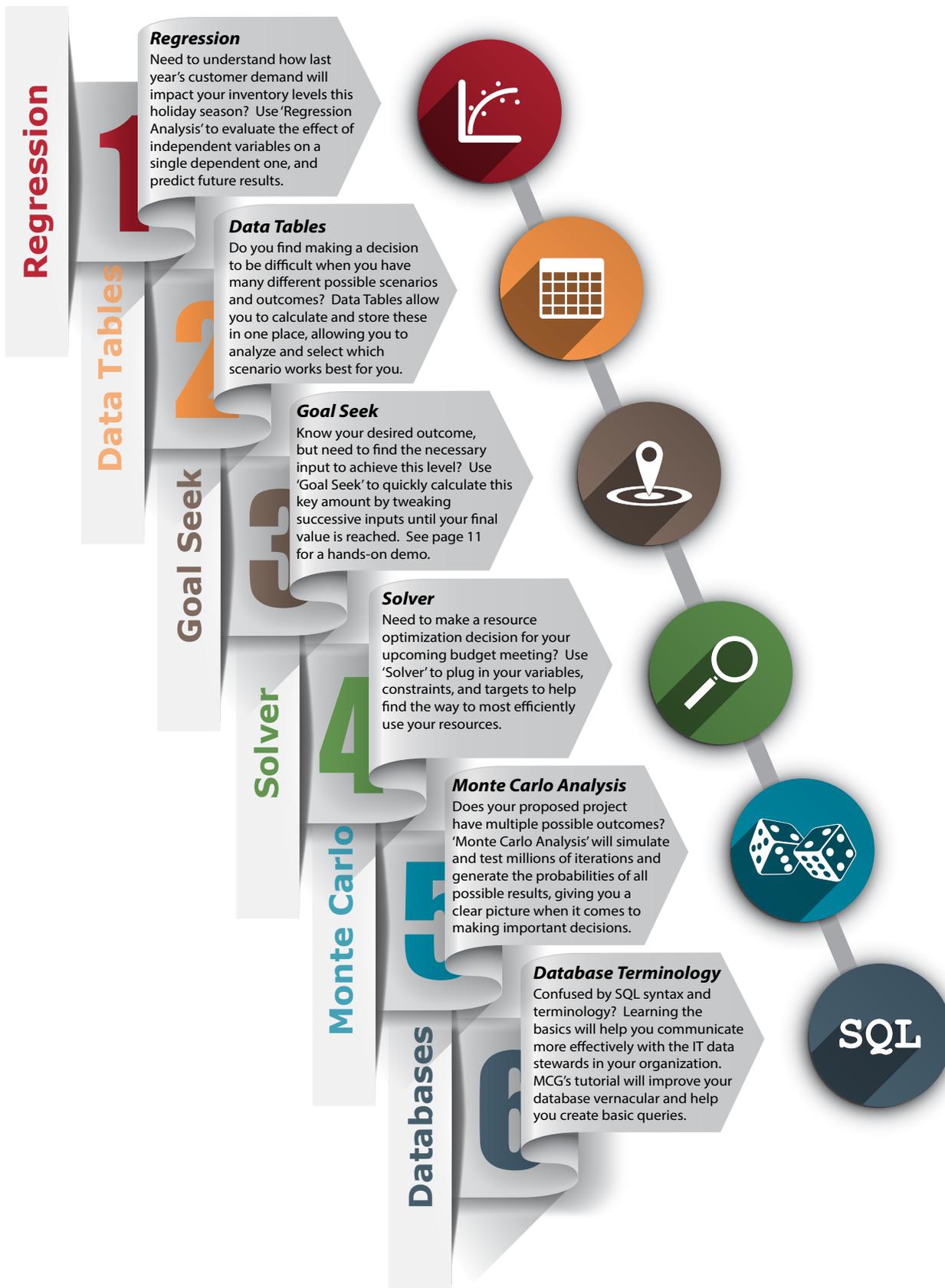
<sup>1</sup> Name changed to preserve anonymity of the consultant.

Intel's co-founder, Gordon Moore, first published his prophetic observation in 1965 that the performance of computers would double every 18 months. Since that time, the chip industry has used Moore's Law as its guidepost—driving the industry's roadmap for developing successive generations of microprocessors. The benefits of this technological progress cannot be understated, and have arguably given rise to one of the greatest economic and social drivers of change in human history. The figure below illustrates the exponential increase in speed and performance of Intel's widely installed CPUs—used in devices spanning personal computers, laptops, and more recently the tablet market—over time.



# Time-tested Data Analysis Tools, On the Go

In our consulting work for senior executives, we are often asked to provide summaries of data and to identify patterns to enable our clients' decision making. We frequently employ simple—yet effective—tools in order to accomplish our data mining and analysis goals. Below is a list of practical tools and short tutorials showcasing how powerful they can be for the business executive's day-to-day decision making. Most of these tools are available in MS Excel.





For over a century, horse racing aficionados have reveled in devising strategies to pick out the winning thoroughbred before the gates swung open. At their disposal has been what data analysts today refer to as *small data*— easily quantifiable figures such as race results, lap times, and track conditions. Recently, advancements in technology have allowed experts to track horses in real-time and to gather information previously thought unattainable. This *big data* may gradually help bridge the gap between likelihood and certainty.

# Selecting the Right Database for the Job

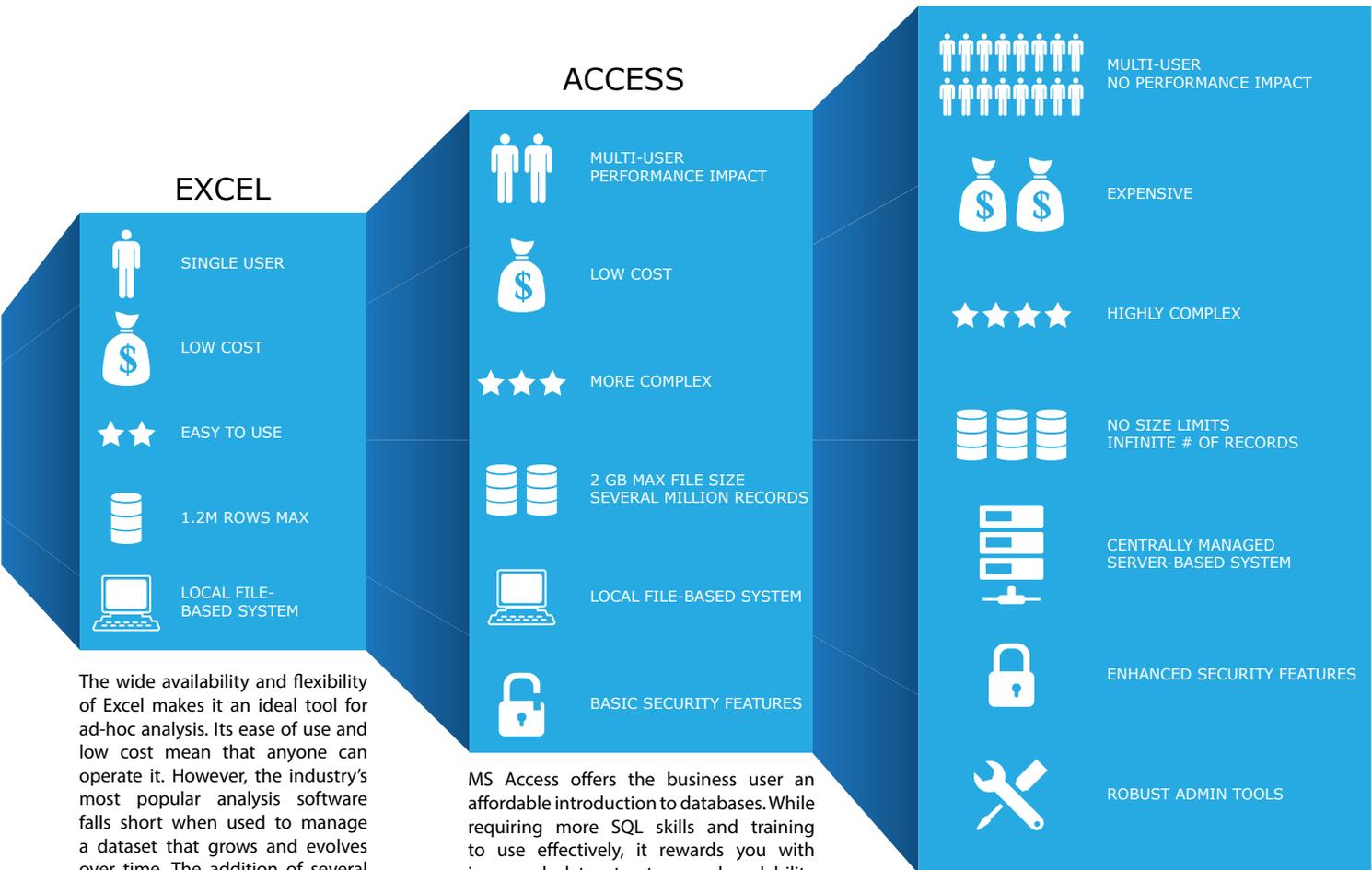
Most businesses work with data in various formats as part of their operations. Inherent to this are the challenges involved in deriving value from their data to benefit the organization. Many solutions exist to help firms cope with these challenges—among them a collection of tools known as *databases*. These prove to be especially useful in scenarios involving large volumes, multiple users, and concerns for security. Monticello Consulting Group can help you discover the solution that best fits your organization, as well as partner with you to make that solution a reality.



## Did You Know?

A relational database houses data in multiple tables joined by unique keys. The ability to create relations eliminates the need to duplicate data, making relational databases organized and efficient. Two popular examples of relational databases are SQL Server and ORACLE.

## Growing Needs



The wide availability and flexibility of Excel makes it an ideal tool for ad-hoc analysis. Its ease of use and low cost mean that anyone can operate it. However, the industry's most popular analysis software falls short when used to manage a dataset that grows and evolves over time. The addition of several users often results in many *similar* versions of a spreadsheet that become inconsistent, difficult to maintain, and prone to risky control issues.

MS Access offers the business user an affordable introduction to databases. While requiring more SQL skills and training to use effectively, it rewards you with improved data structure and scalability when compared to Excel. This, in turn, makes it much more reliable for managing a *living* dataset with less risk. And though Access ultimately is limited by its file-based design, its ubiquity in the workplace makes it an excellent stepping stone to more advanced database solutions.

One can think of SQL Server as being a more robust version of Access. But while the former sits on a local PC, SQL Server is housed in a centrally managed server farm. This allows it to be infinitely scalable to support hundreds, even thousands, of users and endless amounts of data. The power it offers organizations has made SQL Server, along with ORACLE, an industry standard solution for managing data. And while the logistical costs of building the underlying infrastructure for SQL Server can be sizeable, one may find that it is often already in place within major corporations.

## A Hybrid Solution

Oftentimes one application does not satisfy your data needs. You may find yourself gravitating toward several different solutions which best address your specific requirements. This is a perfectly logical approach, and can be made even better by linking your tools. The following are just two ways to create a hybrid software solution that optimizes your strengths while maintaining a fluid exchange of data:



### Excel + Access

The transition from Excel to Access is not always an easy one. And while the popular database solution may fare better with complex data manipulation, Excel remains the go-to option for intuitive graphing and ad-hoc analysis. Linking the two (via an ODBC connection) can provide you with the benefits of both tools in one harmonious package.



### Access + SQL Server

A pairing of Access and SQL Server is a widely used technique for building tactical tools. While not as flashy as a purpose-built software interface, the forms in Access provide users with an adequate medium to run the tool. And thanks to its ubiquity on business PCs, the Access front-end can be deployed to multiple users without much fuss. The back-end, meanwhile, runs on a central SQL Server farm—performing intensive calculations server-side and allowing modifications to be done in one centralized location.

### Did You Know?

Structured Query Language (SQL) is used to manipulate data in relational databases. While minor syntax variations exist between popular systems, SQL is a largely universal language.

## A Note on Big Data

Our clients often ask us for our thoughts on Big Data, or ask us when Big Data is going to impact their organizations. The reality is, if you are working for a Fortune 500 company today, your organization is likely being impacted by the massive proliferation of data already. With that said, it is difficult to provide a rule of thumb qualifying an enterprise's data needs as falling into the realm of 'Big Data,' as what is considered 'big' today will surely be surpassed in just a few short years. However, the following determinants play important roles:

### Data Collection Sources

With the coming of the Internet of Things (IoT), supported by the proliferation of inexpensive sensors streaming data to the Internet, along with many existing channels for gathering data such as social media and large transactional databases, the proliferation of new data sources will continue to accelerate and create the need for enterprises to employ Big Data technologies to capitalize on this information.

### Data Extraction & Organization

Extracting actionable intelligence from disparate data sources, as well as storing this data, requires technologies such as Hadoop, using distributed and scalable processing of large data sets across many different servers.

Since the early 1980s, Relational Database Management Systems (RDMS) had dominated the data storage market. However, with the arrival of the volume and velocity demands of Big Data, the need to leverage massively parallel hardware and software architectures to store and process data has challenged RDMS players such as Oracle and Microsoft SQL Server to keep up with the tide of available data. As a result, Big Data technologies such as Hadoop, which leverage distributed storage and large-scale processing, are quickly being adopted within the corporate IT landscape.

One area Big Data is being deployed in financial markets is by regulators such as the CFTC, SEC, and FINRA. Regulators are deploying sophisticated surveillance technology to mine the terabytes of trade data they are receiving from market participants as part of the complex regulatory reform requirements stemming from the 2008 financial crisis. With these tools, regulators have the ability to identify patterns of troublesome trading and drill down on trading activity for further investigation. This is a new capability in the hands of regulators that will aid in the smooth operation of financial markets, and was not available only a few short years ago—prior to the implementation of Big Data technologies. ■



*“There was five exabytes of information created between the dawn of civilization through 2003, but that much information is now created every two days.”*

- Eric Schmidt, Google (2010)

# Case 1:

## A Monticello Consulting Case Study for Improving Data Quality, a Regulatory Reporting Mandate

The Dodd-Frank Act Title VII as implemented by the CFTC in late 2012 requires regulatory reporting by all registered swap dealers (SDs) and major swap participants (MSPs) trading in eligible product types. These eligible trades are reported to a designated global trade repository (GTR) maintained by one of several industry bodies such as the Depository Trust & Clearing Corporation (DTCC). Because of the scope and difficulty of reporting both new and historical trading activity—often referred to in the industry as back-loaded trades—the information sent to a GTR may fail at meeting quality benchmarks established by the regulators along with a firm's internal compliance stakeholders. Compounding the challenge of reporting can be the ambiguity in the technical requirements published by a GTR for successful acknowledgement (ACK) of real-time reported eligible trades.

For over three years, Monticello Consulting has been actively engaged in the industry, partnering with our swap dealer clients to remediate data quality issues and to improve the ACK rates for the trades associated with real-time reporting. Challenges posed in these projects have included:

- Very large sets of data containing millions of trades and message types, combined with a dearth of robust tools for managing this data.
- Ambiguity of the GTR's requirements for sending a correctly formatted message to receive an ACK.
- Multiple front office systems supplying raw trade data in varying formats.

- Overworked IT staff being stretched to support development for non-U.S. jurisdictions which went into effect around the same time periods.

For one particular engagement, a swap dealer client was experiencing several data quality issues resulting from the original back-loading of the dealer's trades and the ongoing reporting of new trades. Issues included:

- Reporting Logic: Incorrect processing logic in the front office systems and the client's reporting engine.
- Data Completeness: Trade messages were sent to the GTR missing values in multiple fields that were required for receiving an ACK.
- Data Quality: Some trade messages were sent to the GTR with incorrect values in their fields.
- Inaccurate Static Data: For specific trade populations, static data such as a client's legal entity identifier (LEI) was incorrect or missing.
- Orphan Positions: Subsequent to the initial transmission of trades to the GTR, a subset of these trades were exited (a type of CANCEL event) improperly. Because of this *failed exit* scenario, a number of *orphaned positions* remained in an open and active state in the GTR.
- Reconciliation Issues: Reconciling internal trade data versus client data.
- Back-loading Challenges: Identifying and managing populations of live trades versus historic trades for back-loading to the GTR.

The Monticello team solved for these data quality issues by designing a suite of custom-built tactical tools, including:

- An MS-Access tactical tool to analyze the different data fields and determine target populations for cleanup.
- A custom SQL Server database for initial trade analysis of large populations of trades, feeding down to several smaller databases for trade back-load queueing.
- Several bespoke utilities written in C# for data analysis, population culling, XML parsing, and FTP processing.
- A transformation tool written in C# for mapping raw CSV trade files to create properly formatted input files required for back-loading.

The benefits of Monticello's work in this area to the client's regulatory reporting platform have included:

- Pinpointing incorrect logic for rules processing within the front office systems and the client's regulatory reporting engine.
- Cleaning up multiple data fields for trades in both the front office systems and our client's reporting engine.
- Removing a substantial population of orphaned positions.
- Cleaning up static data issues in the reported trade population.

The direct impact of our work in this area has aided our clients in significantly improving their data quality ranking in the industry vis-à-vis other large swap dealers on the Street. ■



Robert Adams is a senior manager with significant expertise in data analysis for complex change programs. Bob has advised clients on improving data quality for CFTC and JFSA regulatory reporting.



Following the financial crisis of 2008, demand for reform on Wall St. culminated in the passing of the Dodd-Frank Act in 2010. As a result, large investment banks are now under increased scrutiny by governing bodies and must comply with various new regulatory requirements.

## Case 2:

### Brokerage Payables Remediation Engagement

A top investment banking client recently called on Monticello to help with a process fit for automation. The engagement called for building a suite of tactical tools to address a particularly thorny reconciliation dilemma. Acting as the clearing broker in a futures and options give-up scenario, the bank would take on the responsibility of managing fee disbursements between involved parties in a trade (Exhibit A). Over the past seven years, our client would repeat this process for millions of transactions, utilizing GMI—an industry standard system for futures and options back-office processing.

For each trade, the bank would receive an invoice from the executing broker—one who had given up a trade to our client. This would be reconciled to the theoretical fee amount in GMI before payment was sent. When the bank couldn't agree with the executing broker on the proper fee amount for the transaction, breaks were created in the process and brokers did not receive payments. This dilemma spawned from several concurrent factors.

#### Factor 1: Voice vs. DMA

One of the primary root causes was a fault in the GMI platform itself. This legacy back-office processing system used by many top broker dealers—including our client—lacks a rudimentary indicator to distinguish between trades booked via *voice*<sup>1</sup> and those booked through *direct market access (DMA)*<sup>2</sup>. As a result, there were many instances where the clearing broker's staff were unable to identify voice trades from DMA trades. Because voice and DMA trades carry different fees, this led to frequent disagreements in the fee invoicing process between the clearing and executing brokers.

#### Factor 2: An Industry Problem

Not all broker-dealers use GMI to track their trading activity. In fact, there are three platforms that have dominant

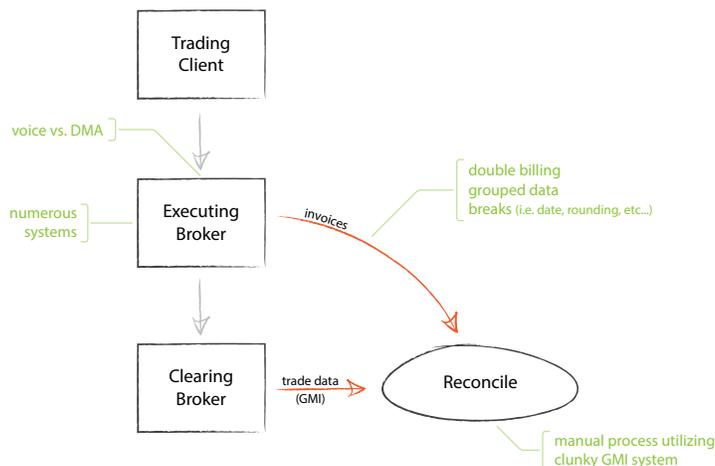
- <sup>1</sup> *Voice* trades are phoned in to the executing broker and require trading assistance by a broker. Trades booked via *voice* garner a significantly higher executing fee than those booked thru *DMA*.
- <sup>2</sup> *Direct market access (DMA)* allows trading parties to use the executing broker's internal trading platform without requiring assistance from a broker. For this reason, trades booked via *DMA* usually carry a smaller executing fee.

#### Exhibit A: F&O Trade Flow and Fee Structure



Ayrat Safine is a Business Analyst and Project Manager with particular skills in quantitative investigation. Ayrat specializes in devising technology solutions for tackling his clients' most critical data challenges.

#### Exhibit B: Several Factors at Play



market share in the industry for back-office futures and options processing.<sup>3</sup> Not surprisingly, these systems lack consistent data standards. The various nuances at play when comparing these trade repositories require complex mapping routines to translate the data from one system prior to matching it to trades in their native platform (GMI in the case of our client).

#### Factor 3: Breaks

Upon reconciling invoices with data in GMI, analysts often encountered

- <sup>3</sup> The most common F&O platforms include GMI, Rolfe & Nolan, and RANsys.

a range of discrepancies. Duplicate invoices were frequently received from executing brokers. Time and again, fees were found grouped together by various categories, and a vast majority of fees were prone to breaks for reasons ranging from precision to timing issues.

#### Factor 4: A Growing Problem

Finally, this reconciliation process for the brokerage department staff at our client was a painstakingly manual one. Every day, the staff had to choose which breaks to tackle but had limited ability to prioritize based on hard data. Without proper data analytics,

#### Did You Know?

A **unique identifier (UID)** is a numeric or alphanumeric string assigned to a single entity or trade within a system. Leveraging unique identifiers allows us to reconcile trades from multiple sources systematically. Consequently, in their absence, a more cumbersome process is required.

#### Did You Know?

ETL (Extract, Transform, Load) is a process by which data is moved from one source to another in a controlled fashion. An ETL tool, such as SSIS or Informatica, is essential for data warehouses.

this often meant reconciling either the oldest items first or prioritizing those brokers who were demanding to be paid on an immediate basis. As one can imagine, this led to an inefficient state where quick wins and easy matches went unexploited—lost in a sea of ever-growing unpaid invoices.

**The Monticello Solution**

To address this problem, the Monticello team implemented a tactical reconciliation tool capable of matching aged transactions en masse. The technology underlying this tool included a MS SQL Server database capable of processing millions of records loaded daily, along with a user-friendly front-end interface built in MS Access. The Microsoft tech stack was chosen for this solution because it was supported within our client’s IT organization, and also due to the omnipresent nature of MS Access on

the business user’s desktop—facilitating quick and easy deployment to new users.

While the use of a unique identifier (UID) is the preferred method of running a reconciliation of this magnitude, no tracer fields could be utilized for the data in question. In the absence of a UID, Monticello’s team of analysts performed a judicious data scrubbing exercise on the incoming broker invoices. Subsequent to the scrubbing, the transaction data was run through a complex hierarchical matching sequence to automatically pair invoiced trades with their respective counterparts in the client’s in-house GMI trade repository.

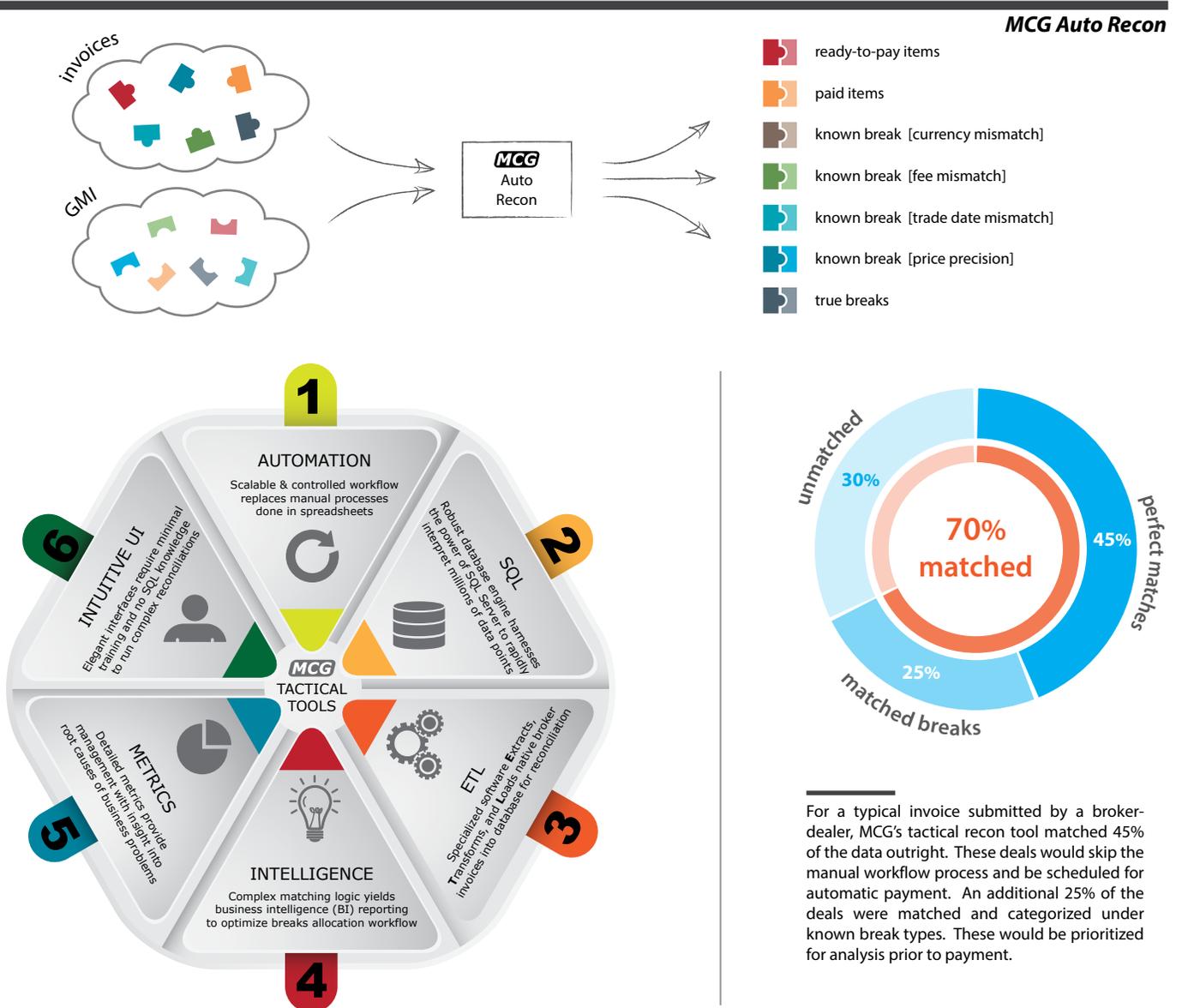
In less time than it would take an analyst to manually locate and reconcile a broken fee using the GMI interface, the MCG Auto Recon tool processes

millions of transactions and produces an output set of business intelligence (BI) reports to direct analysts toward problem areas in need of attention. In addition, the tool suite segregates and prepares perfectly matched brokerage fees for prompt payment, thus further alleviating the analysts’ workload and satisfying the executing brokers waiting for their funds.

This fully automated solution allowed our client to make significant strides toward reducing aged payable balances owed to their peer broker-dealers, while also minimizing the time and resources required to reconcile F&O brokerage fees. Not surprisingly, the MCG tactical recon tool has become an integral component of our client’s back-office futures and options processing, and continues to produce tangible benefits long after our engagement has come to a close. ■

**Did You Know?**

Data scrubbing is the process of cleansing data by fixing errors, removing duplicates, and correcting formatting. A critical step in an analysis effort, data scrubbing can either be done manually using tools like Excel or systematically using special programs and algorithms like SSIS.



For a typical invoice submitted by a broker-dealer, MCG’s tactical recon tool matched 45% of the data outright. These deals would skip the manual workflow process and be scheduled for automatic payment. An additional 25% of the deals were matched and categorized under known break types. These would be prioritized for analysis prior to payment.

# Time-tested Data Analysis Tools, On the Go

## Goal Seek

### What is it?

Goal Seek allows the user to set the value in a cell containing a formula by changing the value in another.

### When to use it?

Use Goal Seek to find the result you want by changing a single input variable. Say, for example, you believe the S&P is currently overpriced, knowing that since the 1870s the index has historically traded with a price-to-earnings ratio (P/E) of 15 and it is currently trading at 20.58. Before dipping into the market you would like to see the index's P/E ratio return to its historical average. Here's how you can use Goal Seek to identify the price the S&P index would need to drop to in order to return to a P/E of 15: Recent S&P Price = 2,105.33. S&P trailing 12 month (TTM) earnings = 102.31 (for Q4 2014).

### How it works

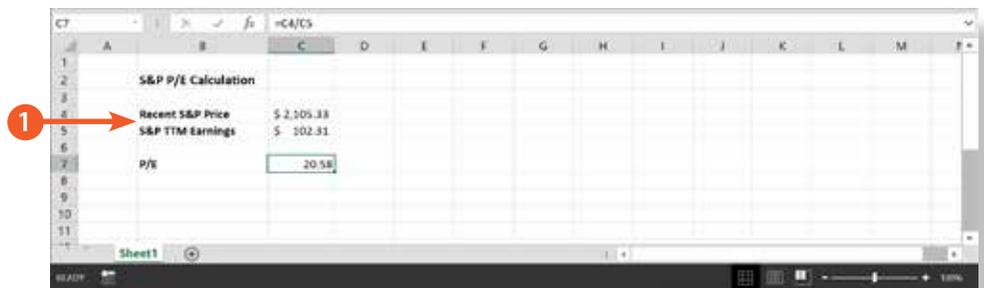
Goal Seek uses the techniques of linear algebra for solving equations.

### What's great about it?

Goal Seek saves you time, eliminating the seemingly endless toggling of values in our spreadsheets to get the formulas to achieve our desired results. Goal Seek is great for tweaking successive inputs to our models until the desired result is achieved. In a later tutorial, we will review Solver/Scenario Manager, an Excel-based tool that allows us to change more than one input variable concurrently, making it possible to achieve optimal results across all variables.

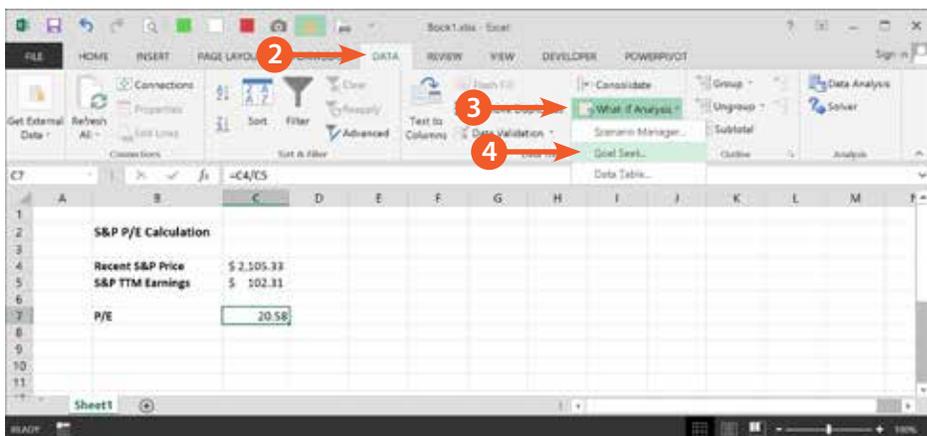
## Using Goal Seek

- 1 Set up your worksheet model.



- 2 Click the **Data** tab.
- 3 Click **What-If Analysis**
- 4 Click **Goal Seek**.

The Goal Seek dialog box appears.



- 5 Click inside the **Set cell** box.
- 6 Click the cell that contains the formula you want Goal Seek to work with.
- 7 Use the **To value** text box to type the value that you want Goal Seek to find.
- 8 Click inside the **By changing cell** box.
- 9 Click the cell that you want Goal Seek to modify.

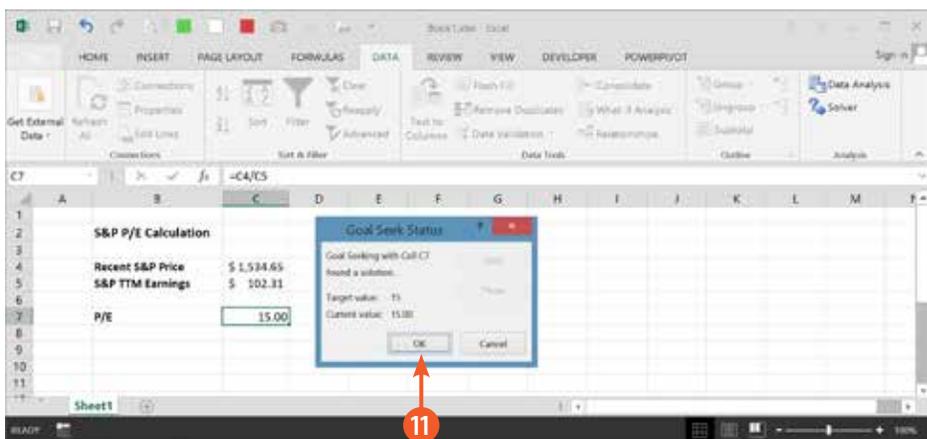
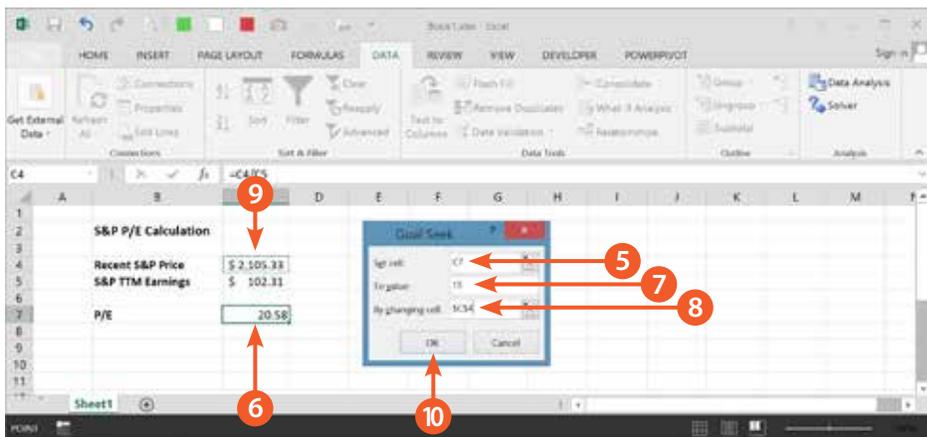
10 Click **OK**.

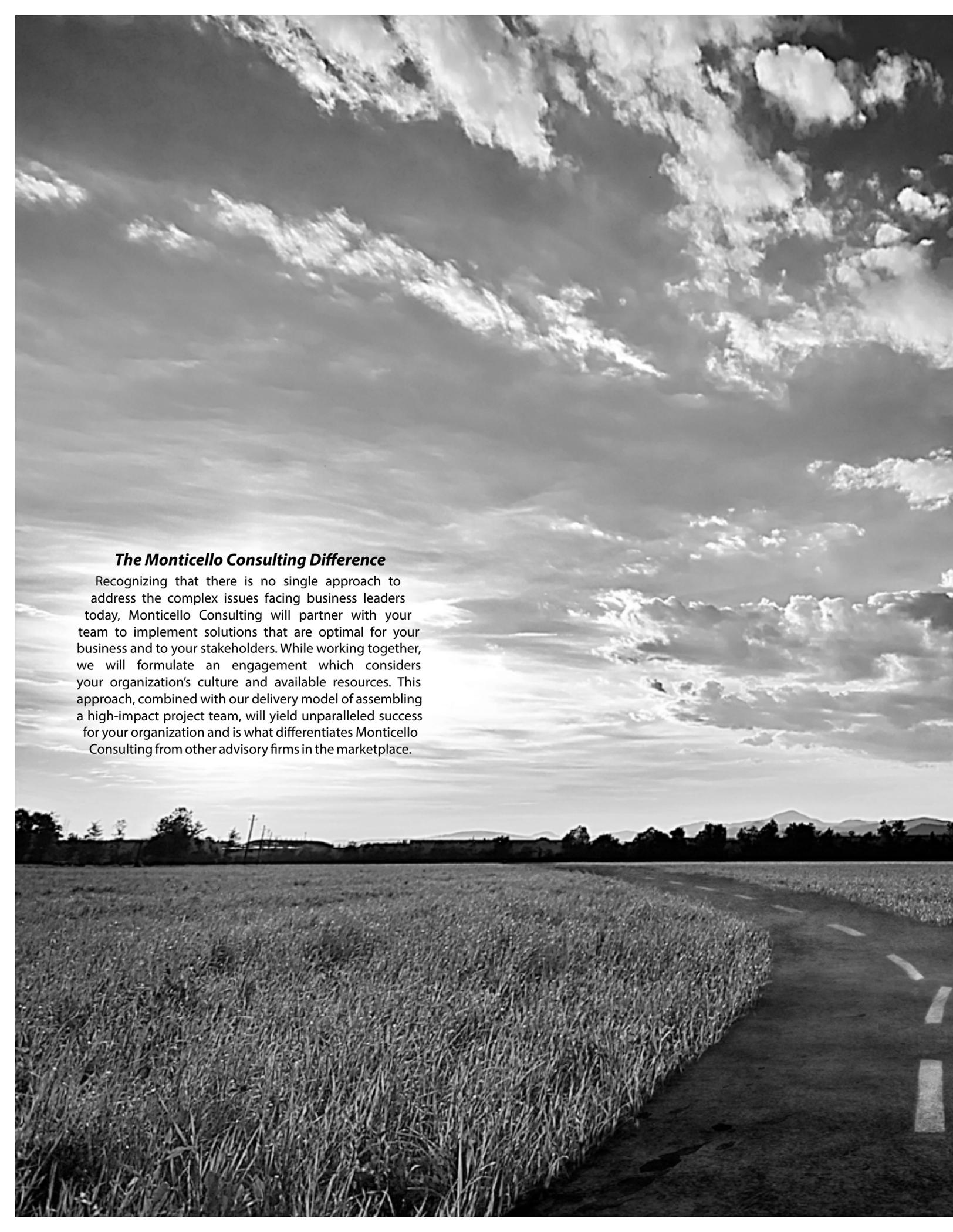
● Goal Seek adjusts the changing cell value until it reaches a solution.

● The formula now shows the value you entered in Step 7.

11 Click **OK**.

Your conclusion? You do not want to start investing in the S&P 500 until the market drops to 1,534.65. Stay on the sidelines for now!





***The Monticello Consulting Difference***

Recognizing that there is no single approach to address the complex issues facing business leaders today, Monticello Consulting will partner with your team to implement solutions that are optimal for your business and to your stakeholders. While working together, we will formulate an engagement which considers your organization's culture and available resources. This approach, combined with our delivery model of assembling a high-impact project team, will yield unparalleled success for your organization and is what differentiates Monticello Consulting from other advisory firms in the marketplace.





*Monticello*  
CONSULTING GROUP

**Virginia Headquarters**

P.O. Box 28037  
Richmond, VA 23228

**New York Office**

35 W 38th Street  
2nd Floor, Suite 2S  
New York, NY 10018

**Contact**

[info@monticellocg.com](mailto:info@monticellocg.com)  
+1 212 278 8012  
[www.monticellocg.com](http://www.monticellocg.com)

**Monticello Consulting Group, Inc.**

