Disruptive Analytics

Charting Your Strategy for Next-Generation Business Analytics

Thomas W. Dinsmore

Apress®

DISRUPTIVE ANALYTICS

CHARTING YOUR STRATEGY FOR NEXT-GENERATION BUSINESS ANALYTICS

Thomas W. Dinsmore

Apress®

Disruptive Analytics: Charting Your Strategy for Next-Generation Business Analytics

Thomas W. Dinsmore

ISBN-13 (pbk): 978-1-4842-1312-4 ISBN-13 (electronic): 978-1-4842-1311-7
DOI 10.1007/978-1-4842-1311-7

Library of Congress Control Number: 2016950565

Managing Director: Welmoed Spahr
Acquisitions Editor: Robert Hutchinson
Developmental Editor: Matt Moodie
Technical Reviewer: Robert A. Muenchen
Editorial Board: Steve Anglin, Pramila Balen, Laura Berendson, Aaron Black, Louise Corrigan, Jonathan Gennick, Robert Hutchinson, Celestin Suresh John, Nikhil Karkal, James Markham, Susan McDermott, Matthew Moodie, Natalie Pao, Gwenan Spearing
Coordinating Editor: Rita Fernando
Copy Editor: Kezia Endsley
Compositor: SPi Global
Indexer: SPi Global
Cover Designer: Isaac Ruiz Soler

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a Delaware corporation.

For information on translations, please e-mail rights@apress.com, or visit www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales–eBook Licensing web page at www.apress.com/bulk-sales.

Any source code or other supplementary materials referenced by the author in this text is available to readers at www.apress.com. For detailed information about how to locate your book's source code, go to www.apress.com/source-code/.

Printed on acid-free paper

Apress Business: The Unbiased Source of Business Information

Apress business books provide essential information and practical advice, each written for practitioners by recognized experts. Busy managers and professionals in all areas of the business world—and at all levels of technical sophistication—look to our books for the actionable ideas and tools they need to solve problems, update and enhance their professional skills, make their work lives easier, and capitalize on opportunity.

Whatever the topic on the business spectrum—entrepreneurship, finance, sales, marketing, management, regulation, information technology, among others—Apress has been praised for providing the objective information and unbiased advice you need to excel in your daily work life. Our authors have no axes to grind; they understand they have one job only—to deliver up-to-date, accurate information simply, concisely, and with deep insight that addresses the real needs of our readers.

It is increasingly hard to find information—whether in the news media, on the Internet, and now all too often in books—that is even-handed and has your best interests at heart. We therefore hope that you enjoy this book, which has been carefully crafted to meet our standards of quality and unbiased coverage.

We are always interested in your feedback or ideas for new titles. Perhaps you'd even like to write a book yourself. Whatever the case, reach out to us at `editorial@apress.com` and an editor will respond swiftly. Incidentally, at the back of this book, you will find a list of useful related titles. Please visit us at `www.apress.com` to sign up for newsletters and discounts on future purchases.

The Apress Business Team

To Ann

Contents

About the Author

Thomas W. Dinsmore is an independent consultant and author who specializes in advanced analytics and machine learning.

In his consulting career, Mr. Dinsmore has served in expert roles for The Boston Consulting Group, PricewaterhouseCoopers, Oliver Wyman, IBM Big Data Solutions, and the SAS Institute. He has also served as Director of Product Management for Revolution Analytics (now a division of Microsoft.)

Mr. Dinsmore has more than 30 years of experience in advanced analytics. He has led or contributed to solutions for AT&T, Banco Santander, Citibank, Dell, J. C. Penney, Monsanto, Morgan Stanley, Office Depot, Sony, Staples, United Health Group, UBS, Vodafone, and many other clients in the United States, Puerto Rico, Canada, Mexico, Venezuela, Brazil, Chile, the United Kingdom, Belgium, Spain, Italy, Turkey, Israel, Malaysia, and Singapore.

Mr. Dinsmore has working experience with most of the leading tools for advanced analytics. He is the co-author of *Modern Analytics Methodologies* (FT Press, 2014) and *Advanced Analytics Methodologies* (FT Press, 2014) and publishes The Big Analytics Blog. He earned an MBA from the Wharton School, The University of Pennsylvania, and a BA from Boston University.

About the Technical Reviewer

Robert A. Muenchen is the author of *R for SAS and SPSS Users* and, with Joseph M. Hilbe, *R for Stata Users*. He is also the creator of r4stats.com, a popular web site devoted to analyzing trends in data science software and helping people learn the R language. Bob is an ASA Accredited Professional Statistician™ with 30 years of experience and is currently the manager of OIT Research Computing Support (formerly the Statistical Consulting Center) at the University of Tennessee. He has taught workshops on research computing topics for more than 500 organizations and has offered training in partnership with DataCamp.com, Revolution Analytics, RStudio, New Horizons Computer Learning Centers, and Xerox Learning Services. Bob has written or co-authored over 70 articles published in scientific journals and conference proceedings, and has provided guidance on more than 1,000 graduate theses and dissertations.

Bob has served on the advisory boards of SAS Institute, SPSS Inc., Intuitics OOD, StatAce OOD, the Statistical Graphics Corporation, and *PC Week Magazine*. His suggested improvements have been incorporated into SAS, SPSS, JMP, STATGRAPHICS, and several R packages. His research interests include statistical computing, data graphics and visualization, text analytics, and data mining.

Acknowledgments

Many thanks to Bob Muenchen of the University of Tennessee; Bob graciously agreed to serve as the technical reviewer of this book and spent many hours reading and commenting on chapter drafts.

Also thanks to Oliver Vagner, Senior Director, Data Analytics at TGI Fridays, and to Professor Dr. Diego Kuonen of the University of Geneva and Statoo Consulting, each of whom provided valuable suggestions and guidance for the book's recommendations to managers.

Thanks also to many other people in the industry who have provided insight into the ideas and topics covered in this book, including Jeremy Achin, Data Scientist and CEO, DataRobot; Bruno Aziza, Chief Marketing Officer, AtScale; Charlie Berger, Oracle; Michael Berthold, President, KNIME.com AG; Arno Candel, Chief Architect, H2O.ai; David Champagne, Principal Software Engineering Manager, Microsoft; Michael Chu, Project Leader, The Boston Consulting Group; Davide Consiglio, Principal, The Boston Consulting Group; Boxuan Cui, Data Scientist, Smarter Travel; David Erdreich, Knowledge Expert, The Boston Consulting Group; Lee Edlefson, Principal Software Engineer, Microsoft; Ali Ghodsi, CEO, Databricks; Mario Inchiosa, Principal Software Engineer, Microsoft; Bill Jacobs, Microsoft; Paul Kent, VP, Big Data, SAS; Josh Klahr, Vice President of Product Management, AtScale; Bill Lehmann, Investor, Bain Capital Ventures; Matthew Madden, Director of Product Marketing, Alteryx; Xiangrui Meng, Software Engineer, Databricks; Ingo Miersewa, Founder and CTO, RapidMiner; Derek Norton, Senior Data Scientist, Microsoft; Thomas Ott, Marketing Data Scientist, RapidMiner; Sean Owen, Director of Data Science, Cloudera; Krishnan Parasuraman, VP Sales, Splice Machine; Zoltan Prekopcsak, VP Big Data, RapidMiner; Peter Prettenhofer, Data Scientist, DataRobot; Dan Putler, Chief Scientist, Alteryx; David Rich, Strategic Advisor and Executive Coach; Razi Razuddin, VP, Strategic Business Development, DataRobot; Vincent Saulys, Senior Director, Advanced Surveillance Development, Financial Industry Regulatory Authority; Joseph Sirosh, Corporate Vice President, Data Group, Microsoft; Dan Steinberg, President, Salford Systems; Ben Strauss, Alteryx and Tableau Evangelist, The Boston Consulting Group; Gregory Todd, Principal, PricewaterhouseCoopers; Deenar Torasker, Founder, ThinkReactive; Tom Ventura, Global IT Director, The Boston Consulting Group; Reynold Xin, Chief Architect, Databricks; David Wang, Director of Product Marketing, Databricks; and Bill Zanine, Thought Leader, Netezza Analytic Advisory Services, IBM. If I have inadvertently omitted anyone from this list, I apologize in advance.

To my clients, who I cannot name, who have made it possible for me to devote a significant amount of time to this book.

Finally, many thanks to the people at Apress who agreed to publish this book, and to members of the editorial and production staff who contributed to the final product.

Introduction

Disruption: In business, a radical change in an industry or business strategy, especially involving the introduction of a new product or service that creates a new market.

From its birth in 1979, Teradata led the field in data warehousing. The company built a reputation for technical acumen, serving customers like Walmart and Citibank; analysts and implementers alike rated the company's massively parallel databases "best in class." After a 2007 spinoff from NCR, the company grew by double digits.

On August 6, 2012, Teradata released its earnings report for the second quarter. Results excelled; revenue was up 18% and earnings per share (EPS) up 28%. Teradata stock traded at $80, five times its value four years earlier.

"We are increasing our guidance for constant currency revenue growth and EPS for 2012," wrote CEO Mike Koehler.

In retrospect, that moment was Teradata's peak. Over the next three and a half years, the company lost 75% of its market value, as it repeatedly missed revenue and earnings targets. In 2015, Koehler announced a restructuring and sale of company assets; several top executives departed. Finally, after a brutal first quarter earnings report, Koehler himself stepped down in May 2016.

Management blamed many factors for the sluggish sales: long sales cycles, a sluggish economy, and unfavorable currency movement. But worldwide spending on business analytics *increased* during this period and some vendors reported double-digit revenue growth.

Blaming Teradata's struggles on poor leadership would be easy. But the company's growth problems in the last few years are not unique: in the same period, Oracle and IBM suffered declining revenue; Microsoft and SAP failed to grow consistently, disappointing investors; and SAS had to walk back embarrassing projections of double-digit growth, recording low single-digit gains.

In short, while businesses continue to invest in analytics, they aren't buying what the industry leaders are selling.

Meanwhile, a steady stream of innovation creates new value networks in the business analytics marketplace:

Open Source Analytics. With substantial gains in the last several years, open source software makes deep inroads in the analytics community. Surveys show that working data scientists prefer open source R and Python over commercial software. Technology leaders like Oracle, IBM, and Microsoft rush to get on the open source bandwagon.

Hadoop and its Ecosystem. As Hadoop matures, it competes successfully with data warehouse appliances, even displacing them. Technology consultant Gartner estimates that 42% of all enterprises now use Hadoop. A few years ago, data warehousing vendors laughed at Hadoop; they aren't laughing today.

In-Memory Analytics. As the cost of memory declines, fast and scalable high-performance analytics are within reach for any organization. Adoption of open source Apache Spark, an open source project for scalable in-memory computing, increases exponentially. With more than a thousand contributors, Spark is the most active open source project in Big Data.

Streaming Analytics. Organizations face a growing volume of data in motion, driven in part by the Internet of Things (IoT). Today, there are no less than six open source projects for streaming analytics in the Apache ecosystem. In-memory databases position themselves as streaming engines for hybrid transactional/analytical processing (HTAP).

Analytics in the Cloud. When Amazon Web Services introduced its Redshift columnar database in 2012, it lacked many of the features available in competing data warehouses. For many businesses, however, Amazon offered a compelling value proposition: "good enough" functionality, at a fraction of the cost of a Teradata warehouse. The leading cloud services all report double-digit revenue growth; Gartner estimates that 44% of all businesses use the cloud.

Deep Learning. Cheap high-performance computing power makes Deep Learning practical. NVIDIA releases its DGX-1 chip for Deep Learning, with the power of 250 servers; Cray announces its Urika-GX appliance with up to 1,728 cores and 35 terabytes of solid-state memory. Meanwhile, Google releases its TensorFlow framework to open source and declares that it uses Deep Learning in "hundreds" of applications.

Self-Service Analytics. With an easy-to-learn user interface and robust connectors to data sources, Tableau turns the business intelligence software industry upside down and grows its revenues tenfold while established Business Intelligence vendors struggle to adapt. Other startups position themselves to bring the self-service model to other disciplines, such as OLAP and machine learning.

This is not another book that hypes Big Data. Petabytes of data are worthless unless they answer a business question; the tsunami of data produced by the digital economy is simply a fact of life that managers must address. Whether you manage a multinational or drive a truck, your business produces more data than ever; you will either use it or discard it, but one way or the other, you must make an informed decision.

In a disrupted business analytics market, managers must focus ruthlessly on needs for insight, then build systems and processes that satisfy those needs. Understanding the innovations described in these chapters is a step toward that end, but the focus must remain on the demand for insight and the value chain that delivers it.

Innovations do not spring fully formed from the mind of an inventor; they are the end result of a long process of tinkering. Many of the most significant innovations we describe in this book are more than 50 years old; they emerge today for various reasons, such as the long-run decline of computing costs. We present a historical perspective at several points in this book so the reader can distinguish between that which is really new and that which is simply repackaged and rebranded.

In the middle chapters of this book, we present a survey of a key innovation in business analytics. These chapters include detailed information about available software products and open source projects. In general, we do not cover offerings from industry leaders, under the premise that these companies have ample marketing budgets to build awareness of their products.

We close the book with a handbook for managers: specific strategies to profit from disruptive innovation. Some of these strategies may seem radical; if this disturbs you, put this book down—it's not for you. But if you are ready to embrace disruptive innovation, and profit by it, read on.

CHAPTER

1

Fundamentals

Disruption in the Analytics Value Chain

The analytics business is booming. Technology consultant IDC estimates[1] total spending for analytic services, software, and hardware exceeded $120 billion in 2015; through 2019, IDC forecasts that spending will increase to $187 billion, an 11% compound annual growth rate[2].

So, if analytics is such a hot field, why are the industry leaders struggling?

- Oracle's cloud revenue growth[3] fails to offset declining software and hardware sales[4].
- SAP's cloud revenue grows, but total software revenue is flat[5].
- IBM reports[6] 16 straight quarters of declining revenue. Mass layoffs ensue[7].

[1] https://www.idc.com/getdoc.jsp?containerId=IDC_P33195
[2] http://www.cio.com/article/3074238/analytics/big-data-and-analytics-spending-to-hit-187-billion.html
[3] http://www.forbes.com/sites/laurengensler/2016/03/15/oracle-third-quarter-earnings/#286720039d5d
[4] http://investor.oracle.com/financial-news/financial-news-details/2016/Oracle-Reports-GAAP-EPS-of-050-Non-GAAP-EPS-of-064-Without-the-Effect-of-US-Dollar-Strengthening-Both-Would-Have-Been-4-Cents-Higher/default.aspx
[5] http://go.sap.com/docs/download/investors/2016/sap-2016-q1-statement.pdf
[6] https://www-03.ibm.com/press/us/en/pressrelease/49554.wss
[7] http://fortune.com/2016/05/20/ibm-layoff-employees-may/

T. W. Dinsmore, *Disruptive Analytics*, DOI 10.1007/978-1-4842-1311-7_1

- Microsoft underperforms[8] analysts' expectations despite 120% growth in Azure cloud revenue.
- Predictive analytics leader SAS reports[9] five years of low single-digit revenue growth; EVP departs[10].
- Data warehousing leader Teradata shuffles its leadership team after four years of declining product revenue[11].

Product quality is not the problem. Each company offers products that industry analysts rate highly:

- Forrester and Gartner both[12] recognize[13] IBM, SAS, SAP, and Oracle as leaders in data quality tools.
- Gartner rates[14] Oracle, SAP, IBM, Microsoft, and Teradata as leaders in data warehousing.
- Forrester rates[15] Microsoft, SAP, SAS, and Oracle as leaders in agile business intelligence.
- Gartner recognizes SAS and IBM as leaders in Advanced Analytics[16].

The answer, in a word, is *disruption*[17]. Powerful forces are rearranging the industry:

- Digital transformation of the economy and rapidly declining storage costs produce a data tsunami.
- The number of data sources is exploding. Data sources are everywhere: on-premises, in the cloud, in consumers' pockets, in vehicles, in RFID chips, and so forth.

---

[8] http://www.reuters.com/article/us-microsoft-results-idUSKCN0XI2NG
[9] http://www.sas.com/en_us/company-information.html#stats
[10] http://www.newsobserver.com/news/business/article48668040.html
[11] http://www.mydaytondailynews.com/news/news/teradata-leadership-change-comes-as-company-strugg/nrHwg/
[12] http://www.sas.com/en_us/news/analyst-viewpoints/forrester-names-sas-leader-in-data-quality-solutions.html
[13] http://www.sas.com/en_us/news/analyst-viewpoints/gartner-names-sas-leader-in-data-quality-tools.html
[14] http://www.gartner.com/doc/reprints?id=1-2ZFVZ5B&ct=160225&st=sb
[15] http://www.forrester.com/pimages/rws/reprints/document/116447/oid/1-SFDMEH
[16] http://www.sas.com/en_us/news/analyst-viewpoints/2016-gartner-magic-quadrant-advanced-analytics.html
[17] http://blogs.forrester.com/brian_hopkins/15-11-03-ibm_and_teradata_a_tale_of_two_vendors_struggle_with_disruption

- Data *governance* is complicated by decentralized data *ownership* as functional executives control an increasing share of technology spending.
- The open source software business model offers an increasingly attractive alternative to commercial software licensing.
- Increasingly, the Hadoop ecosystem displaces conventional data warehousing; R and Python displace commercial analytic software.
- The elastic business model made possible by cloud computing undercuts conventional software licensing and provisioning.
- Widely available and inexpensive computing power make computationally intensive techniques like Deep Learning practical.

Consider what has happened to Teradata. Late in 2012, the company started missing sales targets; in early 2013, it stunned investors by reporting an absolute decline in sales. Management offered excuses; Wall Street punished the stock, driving it down by half in the face of an overall bull market.

From 2013 through early 2016, Teradata continued to miss sales and earnings targets; Wall Street drove the stock price down to a fraction of its 2012 peak. While it is tempting to blame the problem on poor leadership, Teradata's persistent failure to forecast its own sales and earnings indicates something amiss. The world changed; the value networks created in Teradata's rise to leadership no longer exist; the mental models managers used to understand the market no longer work.

Disruptive Innovation

Clayton Christensen of the Harvard Business School outlined[18] the theory of disruptive innovation in 1997. We summarize the theory briefly; for an extended discussion, read Christensen's book:

- Industries consist of value networks, collections of suppliers, channels, and buyers linked by relationships.
- Innovations disrupt industries when they create a new value network.

[18]Christensen, Clayton M. (1997), The innovator's dilemma: when new technologies cause great firms to fail, Boston, Massachusetts, USA: Harvard Business School Press, ISBN 978-0-87584-585-2.

- Not all innovations are disruptive. Many innovations are introduced by market leaders to sustain a competitive position.
- Disruptive innovations tend to be introduced by outsiders.
- Purely technological innovation is not disruptive; what matters is the *business model* enabled by the new technology.

Christensen identified two forms of disruption. *Low-end disruption* occurs when industry leaders enhance products faster than customers can assimilate the enhancements; the disruptor enters the market with a "good enough" product and a better value proposition. The disruptor's innovation makes it possible to serve customers at a lower cost than the industry leaders can deliver.

New market disruption takes place when the disruptor innovates in ways enabling it to serve customers that are not served by the industry leaders.

In this book, we discuss two kinds of disruption. The first is disruptive innovation *within* the analytics value chain (a concept we explore later in this chapter). The second is industry disruption *by* innovations in analytics.

There are many examples of disruption *within* the analytics value chain:

- Hadoop disrupts the data warehousing industry from below. Hadoop does not do everything a relational database can do; but it does just enough to offer an attractive value proposition for the right use cases. When first introduced, Hadoop's capabilities were quite limited relative to data warehouse appliances. But Hadoop's flexibility and low cost were highly attractive for applications that did not need the performance and features of a data warehouse appliance. While established vendors struggle to maintain flat and declining revenue, Hadoop distributors grow at double-digit rates.
- Tableau virtually created the market for agile self-service discovery. Tableau has no charting and visualization features not already available in mainstream business intelligence tools. But while business intelligence vendors targeted the IT organization in large enterprises and continuously added features, Tableau targeted the end user with a simple, easy to use, and versatile tool. As a result, Tableau has increased its revenue tenfold in five years, leapfrogging over many other BI vendors.

Examples of disruption *by* analytics are less prevalent, but they do exist:

- General-purpose credit scoring introduced by Fair, Isaac and Co. in 1987 virtually created a national market in credit cards. Previously, banks issued credit cards to their local customers, with whom they had an established relationship. Uniform credit scoring enabled a few large issuers to identify creditworthy customers in the general population, without a prior relationship.
- When the U.S. Securities and Exchange Commission authorized electronic trading in regulated securities in 1998, market participants quickly moved to develop algorithms that could arbitrage between markets, arbitrage between indexes and the underlying stocks, and exploit other short-term opportunities. Traders that most effectively deployed machine learning for electronic trading grew at the expense of other traders.

The relative importance of the two kinds of disruption depends on the reader's perspective. Disruption within the analytics value chain is pertinent for readers who plan to invest in analytics technology for their organization. Technologies at risk of disruption are risky investments; they may have abbreviated useful lives, and their suppliers may suffer from business disruption. Taking a "wait-and-see" attitude toward disrupted technologies makes good sense, if only because prices will likely decline in the future.

For startups and analytics practitioners, disruption *by* analytics is key. To succeed, startups must disrupt their industries. Using analytics to differentiate a product is a way to create a disruptive business model or to create new markets.

To understand disruptive *analytics*, we must first understand the current state of analytics and its drivers. In the remainder of this chapter, we present a discussion of what drives the demand for analytics, and an overview of the analytics value chain. We close the chapter with an outline of the rest of the book

The Demand for Data-Driven Insight

The key to survival in a disrupted world is to ruthlessly re-examine business processes, working backward from a problem.

Analytics is *the systematic production of useful insight from data*. In business, people use insight to solve one of five core problems:

- Develop a business strategy.
- Manage a business unit.
- Optimize a business process.

- Develop products and services.
- Differentiate products and services.

Each of these problems needs a different kind of insight, whose delivery requires distinctive people, processes, and tools.

Developing a Business Strategy

We define "strategy" narrowly to mean choices made by the top leadership of an organization: the "C-Suite". Many people may participate in the development of strategy, but in every organization, the buck stops somewhere. Strategic analytics are any analytics that support strategic decisions.

What makes an issue "strategic?" Strategic questions and issues have four distinct characteristics:

- The stakes are high; there are major consequences that depend on making the right choice. (Otherwise, the issue will be delegated.)
- The issue falls outside of existing policy; no established rule enabling decisions at a lower level. (There may be a conflict of policies, or the situation may be unprecedented.)
- Strategic issues are non-repeatable; in most cases, the organization addresses a strategic question once and never again. (Repeatable decisions are handled at lower levels through policy.)
- There is no clear consensus about the best choice. (If everyone agrees on the best choice from the outset, there is no need for analysis).

Examples of strategic topics include:

- Technology or product investments
- Mergers and acquisitions
- Business portfolio restructuring
- Business reorganization
- Branding, rebranding, and product positioning
- Crisis management

Since the stakes are high for strategic analytics, so is the sense of urgency; some decisions, like merger proposals, may be strictly bounded in time. Crises provoked by product failure, natural disasters, or other issues may have actual life and death implications.

Deliverables for strategic analysis include reports, charts, visuals, and presentations. Owing to the high stakes of the decision, executives closely scrutinize the presented analysis. Analysis must be "bullet-proof," especially if the results do not square with leadership's prior beliefs. The methods used to produce the analysis must be clear.

Due to the ad hoc and non-repeatable nature of strategic analytics, enterprise data warehouses (EDWs) play at most a supporting role. In most cases, the data in EDWs is internal and supports existing processes with well-defined requirements. The data needed to support strategic decisions often comes from external sources, may be difficult to access, and may be needed only once.

Enterprises frequently engage outside consultants to deliver strategic analysis. While organization insiders may have no experience in a particular type of problem, outside experts have deep experience with similar problems. Firms also prize consultants' independence and neutrality, since strategic decisions require resolving competing internal interests.

Managing a Business Unit

Managerial analytics support decisions a level down in the organization from top leadership. At this level, needs for analysis link to specific functions, such as Treasury, Product Management, Marketing, Merchandising, Operations, and so forth.

There are three distinct applications for managerial analytics:

- Performance measurement
- Performance optimization
- Business planning

Performance measurement is the sweet spot for enterprise business intelligence (BI) systems. BI is highly effective when the data is timely and credible, reports are easy to use, and metrics align with business objectives. Most organizations want to measure business units in a consistent manner, so they ordinarily implement reporting systems centrally rather than letting business unit managers measure themselves.

Metrics tell the manager which entities (e.g., brands, products, campaigns, stores, and sales reps) performed well and which entities performed poorly. Optimization delivers guidance on how to improve or optimize performance by shifting budget investments. Marketing mix analysis, for example, estimates the revenue impact of spending on different channels and programs, so the organization can shift the marketing budget to its most productive uses.

Finally, business planning is a process of goal setting and goal alignment across functions, where the manager justifies operating and capital spending. In large organizations, the business planning process is highly templated and structured. Forecasting is an important tool for business planning.

Deliverables for managerial analysis are similar to strategic analysis. Detailed analysis and forecasts may be in the form of queryable "cubes" or interactive tools.

Optimizing Business Processes

Optimization at this level is much more granular than optimization for functional leadership. In marketing, for example, the CMO needs summary information about the effectiveness of all major programs; the CMO's optimization problem requires shifting budget among programs. The program manager, on the other hand, seeks to optimally match programs, value propositions, and creative treatments to individual customers and customer segments.

There are many ways that analytics can optimize a business process. Examples include:

- Automated decision engines
- Targeting and routing systems
- Operational forecasting systems

Automated decision engines apply consistent rules designed to balance risks and rewards. Embedded analytics help optimize criteria and ensure that decision rules reflect actual experience. Decision engines are faster than human decision-makers and make better decisions. Examples include payment authorization systems and credit approval systems.

Targeting and routing systems evaluate the characteristics of an incoming message or request and direct it to the appropriate agent or subsystem. Analytics extract essential information from the request, eliminating manual evaluation and triage. Examples include e-mail routing systems in customer service operations and SAR investigation routing systems in bank anti-money-laundering systems.

Operational forecasting systems project key metrics that affect operations, enabling the organization to align resources accordingly. Analytics leverage historical data to detect traffic patterns and shift resources to locations or shifts where they are most needed. Examples include retail staffing systems that plan shifts based on expected floor traffic, and police patrol routing systems that direct officers to projected high-crime areas.

Analytics that optimize business processes are ordinarily embedded in production systems, and usually must operate in real time. This implies a need for streaming analytics, which we cover in Chapter Six. Analytic deliverables are machine-consumable models implemented in software.

Developing Products and Services

The development process in organizations runs the gamut from creative brainstorming to formal scientific research, as in pharmaceutical laboratories, to "skunk works" prototyping. As such, the range of possible analyses is extremely broad. Developmental analytics fall into two broad categories:

- Analytics for generating hypotheses
- Analytics for testing hypotheses

Managers perform or commission hypothesis-generating analysis to identify unmet consumer needs or gaps in existing products. This can include activities like analyzing external data consumer surveys and consumption data; analyzing operational data; or evaluating clinical reports of treatment for a certain disease.

At a later stage in the product development process, managers test hypotheses about specific product concepts, prototypes, or small production run. Analysis at this stage can include analyzing clinical trial data to determine the efficacy of a drug; and analyzing test market data to assess the value of a product feature, or similar activities.

For practitioners, specialized domain expertise dominates purely analytical skills in this area. (One would not expect a biomedical specialist who specializes in Parkinson's disease to easily switch to developing trading algorithms for a hedge fund.) Analytic processes must be highly flexible and agile, adapting to the particular problem at hand based on the product development cycle.

Differentiating Products and Services

We distinguish between analytics that *support* product development, and analytics that *are* the product, or embedded analytics.

For the previous four use cases, the "consumer" of insight is inside the organization—a top executive, functional manager, process participant, or product developer. Increasingly, however, analytics provide insight to end consumers outside of the organization. In these cases, analytics differentiate the product and make it stand out in the marketplace.

As the volume and variety of information available to consumers explodes, insight itself becomes a valued commodity. In this world, the most powerful analytic applications aren't often viewed as analytics. Is Google an analytic application? Google uses analytics technology, including content analytics and graph analytics, and it produces a particular kind of insight.

Online retailing's ability to carry a vastly larger number of unique items than brick-and-mortar retailers creates a shopping problem for consumers; with so many items from which to choose, what should we buy? Recommendation engines, which use machine learning to optimal products for an individual customer, are widely used. Most readers will be familiar with some salient examples, all of which use machine learning:

- Facebook leverages a user's profile and likes to optimize the news feed.
- Streaming video sites like Netflix leverage the user's ratings and other information to personalize recommendations.
- Tinder pairs users based on profile and "swipes."
- Amazon.com uses data-driven similarity ratings to display products that are compatible to what a user has selected.
- Spotify leverages a user's prior preferences and content analytics to optimize the music stream.

Success in embedded analytics is a matter of software engineering; the end product must be tightly packaged for reliability and usability; in most cases it must operate in real time.

The Analytics Value Chain

Once we understand the demand for insight, we can define a value chain. The analytics value chain begins with data and ends with insight, progressively transforming data from low value to high value in a sequence of steps, as Figure 1-1 shows.