Companies in the capital and commodity markets are looking at analytics for opportunities to improve revenue and cost savings. Yet, many firms are struggling to develop a comprehensive, cohesive and sustainable analytics strategy to support their uses. In this article, Rashed Haq introduces the key components of an analytics strategy and how to use it to develop a strategic roadmap.

According to Forrester, analytics is the top priority in terms of investment in business applications for firms, up from the fourth priority in 2010. Their survey of almost 1,100 companies shows that 49% of firms have near-term plans to adopt analytics, and this number is 10% higher for top-performing companies.

These statistics indicate a market trend toward using analytics for differentiated business growth and performance. But, with all the different models and tools available, using analytics is not without its challenges. These can be grouped into three primary areas: credibility, variety and complexity.

- The **credibility** of analytics usage is low, but improving. The term “analytics” has become somewhat of a catchall phrase and without understanding the different types, it becomes difficult to pursue. The hype surrounding “big data” is adding to the confusion as firms may jump in without knowing when analytics should be used—and how.

- The **variety** of use cases within a firm can be high. These appear as disparate and disconnected solutions. Without a view into the similarities of these different methods and the potential for standardization, firms struggle to develop a comprehensive path forward.

- The **complexity** of modeling and data management is high. Fine-tuning the model for the relevant decision variables takes skill, patience and time. Additionally, there is a recent proliferation of different types of technologies, best suited for different requirements and with limited applicability in each.

These challenges make it difficult for firms to proactively pursue a sustainable analytics roadmap, forcing many to implement analytics in silos or to avoid it altogether. To overcome the above challenges, firms will need to develop a cohesive analytics strategy and roadmap.

**STRATEGY COMPONENTS**

A strong analytics strategy relates business goals and use cases with how analytics will support employees and the business. To ensure a cohesive and sustainable strategy, the following components should be included (see Figure 1):

---

**Figure 1: Analytics Strategy Framework Components.**
1. **Goals** set out the purpose and vision for analytics within the firm. This may include sustained competitive advantage, incremental revenue opportunities or cost reduction.

2. **Use cases** identify the potential short- and long-term uses of analytics to drive the goals. They will be used to:
   a. Understand the overall business case and show the different types of value that can be derived from analytics
   b. Define holistic requirements within the firm for the types of analytics that will be required and the associated information management to support them
   c. Understand the user groups and business processes that may be impacted and whether the usage will be a one-time process or an ongoing operational process

3. **Quantitative methods** define the different types of analytics that will be required over time to support the different use cases.

4. **Architecture** encompasses the technology components and platforms that will be used to support analytics. It covers information gathering, storage and processing, analytics modeling, visualization, user experience and history maintenance.

5. **Data readiness** defines the strategy that will enable firms to ensure that all relevant data is available with the appropriate quality and timeliness. This may include activities such as data quality assessment and remediation, data lifecycle governance, etc.

6. **Organizational capability** establishes the organizational architecture required to leverage analytics. This includes decisions around structure (e.g., should analytics reside within one group, be embedded in business groups or maintain a community of practice?), the processes to sustain the analytics lifecycle and the necessary group capability.

7. **Governance** defines the structure and processes required to sustain the analytics capability and strategy. It will determine owners, standards, value measurement, project approval and prioritization, etc. Governance will also define the research agenda in terms of market analysis, competitive assessment and vendor assessment.

---

### USE CASES

Use cases are the areas of business operations where analytics could be used within a company. They help to define the depth and breadth of how analytics will be used in the business and how that will help the firm. The use cases will be specific to the type of company and the strategy they are following. Below are some examples of use cases across different companies:

- **Portfolio optimization** can help the firm to improve commercial decisions for transactions within a physical or financial portfolio by leveraging an equation-oriented, data-driven toolset like a chess simulator (see CROSSINGS Edition 7, Fall 2012, page 42)

- **Energy intelligence** can be used to help the company gain a qualitative & quantitative understanding of near-term market operations, hence improving the organization’s ability to anticipate or quickly identify arbitrage or risk mitigation opportunities (see CROSSINGS Edition 8, Spring 2013, page 18)

- **Surveillance, compliance and fraud detection** enable the firm to correlate data from multiple sources to identify potential fraudulent activities closer to real time (see CROSSINGS Edition 7, Fall 2012, page 6)

- **Risk-based maintenance** can help the organization to proactively identify what inspection or repairs are required for their facilities and assets based on risk factors, hence minimizing downtime, costly reactive maintenance and potential brand and legal ramifications (see page 33)

Identifying the potential use cases within the firm allows for a deeper understanding of the business needs, the requirements for the technologies and quantitative methods that will be required and any commonality that may be leveraged across the different use cases. Additionally, these may be used as the foundation for the overall business case to develop the firm’s analytics capabilities.

---

### QUANTITATIVE METHODS

Quantitative methods are the heart of how an analytics strategy is different from other strategies. This is what gives analytics its power and utility. The primary focus is to “model” real-world systems, such as refineries, financial portfolios or workers’ schedules. The four key ways that models may be represented are as an operational exercise, a game, a simulation or an optimization model.
Operational exercises represent a model by executing experiments with the real-world system and leveraging the results to make decisions about how to operate in the future. As a simplified example, a refinery operator may say that next week it will run light crude and the following week it will run midgrade sour crude. The operator’s goal is to understand which type of crude might be more profitable. It’s important to hold as many of the other variables as fixed as possible, and then measure, analyze and interpret the results of the “experiments”. This model representation is the most realistic because the real-world system (e.g., refinery) is being used. It leverages the least information technology (e.g. historical reports) and is slower and more costly.

Gaming represents a model by creating a simplified response to various scenarios or strategies. Continuing the previous example, this approach may entail getting the decision makers at the refinery in the same room to talk through the different scenarios. The goal would be to solicit everyone’s responses in terms of what actions they would take and what decisions they would make about how much to produce, what the expected costs would be, etc. This is similar to the previous representation, but instead of running the “experiment” through the real-world system, it is abstracted in a meeting with whiteboards, spreadsheets, etc. Some of the advantages of this approach are that it pulls on the intuition and experience of the experts and provides the opportunity to spot conflicts. Gaming is generally used as a management-learning tool to teach the complexities in the decision-making process. This model representation is less realistic, but also does not leverage significant information technology, making it faster and less costly.

Simulation is similar to gaming, but in it the decision makers are augmented with quantitative models. This can include creating a simulation of a real-world system that has not yet been run (e.g., simulating the refinery operations for the next month), or it could be a simulation of various scenarios as a diagnostic to understand why something has happened. The model evaluates the performance of the alternatives. This model representation is more abstract, uses significant quantitative methods and requires information technology; therefore, it is faster and less costly than using the real system.

Optimization represents a model completely in mathematical terms, usually by setting an objective that needs to be maximized or minimized under different constraints. The model finds the best possible value of the objective function that also satisfies all the constraints. Unlike the previous types of model representations, this model generates the best alternatives rather than evaluating one supplied by the decision makers. Similar to simulation, this model representation is more abstract, uses significant quantitative methods and leverages information technology; therefore, it is faster and less costly.

The latter two model representations are considered primarily quantitative, while the first two require relatively simple quantitative approaches. The latter two require thoughtful, quantitative articulation of the problem to be solved. This takes considerable deep business expertise, fine tuning and back testing.

In addition to the model representation, the other key consideration for quantitative methods is whether some—or all—of the inputs into the model representation are deterministic or stochastic. In other words, are there unpredictable fluctuations in the inputs? This is important because the quantitative methods that need to be employed will be significantly different if the parameters or values are stochastic, or random in nature. Figure 2 shows examples of the types of quantitative methods that must be applied for the different model representations and problem types. It is important to note that even with a large variety of use cases, the quantitative methods that may need to be utilized are generally comprised of a small, finite set. Knowing this makes it easier to plan for—and leverage—similar quantitative capabilities across the firm.

![Figure 2: Categories of Major Quantitative Methods.](image-url)
ARCHITECTURE FOR ANALYTICS

The architecture required to support analytics has many similarities with non-analytic architectures, but with a few key differences. As shown in Figure 3, the different components of the architecture framework include:

- **Data management**, which is the platform for managing data sources and integration
- **Data grid**, which is the high-performance data storage or in-memory grid
- **Compute grid**, which is the corresponding high-performance and parallelized data processing and computation component
- **Usage component**, which includes reporting, business intelligence and visualization

The key differences for analytics include the addition of the compute grid and the considerations for combining structured and unstructured data—potentially with real-time streaming data at high volumes and high speeds. Real-time information generally requires very high performance. The last decade has seen significant cost improvements in all areas relevant to analytics, bringing previously cost-prohibitive opportunities within reach for most firms.

Data gathering includes finding the **data sources**, potentially acquiring the data through a data service agreement, performing any validations and formatting necessary and storing the data. This data will eventually be used for exploration, analytics, model-building and back testing. And it may come in a variety of formats, such as relational, structured, unstructured and real-time streaming.

The **data integration** layer provides a complete set of capabilities for data management across relational, non-relational and streaming data throughout the full data lifecycle with the ability to:

- Ingest several varieties of data, massive volumes of data and real-time data (or events)
- Seamlessly move data from one type to another
- Cleanse and refine data in a business context, including eliminating redundancies, removing obsolete data, correcting inaccurate data and enriching missing data
- Correlate datasets through master data
- Address security and availability concerns

Similar to quantitative methods, the architecture components can be standardized based on the types of use cases within the firm. Not all firms will require all components, but as the breadth of use cases are identified, a standardized architecture can be developed. Figure 4 shows some of the key components.
The data grid provides the software and hardware for managing the data necessary for analytics. This includes everything from traditional enterprise data warehouses and data warehouse devices to more modern technologies, such as:

- Aggregate databases that allow seamless integration between structured and unstructured data forms
- Hadoop MapReduce for large-scale batch-oriented analytics
- In-memory data for real-time analytics using in-memory databases (IMDB) or in-memory data grids (IMDG)

The compute grid provides mathematical and statistical methods for calculations. And it encompasses the software infrastructure needed to transparently distribute and parallelize computation-intensive tasks across groups of networked multicore computers to optimize for efficiency or time. This supports scheduling work requests to all available computing resources, including local grids, remote grids, virtual machines and dynamically provisioned cloud infrastructures. It may also include specialized computational software, such as CPLEX or Algorithmics, or custom software on top of the grid software.

Finally, usage applications cover the following:

- Reporting and business intelligence, with static reports, parameterized reports, ad hoc queries and automated or on-demand, web-based or mobile delivery
- Interactive, multidimensional OLAP-style slicing and dicing with drill up and down capabilities, including hierarchical aggregation, averages, comparisons, pivots, etc., for pattern discovery and forensic analysis
- Visualization with interactive graphs, charts, real-time monitoring dashboards and maps

These components work together to bring analytics capabilities online, for all types of use cases and analytics needs.

CREATING A ROADMAP
Companies currently leveraging analytics capabilities are already gaining an advantage over their competition in terms of being able to make better near-term operational and commercial decisions in response to market changes. These firms will need to standardize and cross-leverage quantitative methods and architectures across use cases to prevent runaway costs.

For companies starting on their analytics roadmap, it will be critical to leverage the patterns discussed in the analytics strategy framework to help define a sustainable path forward. This involves selecting the relevant set of quantitative methods and establishing the appropriate infrastructural framework to power it. Given the finite set of quantitative methods that support the vast majority of business use cases and the recent evolution of cost-effective technologies, robust analytics is now within reach of most firms.

Rashed Haq is Vice President of Energy Commerce at Sapient Global Markets. Based in Houston, Rashed specializes in trading, supply logistics and risk management. He works with oil, gas and power companies to create innovative capabilities, processes and solutions for their most complex challenges in business operations. rashed@sapient.com