

8

Deployment and Life Cycle Management

Arnie Greenland

Robert H. Smith School of Business, University of Maryland, College Park, MD, USA

8.1 Introduction

This chapter is appropriately placed near the end of this book because it pulls together all of the various ideas presented in earlier chapters. Ultimately, the goal of analytics professionals is to create and implement meaningful, successful, and sustainable analytics solutions using the tools of analytics, which have been discussed in Chapters 5 and 6. The successful implementation of an analytics project relies on strong project management skills and on leveraging the analytic and data insights that are the unique contributions of analytics professionals. These have been covered in Chapters 1–4. The life cycle described in this chapter focuses on joining those pieces together in a structured and ordered way, while detailing the special nature, complexities, and challenges of delivering analytics projects.

A life cycle is defined as a sequence of phases in the process of developing an analytics model or system. It is very similar to the term used in the more general context of a software or systems development life cycle common in the discipline of information technology (IT) management. The intent of this chapter is to focus very specifically on those phases that have been identified by the developers of the CAP certification process and that are also closely related to standard methodologies accepted broadly by practicing professionals. The analytics project life cycle may actually be a component of, or integrated with, a larger information technology life cycle.

The analytics system/model life cycle is composed of several phases: initial design, development, testing, implementation, deployment, and postdeployment monitoring. The actual calendar time that transpires in each of these various phases can vary widely but depends on the specific characteristics of the model or system being developed. Total time for this entire process can range from months

to years, depending on complexity of the data relationships being modeled. The postdeployment monitoring may span multiple years.

As with any complex project, the analytics professional (AP) leading a project shoulders the responsibility of understanding the specific steps or phases and the order of those steps he/she will traverse to accomplish the stated goals. Typically, professionals refer to such a set of steps, and any complex interactions, as a *methodology*. There are a multitude of methodologies in use for project management in general, and for analytics projects in particular.

8.2 The Analytics Methodology: Understanding the Critical Steps in Deployment and Life Cycle Management

A popular and accepted methodology in the analytics community is CRoss Industry Standard Process for Data Mining (CRISP-DM). This methodology was created initially as a cooperative effort of a number of companies that were interested in data mining, including SPSS (IBM), Teradata, Daimler AG, NCR Corporation, and OHRA, an insurance company. While it originally focused on data mining, analytics professionals in many fields have found it useful for projects of all types and is the closest thing at this point to a “standard.” For example, a recent article in the online site KDNuggets, written by George Piatetsky,¹ states “CRISP-DM remains the top methodology for data mining projects, with essentially the same percentage as in 2007 (43% vs 42%).” The only other data mining standard named in this article was SEMMA,² though its use reported in the article fell from 13% in 2007 to 8.5% in the 2014 survey. While not used universally, it is about five times more often used than the next methodology mentioned, so it is the closest thing we have at the moment to a “standard.” The CRISP-DM methodology is captured in the following commonly available diagram.³

There are six major components of the methodology:

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Evaluation
- Deployment

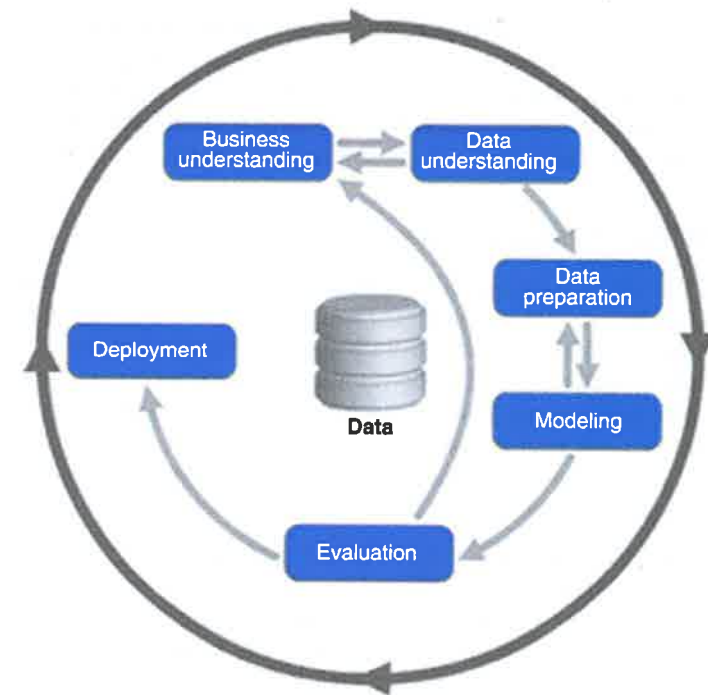


Figure 8.1 CRISP-DM diagram.

A very important feature of the CRISP-DM, shown visually in Figure 8.1, is that there are many feedback loops in the process. In simple terms, what you learn or encounter in one step of the process often impacts an earlier phase of the process, so the diagram shows either an arrow in both directions or an arrow from a later phase coming back to an earlier phase. For example, the dual arrows between business understanding and data understanding emphasize the important notion that you cannot understand the data unless you also have a very deep understanding of the business issues, and as you have questions about the data, it is best to reach back into the sponsoring organization to ask those questions and obtain understanding.

A very important point here is that these six components map nicely to the CAP Job Task Analysis (JTA),⁴ so throughout this description of the phases of CRISP-DM, we will freely integrate detail from the JTA as a way to explain the phases but also as a way to link the methodology directly to the types of information that analytics professionals, working toward the CAP Certification, need to know from the JTA.

1 Piatetsky, G, KDNuggets (2014) CRISP-DM, still the top methodology for analytics, data mining, or data science projects. Available at <https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html> (accessed October 2014).

2 A data mining methodology developed by SAS Institute, Inc.

3 This specific version of diagram on CRISP-DM is obtained from https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

4 See Appendix A of this document.

The CRISP-DM methodology has been repeatedly praised because of its clear recognition of the importance of a strong connection to the sponsoring business or other type of organization that is seeking to create an analytics model or solution, so it is only natural that we begin this discussion with the phase of the CRISP-DM focused on understanding the business situation.

8.2.1 CRISP-DM Phase 1: Business Understanding

The first phase of the methodology in CRISP-DM is business understanding, and this phase corresponds to the first two domains of the JTA: Business Problem Framing and Analytics Problem Framing. Both the CRISP-DM and the JTA recognize the critical importance of understanding the business issues and focusing specifically on how analytics could have a real (and measurable) positive impact on that business situation.

Consider Domain I of the JTA, Business Problem (Question) Framing. The tasks enumerated therein are as follows:

- Task 1: Obtain or receive problem statement and usability
- Task 2: Identify stakeholders
- Task 3: Determine if the problem is amenable to an analytics solution
- Task 4: Refine the problem statement and delineate constraints
- Task 5: Define an initial set of business benefits
- Task 6: Obtain stakeholder agreement on the problem statement

This set of tasks lays out an excellent set to follow. Upon completion of these steps is a clearly defined and documented business problem statement.

8.2.2 JTA Domain I, Task 1: Obtain or Receive Problem Statement and Usability

A business problem statement is a clear and concise description, typically written in business terms, of what the business or organizational objectives the sponsor wants to reach. The business problem statement defines the key outcomes or accomplishments that are desired, how the organization will measure whether these outcomes have been reached (for example by specifying business metrics that would be measured and the levels or targets for those metrics that represent success), and all other relevant business issues, such as time frame, cost constraints, and other business requirements.

Clearly, the best case scenario is that the organization or business that is sponsoring the analytics project can simply deliver a complete and fully thought out Business Problem Statement. In practice, this rarely happens. It is much more common that the analytics professional (or AP) and team must create the Business Problem Statement working together with the sponsor. The work involved in doing this ranges over a wide set of possibilities. It may turn out to

require a small amount of work to take an initial version of the business problem statement to a good initial place, but it is actually more common that it requires creating the document totally from scratch.

We complete this discussion focusing on the common situation of not having a starting version of the document at all. Even if the sponsor can provide an initial version, the same set of activities would be done, though maybe at lower levels of intensity. An important component of creating such a document is conducting detailed interviews to discover and document the business situation. The skills required in such an endeavor are many. As a baseline, the analytics professional needs to know what constitutes a clear and usable problem statement. While project management experience in general is needed for this activity, it is also important to be attuned to issues that impact the use of analytics as a possible solution such as time requirements for completion, availability of data, the structure and form of the data, and availability of business experts at the level required to assist in the process. Next, the AP needs to have the skills to interview knowledgeable individuals, ask appropriate questions, and ultimately document the findings. Successfully obtaining the information needed requires the ability to be persistent in making sure that there is clarity in the understanding of the business issues and that follow-up be done to fill in holes in understanding. The AP will also need to possess basic business knowledge such as how businesses are typically organized, how this sponsoring organization is organized in particular, how business processes are defined, and how they work together to fulfill business objectives.

The deliverable from this step in the process is better described as a starting point for the development of the Business Problem Statement. It is a document that describes the business situation, lays out the problem that is intended to be solved, and the basic business metrics that will be used to measure success. This document, as will be discussed in continually increasing detail in the following sections, will be enhanced by focusing on different components and aspects of successfully documenting the problem to be solved.

8.2.3 JTA Domain I, Task 2: Identify Stakeholders

This task reinforces the notion that success of any project requires involvement of key players in the organization, both at senior levels to provide the organizational backing, funding, and other resources, and to assure involvement of junior stakeholders. It is a commonly accepted understanding that the higher level involvement in, knowledge of, or enthusiastic support for a project, particularly an analytics project because of its complexity, is closely related to project success. It is also critical to have involvement at lower levels in the organization because at those levels you can find individuals with time and access to information (e.g., the data you are hoping to get) that will be essential to proper functioning of the project.

It is recommended to reach out to the full range of stakeholders early in a project, to meet with and interview as many as possible, to acquire and document their views and expectations. In some cases, it is beneficial to bring stakeholders together to share ideas and work toward an organizational consensus. However, one should be very careful in planning such a meeting, especially if there are highly divergent views in the organization that could result in a more difficult situation than if you did not bring stakeholders together.

INTERVIEW WITH ALAN TABER

When asked about what constrains the analytics professional's choice of problem-solving approaches/methods/models, Lockheed Martin Missiles and Fire Control's System Engineer Alan Taber offered the following thoughts.

Time and money are two key constraints—that goes without saying. The third constraint is how data-driven the stakeholders are—or aren't. Obviously the stakeholders are willing to listen or they wouldn't be willing to talk to you in the first place, but some stakeholders are more intuitive, while others are more skeptical. In almost all cases, when you are working with executives, you should bear in mind that they got to where they are because they have been successful in the milieu their companies or organizations have been operating in for the last several years.

This leads us to a discussion of change management. The solution to a problem typically requires change. So change management is inherently bound into analytics. The analytics professional is ultimately going to suggest that someone do something differently, invest differently, organize differently, go after a different market

segment, or go after the same market segment been in a different way. When you are managing change, you have to know how aggressively and how quickly you can guide someone along the path of change. This is a nebulous concept, and understanding and appreciating it comes with experience.

The analytics professional needs to support his or her clients in their decision-making, help them better understand the decisions they are trying to make, help them better understand the vision of where the organization could be if these decisions are made and plans carried out effectively. The analytics professional has to do so without burning out clients and making them dread the sight of him or her walking through their door. If the client feels the analytics professional always wants more time, more money, more this, more that—or is pushing them faster than they want to go—communication starts to shut down. And if the client and the analytics professional are not communicating, the analytics professional cannot solve the client's problems because the client has been open about his or her business problems.

This is an excerpt from one of a series of interviews with analytics professionals and educators commissioned by the *INFORMS Analytics Body of Knowledge Committee*.

8.2.4 JTA Domain I, Task 3: Determine if the Problem Is Amenable to an Analytics Solution

This is one of those project tasks that are unique to an analytics project. As the AP begins to understand the problem, it is important to begin thinking about, but not finalizing, whether analytics makes sense for this problem or business situation at all. Since nearly every analytics solution (discussed in more detail in Chapters 5 and 6) requires data, this is the place where the AP begins to determine whether there are data available, whether they are easily accessible, and whether there are issues relating to obtaining the data (classification of data as confidential or higher levels of classification, or whether there are privacy issues). Assuming, the AP believes that there are data and that they could be obtained, it is also important to begin thinking at this point about whether the possible size and scope of the problem is tractable. For example, is it possibly too big, too complex, or requiring more time and resources than are available? Finally, the AP should begin considering quality and timeliness of a possible solution. It might be that to perform the analysis that the key stakeholders expect will take more time and consume more resources than that organization is willing to commit. If that is the case, the AP should be very careful in committing to taking on the project.

8.2.5 JTA Domain I, Task 4: Refine the Problem Statement and Delineate Constraints

This task is essentially a recognition that as one learns more about the business problem; the problem statement may change or become more clearly understood. With additional information obtained from Tasks 2 and 3, the document created in Task 1 should, therefore, be modified. Many practitioners see a Business Problem Statement as an organic document, meaning it is a document that develops or grows over time as the authors learn more about the business situation, the expectations of key stakeholders, and other critical factors. We will see that the problem statement document may be revisited many times, further justifying considering it, from the beginning, as an organic document.

8.2.6 JTA Domain I, Task 5: Define an Initial Set of Business Benefits

One of the most critical initial steps in any project, and a key success factor, is to make crystal clear, right up front, the business goals and accompanying business benefits that are expected. It is possible that the organization seeking analytics expertise has created a written document to define the expected benefit; but it is also common that no such document exists, and even if one does exist, it may require work to clearly understand the proposed benefits. The analytics professional will very likely need to rely heavily on communication skills, such as

interviewing and information extraction skills, to obtain and document the expected business benefits.

Typically, the business goals and benefits will be discussed, and more importantly measured, with reference to the business metrics that the organization as a whole uses to manage the business. It is not uncommon for businesses to have dozens, sometimes many hundreds, of metrics that are obtained, studied, and communicated, as they see fit, in their organization. So, it is also likely that the goals of the project you are in the process of designing will measure the benefits in terms of those same standard business metrics. For example, if you are working on a project to improve the efficiency of a fulfillment process within a retail business that is linked to a carefully managed inventory system, you would expect goals such as the following:

- 1) Lower the rate at which orders requiring a backorder to be required by a stated percent
- 2) Lower the error in demand forecasts by some stated percent or absolute error bound
- 3) Lower personnel costs in the warehouse (through better demand forecasting and more efficient staff scheduling) by a stated percent

The specific goals and associated metrics will be tightly linked to the goals of the organization or business and will use the same set of metrics that are fundamental to the efficient operation of that organization or business. The stakeholder interviews are the right place for the AP to delve into how the business metrics are determined or computed, where the data to do so come from, and how the resulting metrics are used to manage the business.

The outcome of this task is a clearly written section of the business statement that describes the expected business benefits. This section should be communicated in terms of the core business metrics documented in this task but would go further to lay out how, after the analytics project being considered is complete, those business metrics would be altered and, hopefully, improved. This is also a place for the analytics professional to proceed very carefully. The planned or expected business benefits need to be reasonable to achieve, considering all of the factors already discussed. These include the availability and quality of data required for the project as well as the time and resources that the organization has to dedicate to the project.

8.2.7 JTA Domain I, Task 6: Obtain Stakeholder Agreement on the Business Statement

The final task within Domain I closes the loop with the sponsoring organization or client by communicating all that was learned during prior five tasks. A critical success criterion here is that the language and terminology of this Business Problem Statement be consistent with that typically used by the stakeholders in

managing their business. Using analytics jargon, complex mathematical or statistical words, or concepts that are foreign to the sponsoring organization may create a feeling among the stakeholders that the analytics team does not understand their business, or worse that the stakeholder community may lose faith in the ability of the analytics team to solve their business problem. It should be a document that is readily understandable to a mid-level manager in the sponsoring organization. The AP relies primarily on their information harvesting, learning, and communication skills, either written or verbal, to communicate the essential aspects of the business problem and the business needs to be satisfied in the project to the key stakeholders identified in this phase. A good strategy to avoid communication problems is the kind of regular communication throughout this phase, which has already been discussed. The goal in this task, of course, is to receive the written approval from the primary contact or key stakeholder on the project to proceed to the next step.

Creating the Business Problem Statement is a very important milestone for every project, whether it be an analytics project or not; and we have included this domain of the JTA as part of Business Understanding, the initial phase of the CRISP-DM methodology. However, as we will emphasize throughout this chapter, analytics projects have a number of unique characteristics that fall squarely in the hands of the analytics professional. One of those unique characteristics is the second Domain of the JTA: Analytics Problem Framing. We are including this domain along with Business Problem Framing (Domain I) as part of the Business Understanding phase of the CRISP-DM methodology because we see it as a critical aspect of the initial activities of the project and is clearly included in the full understanding the business problem, and more particularly whether there is a credible analytics problem and reasonable solution that is underlying this business situation. As indicated in the JTA, this domain includes the following five tasks:

- Task 1: Reformulate the problem statement as an analytics problem
- Task 2: Develop a proposed set of drivers and relationships to outputs
- Task 3: State the set of assumptions related to the problem
- Task 4: Define key metrics of success
- Task 5: Obtain stakeholder agreement

8.2.8 JTA Domain II, Task 1: Reformulate the Problem Statement as an Analytics Problem

After completing Domain I, we have an initial business problem statement document. To the analytics professional, this begs the question: What is the analytics problem? Indeed, in some cases there may not actually be an analytics problem. It may become apparent after evaluating the information in the problem statement that a different solution, other than an analytics approach, is the

right course of action. For example, it may be clear from the problem statement that the solution will emerge by viewing this business situation from a purely management, information technology, business process, organizational, or personnel perspective. In which case, a different team may be best to carry the project forward.

However, fortunately for the analytics profession, many of the types of business problems we encounter are best solvable by bringing business analytics tools and techniques to their solution, and Domain II of the JTA is then relevant. This step is important to be done primarily by the analytics professional and his/her team of other analytics professionals, but it is also important to make sure that the level of communication started in Domain I continues into this part of the project as well. So, as the tasks in this domain evolve, they should be presented to and discussed with the key stakeholders identified earlier in the Business Understanding phase.

Consider now the specific Task 1 of reformulating the problem statement. The skills and knowledge on the part of the AP of creating the Analytics Problem Statement are similar to those for the phase of the CRISP-DM called Modeling, and there are two domains of the JTA (Methodology Selection and Model Building) that are relevant. There are also two other chapters in this book (Chapter 5 on Solution Methodology and Chapter 6 on Model Building) that cover a great deal of the same territory. To resolve this apparent overlap, we focus in this section only on the life cycle issues in the creation of the Analytics Problem Statement document, as this section is intended only to lay out the modeling approach at a high level. We will leave the activity of selecting the specific model and building that model to later parts of the project life cycle that happen *after* the data have been obtained and carefully analyzed.

The focus of this task is to review the modeling concepts (that have been discussed in Chapters 5 and 6) and, based primarily on knowledge of the capabilities of those modeling approaches and the experience of the analytics team, create a plan. Since this is happening before one has data, or the chance to analyze that data, by necessity, this plan should include a range of possible modeling approaches that will be considered. For example, suppose you are analyzing a problem that appears to require classification of a newly arrived transaction or “case” in some business situation into one of several treatments by the business. From this simple description, it appears that one of the classification algorithms common in the data mining or machine learning literature would be appropriate. But it would be premature to specify which specific model from a broad range of possible models (e.g., logistic regression, classification trees, neural networks, or ensemble models) is best to use. Such decisions are best left to the time in the life cycle where the data are available, properly prepared, and analysis of that data has begun. Therefore, the Analytics Problem Framing should include this range of possible approaches and avoid getting too specific on the solution approach.

Before one can successfully frame the analytics problem, there are a number of important activities to complete, and we suggest that these activities are exactly what are shown in the JTA as Tasks 2, 3, and 4, shown in the list above. We suggest, here, that one cannot actually complete the drafting of Analytics Problem Framing document without completing those tasks. These tasks include the following:

- Thinking through (and enumerating if possible) the key drivers or sets of relationships in the data that will allow the model to reach an acceptable solution
- Enumerating the assumptions that are needed in the modeling activity
- Defining metrics of success

Rather than changing the order of presentation, we postpone the description of those activities to later in this subsection but suffice it to say that we will need them to be completed before attempting to create the Analytics Problem Framing document that is discussed in the next paragraph.

The key contribution of this task is to take the business problem statement created in Domain I, along with the results of Tasks 2, 3, and 4 below, and expand (or append) a description of how analytical methods can be brought to bear to reach a solution. Many of the same skills required for the selection and building of the models, at later phases of the project, are also needed at this stage. Also, the resulting expanded or separate document should be created and then communicated in writing or through a presentation to those in the sponsoring organization for the project, most likely a subset of the key stakeholders mentioned earlier. Because this is a task that needs to be accomplished without having the benefit of analyzing data, it is potentially more difficult (maybe we should say, more risky) than what comes later. It is important that those performing this task have experience in building models of the type they will recommend, to avoid any pitfalls or issues in implementation that might surface at a later phase of the process when data are on hand to work with. Also, it is a good policy to provide a wide range of possible solution approaches if it is not crystal clear which analytical method will work the best prior to getting into the data.

8.2.9 JTA Domain II, Task 2: Develop a Proposed Set of Drivers and Relationships to Outputs

This task focuses on the **structure** of the model that we propose to build. A fundamental aspect of all modeling is the notion of the logical organization and presentation of the things we know (the inputs) being used by an analytics model to obtain the things we want to know at the end (the outputs). A big part of the modeling process is to sort through all of the data and discover the key drivers and important relationships that will be exploited so that the model we build will produce the required output. Such outputs include, for example, a correct

estimate or prediction, an optimal allocation of resources, a more efficient process, or a critical business decision that needs to be made. This task is intended to lay out what we know about the drivers and relationships based on information that we can obtain from prior work, that can be obtained by interviews or discussions with key stakeholders, or can be suggested based on prior experience with modeling similar business situations. The output of this task is a technical write-up that can be included in the Analytics Problem Framing document being compiled for this task. The content of this write-up should be a clear description of these drivers and relationships, and how, when they are put together in a model, the goals of the modeling activity can be achieved.

Analytics skills required to perform this task are knowledge of the modeling tools, some initial knowledge and understanding of the data that are available, and experience in building models of the type being recommended in the past. In addition, as with nearly every task in the JTA, the analytics professional needs softer skills of written and verbal communication and persuasion to be effective in creating and effectively communicating this new write-up that will appear as a section of the Analytics Problem Statement.

8.2.10 JTA Domain II, Task 3: State the Set of Assumptions Related to the Problem

Similar to thinking through the structure of the proposed modeling approach, it is important to make clear to the sponsoring organization what assumptions are being made. The assumptions we are thinking of in this context are primarily technical or analytic assumptions that impact the modeling. For example, if we are applying a predictive tool, such models often have underlying assumptions such as normality of errors. While it is typically not useful to try to explain complex statistical or other modeling assumptions to stakeholders in the sponsoring organization who do not have training in those areas, it is important to communicate that models come with assumptions and focus on assumptions that are generally accessible to a larger audience. An example of such an assumption is that the future (e.g., demand for a product) will behave similar to the past. This is clearly an assumption analytics professionals accept in many models, such as forecasting models and other predictive models; but it is one that business professionals can understand as well. They will also understand the risk associated with that assumption, that is, there is a chance this assumption may not be fulfilled in practice.

As with Task 2, the written output of this task would be a component or section of the larger Analytics Problem Framing document. The skills and knowledge required to perform this task are similar to those required for the model structure task that involves enumerating of drivers and relationships for the proposed model (Task 2).

8.2.11 JTA Domain II, Task 4: Define the Key Metrics of Success

This activity is closely related to Domain I, Task 5, in which we define business benefits. In this task, we focus on how the modeling activity can improve the business situation. The key metrics of success will employ the same set of metrics developed in that prior task but will focus on how the model or set of models being considered operate. This includes how the specific improvements that are contemplated will be measured, reported, and interpreted. Again, the output from this task is a written section to be included in the business problem statement as modified to incorporate the analytics problem framing.

We cannot emphasize enough the importance of this task. Many projects, whether analytics projects or not, run into issues at future stages because the key success metrics were not written down clearly and goals of achievement for each of the key metrics not crisply defined. Therefore, great care and attention to details needs to be focused on this task. Along with the definition of these success metrics, it is also important at this stage to think how difficult it will be to access the data needed to compute or estimate these metrics and whether those data will be available at the time the success metrics are to be presented for evaluation.

8.2.12 JTA Domain II, Task 5: Obtain Stakeholder Agreement

The output from Tasks 1–4 is a revised business problem statement. It includes the original business problem statement as well as the analytics problem statement. These documents may be separate or a combined document based on the preferences and discussions of the stakeholders and the analytics professionals involved. The analytics problem framing is intended to be at an appropriate level of detail so that the organizational stakeholders can confirm that the analytics team understands both the business problem and the analytical solution statement, and that it is consistent and supportive of the type of solution and business outcomes that they are seeking.

The goal of this final task is to present the stakeholders with a written document, possibly along with a presentation, to explain the plan to the stakeholder group and answer questions. This is also an appropriate time to focus very clearly and decidedly on the notion of stakeholder expectations. The set of expectations includes both expectations about the outcomes, as far as business benefits, improved metrics, and so on; and expectations about the process of creating the model and the postdeployment requirements for maintenance.

As far as the process of creating the model is concerned, all analytics professionals are aware that modeling is complex, takes time, requires good data, and cannot promise success (in advance). Also, the modeling process may be foreign to many business people. A good example of this is building an optimization model. Practitioners who work in this area all know that despite

spending a great deal of time working to extract the full set of constraints, the first time the analytics team “runs” the newly created optimization model, it is very common that the stakeholders working closely with the analytics team will say something like: “that cannot happen in our business.” Of course, what really occurred was that the constraints that would preclude that model outcome were not included in the model run. So, the AP and team go back to the model, add these new constraints, and the process continues, possibly with additional such iterations until the right model is created and running.

Experienced professionals know that the process just described for optimization models and, indeed, for nearly every type of model that analytics professionals build, is common and in fact expected. However, it may not be what the stakeholder team monitoring the model building process expects. Therefore, this is the time to begin to communicate those sorts of expectations. The Analytics Problem Statement should have enough understanding of the types of modeling under consideration to set expectations; and it is critical for part of the communication process to the stakeholders at this phase to include examples of the type described above, so that when the need for iteration between analytics and business experts happens, it will not be a surprise or a cause for concern on the part of the sponsoring organization.

One other expectation to set: the postdeployment needs of the model. When the team is thinking about creating a new model, all of the attention is on the present: designing the solution, getting data, building the model, getting the improved business outcomes; but, as we will describe in the last phase of the life cycle, models also need regular maintenance, and this may not be what key stakeholders at the sponsoring organization expect. The Business Problem Statement document is a good place to plant that seed, so that when the focus turns to deployment and postdeployment phases, the stakeholders will have heard this before and are ready to plan for it at that time.

The desired outcome of this task is to receive agreement from the stakeholders in the sponsoring organization that both problem statements are acceptable and will be supported by this stakeholder group moving forward.

8.2.13 CRISP-DM Phases 2 and 3: Data Understanding and Data Preparation

Data have been called the “oil” that both powers and lubricates the analytics engine, so being successful at these phases of the project is very critical to success. This is also part of the analytics project life cycle where, rather than discussing and researching the data with secondary information (e.g., using a data dictionary that describes the data), we actually get our hands on the data, begin to look at it, clean it up and start to discover, and document the relationships that exist between different data items.

The CRISP-DM methodology includes two separate phases dedicated to data. The first phase is called Data Understanding, and the second phase is Data

Preparation. By contrast, the JTA merges all of the activities associated with data into a single Domain called simply, Data. What both of these slightly different methodological structures agree on is that data are a core component of all analytics projects. The CRISP-DM shows this visually (in Figure 8.1) by placing the image of the data in the very center of the diagram, invoking the notion that all of the phases of an analytics project revolve around the data used to create the analytics model and is critical to the success and effectiveness of what is created.

As was mentioned at the introduction of the CRISP-DM, the methodology recognizes the nonlinear characteristic of this type of work, where information obtained at later stages of the process, for example, about how the structure of the data impacts the business situation, may require revisiting some earlier stages of the business understanding activities (and vice versa). The idea that one obtains the data and retreats to his or her office to analyze it has been shown repeatedly, in practice, to be doomed to failure. The understanding of the data requires regular, and often intensive, interaction with business experts to reach the level of data understanding that is required to create a successful and sustainable business analytics solution. In addition, the link to the business needs to continue literally throughout the entire process of planning and implementation of the analytics solution.

We find again that the detail in the JTA for the data domain provides a deeper understanding of both of the CRISP-DM phases of data understanding and preparation. The JTA Data domain does not order the tasks into those that involve data understanding and those that focus on data preparation. To make this discussion simpler, we will discuss the tasks of the JTA in the order they appear in the CAP JTA document, but we will make it clear which aspects of those tasks apply to either understanding or preparation. Frankly, any activity in which we work with data, either directly or indirectly, brings better data understanding. Certainly, some of the tasks focus more on one than the other, for example, Task 3 is called “Harmonize, rescale, clean and share data” is focusing mostly on preparing the data for use.

The tasks included in the JTA Data Domain are as follows:

- Task 1: Identify and prioritize data needs and sources
- Task 2: Acquire data
- Task 3: Harmonize, rescale, clean, and share data
- Task 4: Identify relationships in the data
- Task 5: Document and report findings (e.g., insights, results, business performance)
- Task 6: Refine the business and analytics problem statements

While Chapter 4—The Data—already had provided a great deal of important information about data, in this chapter, we will focus on the process steps and their role in the full life cycle of an analytics project.

8.2.14 JTA Domain III, Task 1: Identify and Prioritize Data Needs and Sources

Before beginning the hands-on part of the data work, the AP pauses one more time to make sure that the data needs are clear and the sources are known and available. This task falls squarely into the CRISP-DM notion of Data Understanding, and the analytics team's emphasis, here, is learning as much as possible from documentation provided by stakeholders as well as interviews and workshops with stakeholders to make sure the data needs are clear and the level of understanding of the data is high. This is yet another task that focuses on the softer skills of communication, interviewing, information acquisition, and business understanding.

The expectation is always that the process of identifying the data that is needed and prioritizing the acquisition process will proceed smoothly and that stakeholders are totally forthcoming with sharing information needed at this point. However, the team should be ready for issues that might arise. One common issue relates to security and privacy, but it is also possible to encounter issues related to who is the keeper of the data, who controls its use (and distribution), and who needs to be involved in approving the process of moving forward. The key stakeholders will be critical in deciding where to put the priorities for data acquisition, which comes next, who to contact (the sources), and when more senior stakeholder involvement may be required.

The output of this task should be a plan, preferably a written document, or possibly a slide presentation or less formal document. The document should lay out exactly which data items are required, where they will come from, what form they will be transferred in, and, if possible, a time frame for accomplishing the data transfer process.

8.2.15 JTA Domain III, Task 2: Acquire Data

This task continues Task 1 toward the process of actually acquiring the data. With a plan obtained from the prior task, this task is the implementation of that plan. The range of experiences in acquiring data is very broad. In some cases, the organization is fully prepared, the appropriate senior management involvement took place, and, when asked to provide the data, it is simply provided—the ideal situation. Another possibility is that the data may be “public data” and the access to data has been set up to be seamless and can be initiated by the analytics team through a known public process—another ideal situation. However, the other end of the spectrum is also possible. It is not uncommon to encounter organizations in which control of data is closely linked to the organization's power structure. So, even if there are senior management approvals, the data owner may not simply jump to provide what is requested. This is one of the places in a project where “the rubber meets the road” in terms of whether the stakeholder involvement is at the

appropriately high or influential enough level. If the stakeholder clout, when carried as far into the organization as is possible, turns out to be insufficient to break through a log jam, it is possible that the project may find its end right at this point. If the stakeholder clout is sufficient, the worst case situation is that a reluctant “data owner” can delay but not stop the sharing of the data that is required. However, experience indicates that the process may be much longer than the newcomer to these situations would expect. Data owner tactics that the analytics team might encounter include the following:

- 1) Seeking additional approvals and invoking a much longer process.
- 2) Seeking to redo the process of justification of the entire project for the data owning group within the organization, again bringing delays.
- 3) Micro managing the specific data items that were requested, hoping to exclude as much as possible from being released.
- 4) Limiting, or extending into the future, the times they are available to meet to actually exchange information about the data request and providing the actual data.
- 5) Setting up complex data usage requirements, for example, that the data can only be accessed on the organization's IT systems, and finding that there are limited physical resources for your team to sit and do this work.

While the prior discussion is clearly unfortunate when it happens, the AP should be heartened in that this “worst-case scenario” is not the norm; but it is always prudent to be prepared for the data acquisition task to take longer than one might imagine and require multiple steps, many meetings, repeated involvement of the stakeholders within the organization that you are working with, and could have possible impacts to schedule and cost of a project. Again, success at this stage relies primarily on soft skills such as communication, persuasion, and negotiation to get the result that is desired.

8.2.16 JTA Domain III, Task 3: Harmonize, Rescale, Clean, and Share Data

Clearing Tasks 1 and 2 of this domain means that you actually have the data you requested and it is time to begin working with it. Nearly every analytics professional has spent endless hours in the process of “cleaning” data, and, as with the acquisition task, the amount of time this can take is both difficult to estimate and almost always longer than you might think. This task falls clearly under the CRISP-DM notion of data preparation, in that we are working to get the data ready to use. It is not reasonable to include all of the types of issues that one might encounter in the process of harmonizing, rescaling, and cleaning data, but the following are some of the most commonly encountered ones:

- *Structure of the Data:* It is typical to get data extracted from a relational database, and therefore in a structure that is efficient for storage and, even,

extraction but may not be structured optimally for analysis. A first step in many data harmonization tasks is to create a “normalized” data structure, typically laid out in a rectangular data structure with rows and columns where rows are the records (cases) and columns are the fields (variables, features). This structure is generally accepted to be the one best suited for analysis, though there are exceptions, such as sparse matrices in an optimization where alternative data structures are actually better.

- *Missing Data:* Real data sets routinely have missing values; sometimes they are marked as such, and sometimes the field is simply empty. In other cases, the lack of a value is “coded,” maybe as 0, when in fact it is missing. The analyst needs to make sure they fully understand how the data were prepared (and how missing values are “coded”) and also must formulate a clear plan to handle missing data (e.g., will a record with missing values be dropped? will missing fields be imputed?). There are many issues that may need to be resolved in this regard.
- *Merging Data Sets:* It is typical that data come from a number of different data sources, and it is also common that the values for a particular data item differ between those sources. The reasons for this are many: the data could be outdated, there could be an error, the data may have been recorded in different units, and so on. In any case, the AP must resolve all of these issues.
- There are many other issues that can arise—many based on very specific knowledge of the data and their use in the business, requiring the depth of business or organizational knowledge that we have discussed often in this chapter. Such issues need to be resolved in close collaboration with the organizational stakeholders.

Successful completion of the cleaning process requires skill in manipulating the data, knowledge of the business, frequent and substantial interaction with key stakeholders (especially those with deep knowledge of the data), and enough time to get it all done. The outcome of this task will be a data set that is much better prepared for the analytical work that is ahead. Though, as with nearly every task in this methodology, it is likely that the analytics team will ALSO obtain greater data and business understanding from the process of cleaning the data, and further, the findings obtained are likely to impact both the business and analytics problem statements. Such data findings should be documented and will be collected (in Task 5) into a full set of data findings.

8.2.17 JTA Domain III, Task 4: Identify Relationships in the Data

With initially clean data in hand, the analyst can turn attention to the modeling and analysis questions that are critical to the success of the project. As discussed earlier (but in the situation when the data was not actually in hand), discovering key drivers and identifying relationships in the data is one of the fundamental

activities that is performed by the AP; and it is squarely in the category of data understanding. The specifics of how to discover the set of relationships depend on the structure of the data and on the proposed model or range of models that are under consideration. Earlier chapters in this book cover a multitude of models, and how they are analyzed or approached, so we will not try to repeat those discussions here. Suffice it to say that the task of finding the key data drivers and understanding data relationships in the data is an important milestone along the analytics project life cycle.

8.2.18 JTA Domain III, Task 5: Document and Report Finding

The data work that has been described, acquisition, cleaning, and data exploration (where we are seeking key internal relationships) represent a great deal of work, and it is a good practice to write down all that transpired in performing those tasks and, then, to communicate those findings to the stakeholder community. The type of communication tool, report, presentation, or meeting depends on the preferences of both the stakeholder community and the analytic professionals involved; but it is critical that the passing of this information happen in one form or another.

8.2.19 JTA Domain III, Task 6: Refine the Business and Analytics Problem Statements

At this phase of the life cycle, the data are acquired, cleaned up, and much better understood. Throughout these tasks, a great deal of data understanding was obtained, and it is likely and expected that this understanding will have an impact on the plan for the project. Therefore, the final task in the domain is to modify both the business and analytics problem statement documents with the updated information, plans, and ideas; and then, present the revised documents to the stakeholder group. The objective of that activity is to seek an updated approval and/or agreement to move forward from the appropriate organizational structure monitoring the analytics project.

8.2.20 CRISP-DM Phase 4: Modeling

Modeling is the phase within the CRISP-DM process that analytics professionals are typically most energized about. It is the place in a project that we are most looking forward to, because this is where we bring to the business situation our “secret sauce.” We are hopeful that with the business and analytics problem statements we have developed and have come to deeply understand, and the data we have acquired and prepared for use, we will create a solution that meets the needs of the sponsoring organization, improves the overall business

performance (following the business metrics set forward in the plan), and turns out to be sustainable for the expected lifetime of the application.

The CAP JTA has two domains that cover the core activities of the modeling phase of the project. Those domains and the tasks encompassed by them are as follows:

- Domain IV: Methodology (Approach) Selection
 - Task 1: Identify available problem solving approaches
 - Task 2: Select software tools
 - Task 3: Test approaches (methods)
 - Task 4: Approaches (methods)
- Domain 5: Model Building
 - Task 1: Identify model structures
 - Task 2: Run and evaluate the models
 - Task 3: Calibrate models and data
 - Task 4: Integrate the models
 - Task 5: Document and communicate findings (including assumptions, limitations, and constraints)

This book contains two very detailed chapters that cover this part of the life cycle in excellent detail; therefore, we refer the reader to those chapters at this time. Chapter 5—Solution Methodology—is focused on JTA Domain IV (Methodology Selection), and Chapter 6 is focused on JTA Domain 5 (Model Building).

8.2.21 CRISP-DM Phase 5: Evaluation

While it is important in any business project to evaluate performance at each stage, it is especially important to focus on evaluation for an analytics project. Analytics projects are driven by data in that we require data to build such models. But, in addition, the output or results of those models are also typically numeric. Of course, each specific model or modeling approach (as discussed in Chapters 5 and 6) is different as to what types of results are produced. A predictive model typically has a resulting error estimate using one of the many standard methods. Classification models are evaluated looking at the percentage of classification errors and other popular measures such as precision, recall, and F-score. Prescriptive models produce “optimal solutions” such as resource allocations or product production plans that can be analyzed by business models that compare the “optimal” solution with a current or other possible solutions. The key is that each model will have a standard set of measures that can be employed for evaluation. As with the prior phase of the CRISP-DM, we will refer the reader to the specifics of Chapters 5 and 6 that cover each of the models and that include clearly defined methods for evaluation for each.

We focus primarily in this life cycle discussion on what the analytics professional does with the evaluation results that are obtained at the end of the modeling process. With respect to alignment with the JTA, it is a little less clean in this situation. Looking at the tasks in the Model Building Domain of the JTA, Task 2 mentions evaluation and Task 4 mentions reporting, both of which are clearly important to evaluation as a phase in the life cycle of a project. However, Domain 5 of the JTA, Deployment, starts off with two similar tasks: These first two tasks are as follows:

- Task 1: Perform business validation of the model
- Task 2: Deliver report with findings

We will discuss both of these important aspects of evaluation in the remainder of this Evaluation phase of the project.

In each of the earlier phases of this life cycle, the documenting of business goals, measured by relevant business metrics, were singled out as critical activities and are important activities in the creation of such documents as the Business and Analytics Problem Statements. Common examples of those metrics are as follows:

- An expected/targeted percent or actual dollar reduction in cost
- An expected/targeted improvement in efficiency
- An expected/targeted increase in revenues (sales) or profits
- And for public sector organizations, an expected increase in coverage, speed of performance, accuracy, or quality of operations

One important point to bring forward at this time is that the evaluation process often produces an ESTIMATE for the performance rather than actual performance. By this we mean, an organization cannot know for sure that a proposed model or “decision,” in the case of an optimization solution, will produce the business benefits promised (e.g., sales, profit, improved quality) until that solution was put into use, data collected over the period of operation of the new approach, and analyzed to see what the actual business improvement was. The evaluation process at the immediate conclusion of the modeling process must, therefore, rely on business models, statistical methods, and other standard methods for assessing performance PRIOR to the actual implementation and use of the model. This is exactly the activity that Task 1 of the Deployment JTA describes as Perform Business Validation of the model. However, when the sponsoring organization decides to move forward with implementation, it is also important to put in place the business infrastructure, methods, tools, and reporting procedures, so the data will be available in a reasonable amount of time, to measure ACTUAL business performance obtained when using the business model, system, or decision that was the result of the analytics modeling activity we are discussing.

We will come back to the notion of evaluation of the models in their operating environments at this critically important later phase in the analytics project life cycle.

After the validation task is completed and the results obtained, it is time to communicate the results. In Task 2 of the Deployment Domain, the description is: Deliver Report with Findings. The results of the evaluation process at the model development phase can have many outcomes. Of course, we all hope that the model developed turns out to be exactly what the sponsor expected, that the results exceed all of the stakeholder expectations, and that the only sensible next step is to move forward with implementation or operation of the model/system. We will pick up the thread of the life cycle when “go forward to implementation” is the decided upon, in the CRISP-DM phase called Deployment.

It is also important now, while we are still discussing the Evaluation phase of the project, to consider what happens if the evaluation produces issues. By “issues” we mean something that requires further consideration, discussion, or analysis on the part of the entire project team, both the analytics team and the sponsoring organization. Such issues, in particular, mean that the project is not ready to move to the next phase. Among the issues that may occur are the following:

- The expected improvement in cost, revenue, profit, or time (efficiency) is below what was expected or required.
- The model takes a much longer time to reach its solutions than expected, and may need continued work to remedy this problem.
- The accuracy, variability, or sensitivity (to model inputs) for the model is such that the answers engender less confidence in the results than the sponsor and the modelers had hoped for.
- There are many other possible issues.

Before moving forward (whether there are serious issues or not), the generally accepted approach is to stop at this point and document the results of the evaluation process to the sponsoring organization. The information to communicate should include findings and recommendations of what should be done next. It is important to make sure that the sponsor continues to see the analytics team as a good partner by being forthcoming, and in particular not withholding or delaying communication, of these intermediate findings at the end of the initial modeling stage. The analytics professional, as the leader of the analytics team, should find the best medium for communication: a report, a presentation, maybe just an agenda for discussion at a meeting, and come together with the sponsor stakeholders to present the findings. If the findings are sufficiently positive, the expected recommendation is to go to the next project phase. If the findings show issues, as already mentioned, we revert back to the structure of the CRISP-DM methodology.

The CRISP-DM methodology includes an important feedback loop that is, for that very reason, located at the Evaluation phase. As seen in Figure 8.1, there is an

arrow from the Evaluation phase to the Business Understanding phase. The message here is take the findings of the Evaluation phase, especially if those findings include any issues, and link back with the sponsoring organization, the business, to reconsider some or all of the assumptions, expectations, business issues, and goals of the project. Projects do not, typically, require scrapping all of the work to that point and starting over. More often the evaluation will point out specific issues, for example, clarifying expectations, getting deeper understanding of some of the data, and formulating alternative business strategies or goals. Here are a few possible outcomes or recommendations typically uncovered:

- Gaps in the data were found, so the team will take on the task of obtaining more or different data (going back to the data phases of the methodology).
- The model did not perform as well as expected, so the team may want to look at alternative modeling approaches (going back to the modeling phase).
- The variability of results has led the key decision-makers in the sponsoring organization to rethink the risk issues underlying the use of the model (going back to the business and data understanding phases).

Of course, there are many other possible outcomes, but the key is that the analytics modeling team and the organizational stakeholders will come together and decide the next stage in the process collaboratively, and then the analytics team will implement that jointly arrived at plan.

Hopefully, after the analytics project team has completed revisiting prior phases of the process, and after completing another round of evaluation, the issues that surfaced in earlier evaluation processes are resolved, that the business performance metrics related to the project are found to be in acceptable ranges for all involved, and the final decision is to move forward to the next project phase of deployment.

8.2.22 CRISP-DM Phase 6: Deployment

This phase of the CRISP-DM process is described simply as deployment, but the JTA has two remaining domains: deployment and life cycle management. Frankly, the linkage between the CRISP-DM methodology and the JTA falls apart somewhat at this point. One reason is that the final domain, model life cycle management, includes essentially everything that is in this chapter, and some of the tasks included there involve topics we have already discussed, such as documentation of the initial model, and other activities include tasks that intended to take place during and after deployment. To simplify the discussion, we will break this section into two parts:

- Activities up to and including delivery of the model (deployment)
- Activities that take place from the time of delivery forward (postdeployment)

In Domains VI and VII, the JTA contains activities that fall into both of those two categories and even some activities that happened earlier in the process. The following bullets contain these two JTA domains:

- Domain VI. Deployment (the ability to deploy the selected model to help solve the business problem)
 - Task 1: Perform business validation of the model
 - Task 2: Deliver report with findings
 - Task 3: Create model, usability, and system requirements for production
 - Task 4: Deliver production model/system*
 - Task 5: Support deployment
- Domain VII. Model life cycle management (the ability to manage the model life cycle to evaluate business benefit of the model over time)
 - Task 1: Document initial structure
 - Task 2: Track model quality
 - Task 3: Recalibrate and maintain the model
 - Task 4: Support training activities
 - Task 5: Evaluate the business benefit of the model over time

The following discussion will refer back to specific tasks in both of these domains, but the discussion will focus simply on before delivery of the model and after.

8.2.23 Deployment of the Analytics Model (Up to Delivery)

Whether the analytics team (led by the analytics professional) is the group responsible for implementation of the model or system they have designed often depends on the size and complexity of the model created, the complexity of the data involved, and whether the analytics model fits into larger or existing IT systems within the sponsoring organization. In cases where the model size is smaller than medium size and when the system would be described as “stand-alone,” the management within the sponsoring organization may decide to have the analytics team assume the responsibility for implementation (building, coding, and delivery) of the model they developed. However, when size, complexity, and organizational interactions are larger (and often no matter what the size), the sponsoring organization may alternatively place the responsibility for implementation to the division or group of the sponsoring organization that is responsible for IT. In that situation, the analytics team would typically move to a support role, rather than the leadership role they had in the design phases.

*Note * here indicates a task in the JTA that is not included in the CAP certification exam questions.

The first activity of deployment is to document production requirements. Since at this phase of the project the model or system is built, fitted, or “learned” from the data, and recently tested, the analytics team should fully understand the production requirements. The goal then is to write them down clearly and completely so that the implementers can move forward with clarity to a successful implementation. Even if implementation is the responsibility of the analytics team, it is important to take the time to create written production requirements as they are also useful for testing and other needs later in the life cycle.

The activity of creating a production requirements document is a standard systems design or system architecture activity that is generally the responsibility of IT professionals to do, but certainly to manage and oversee. That team will make clear what sorts of documents are required, provide a form or template in which to create them, and, often, aid the analytics team in creating the documents so that they are acceptable to the implementation team. One important role in this process is, often, to create a test plan. IT professionals usually set up carefully planned and staged procedures to test whether the software or system developed is working as expected. However, when the system is an analytical model, it often requires someone with deep understanding of that model to create worthwhile tests.

Another typical area of collaboration between analytics teams and IT implementation teams relates to the technical code related to the model. For example, if the model implements an optimization model, the analytics team likely used one of the standard commercial or open-source solvers; and it is likely that the IT professionals are not expert in that software. In those cases, the analytics team may be responsible for coding and testing a module or executable component of the larger system following guidelines and procedures specified by the IT team.

The next activity in the JTA (chronologically) is the Deliver Project Model/System, (Task 4 of Domain VI). You will notice that this task is marked in the JTA as a component that is not included in the specific certification process and, therefore, the analytics professional seeking the CAP certification will not find questions on the certification exam on this topic. This task falls more specifically in the area of IT system implementation, but it is typical that detailed knowledge of the analytics models being deployed will continue to be required by the implementation team doing the coding, testing, and final delivery of the model or system. The JTA does include Task 5 titled “Support deployment,” and we will discuss this task now.

This task is a recognition that the analytics professionals need to be engaged and available as required by the implementation team to solve problems or deal with issues that arise in that activity relating to the analytics model being deployed.

Example of the types of model support that come up in many implementations are as follows:

- The data used in driving the analytics model have errors or other issues that are shutting down or crashing the system.
- The model is taking too long to run or in some other way impeding the operation of the system environment where it is located.
- Interactions with other parts of the larger IT architecture are encountering issues that the IT team cannot diagnose.

Those examples and many others that might occur require knowledge of the data and of the model, and that knowledge resides with the analytics professionals who created the model. There may be a process in place to engage the analytics team when an issue comes up, such as the creation of a “ticket” that the team is required to respond to in a particular manner, or it may be less formal where a manager or member of the implementation team calls or e-mails someone on the analytics team to seek help. In any event, the analytics professionals need to be there to help with implementation all the way through to the final testing and handoff to an operational entity to oversee.

The entire set of skills required for design and building of the model are required for this phase as well. They include all of the modeling and analytical skills it took to create the model, but more importantly they also include all of the communication or softer skills of collaboration and persuasion that an analytics professional needs to be successful in his or her work.

The next topic to discuss is training. The only mention of training in the JTA is in Task 4 of Domain VII (as you see the activities are not mentioned necessarily in the order that they often occur). We are including this as prior to handoff of the system to the operating entity because we see training as a critical and important part of any deployment. The responsibility for training may reside with either the analytics team or the larger IT team, but what is most important here is that the analytical knowledge of using and, more importantly, interpreting and explaining the model outputs will typically reside in the analytics team. Therefore, creating training materials and possibly being the individual to deliver training modules (be the “trainers”), create tests, or other ways to measure how users are faring in using the model are good activities residing with the analytics team.

As with many other aspects of the entire process, successful completion of the training activity requires an appropriate mix of strong analytics input and excellent communication skills. The individuals preparing training materials need to totally understand how the model works and must be able to explain these workings, both in writing and verbally, to the individuals from the sponsoring organization who are responsible for using the model. Further, they should be able to break down the separate aspects of the model into manageable chunks of information so they can be presented in the training activity, intersperse practice scenarios, and provide a method by which a person (if tested) can demonstrate sufficient mastery of the model.

8.2.24 Post-deployment Activities (Domain VI: Model Life Cycle Management)

We come now to the second aspect of deployment: postdelivery monitoring and reporting. Several of the important tasks included here show up in the JTA in domain VII, model life cycle management. It is reasonable for the reader to ask at this point: “Hasn’t this entire chapter been about life cycle management?” And of course that answer is yes. Indeed, it would have made sense to point out the importance of the life cycle right up front in the JTA, but we see that it is mentioned for the first time as the final domain of the JTA. One good explanation is that, as you come to the end of any project and especially an analytics project, it is important to focus attention on the sustainability and continued usefulness of the model that has been created.

Therefore, we move now to the phase of the project AFTER the model is delivered and taken over by the operating organization. Three tasks in particular from Domain VII focus attention to the postdelivery time frame. They include tasks for tracking model quality (performance), recalibrating and maintaining the model, and evaluating the business benefit of the model over time. The second of those (recalibrating and maintaining the model) is marked with an asterisk in the JTA, indicating it is not intended to be included in the certification exam; however, the tasks and skills required to recalibrate and maintain are essentially those same skills that were required to build, test, and deploy the original model. To be consistent with this document being focused on the basic CAP certification process, we will focus on the other two tasks.

Consider first tracking the model performance. At an earlier stage in the life cycle modeling discussion, we mentioned the importance of preparing the stakeholders within the sponsoring organization to the need for model maintenance. Nearly all analytics models require regular maintenance of some sort or another, and that specific maintenance activity depends on what type of model was built. For example, suppose the team delivered a demand forecasting model for a manufacturing or retail entity. It is well known that demand can change over time, and not recalibrating, re-estimating, or redoing such a predictive model may lead to larger and larger errors in forecasts over time. It is important to stay on top of such loss of accuracy or quality in a model quickly as it is not uncommon for users of such a model to blame the concept or process of modeling for the problem rather than the fact that the data or underlying relationships between predictor variables and the quantities being forecasted have changed or evolved over time.

Hopefully, during the earlier communication processes and delivered documents, the sponsoring organization was prepared for the need for maintenance activities and they have planned them into the program on an ongoing basis. A key component of the maintenance process is monitoring and tracking model