The business technology ecosystem is quickly filling with vast quantities of data. IDC’s “State of the Universe” report estimated that the amount of data grew nine-fold between 2006 and 2011, reaching 1.8 trillion gigabytes. However, the problem extends beyond that of raw storage capacity. In fact, analysts have said the number of different files containing information is growing faster than the digital ecosystem itself. This creates a heavy burden for IT staff to manage, particularly as their companies turn attention toward analytics initiatives and business intelligence solutions.

A number of data collection, analysis, and integration tools have emerged to help address this big data trend. The majority of applications come with their own data integration tools, but this doesn’t solve the problem of disparate systems and may create even more work for IT departments. Unified solutions, such as Informatica’s PowerCenter, have emerged to help consolidate data integration (DI) processes and bring much-needed efficiency to the digital ecosystem. Yet, even businesses with a unified DI tool can run into roadblocks that make the process time consuming and difficult.

But there’s hope—these issues can be overcome by centralizing management and job scheduling of IT background processes so that administrators can easily automate and control their systems. However, it is important to understand the challenges that many companies face with DI, so that the appropriate tool can be used to address them.

Data Integration Challenges
Organizations are increasingly facing a growing volume and variety of information. Important assets are no longer neatly packaged in databases, and even those that are could be housed in different systems spread across the corporate IT ecosystem. The U.S. Department of Transportation’s primer on data integration provides a comprehensive list of the DI challenges that have come from this explosion of information.

Heterogeneous Data
Volume may represent one of the big data challenges of modern businesses, but it isn’t the only quality that can make integration difficult. There is also variety. IBM’s analysis of the digital universe touched on this issue, and the reality is that most IT professionals are now expected to deal with a mixture of structured and unstructured data sources. This was confirmed by an Intel survey of 200 IT managers, in which 84 percent of respondents said they were currently analyzing unstructured data. Of the managers who weren’t actively working with unstructured data, 44 percent said they expected to in the next 12 to 18 months.
Currently Analyzing Unstructured Data (n=200)

Q: Will you be asked to process and analyze unstructured data sources in the next 12 to 18 months?

- Yes: 84%
- No: 14%
- Unsure: 2%

Q: Are you currently processing and analyzing unstructured data sources to get new insights?

Among those not currently analyzing unstructured data:

- 44% expect to be asked in the next 12-18 months.
- 12% say they will not be asked to do so.
- 44% are unsure whether they will be asked.

Even as the number of different sources of data expands, expectations regarding time-to-delivery are increasing. Although the survey revealed that batch and real-time delivery methods were used approximately the same amount, respondents predicted they would use real-time delivery close to 66 percent of the time by 2015. As a result of both volume and variety, IT administrators expressed the need for improved data management practices and technology.

**Bad Data**

Data quality can have a widespread impact on business process improvement and compliance management. Accuracy is often the first thing that comes to mind when data quality discussions come up, but there are actually several metrics that factor into whether information is good or bad. The Department of Defense’s Guidelines for Data Quality Management outline the following:

- Accuracy
- Completeness
- Consistency
- Timeliness
- Uniqueness
- Validity

Poor DI practices can have an effect on all of these characteristics. One misstep could affect many of the core metrics used to judge value. For instance, a shutdown of a single DI task will impact both the completeness and timeliness of the information, while relying on manual input may reduce completeness, timeliness, and accuracy. These factors limit the organization’s ability to use its data warehouse. After all, low-quality data is likely to generate poor insight.

**Excessive Costs**

It’s no secret that managing disparate systems is a time consuming task, and the hours of overtime can quickly add up to drive a DI project beyond its allocated budget. If IT professionals are consistently putting more time into DI than expected, it could be a problem with the tools they are using. It’s essential to factor the potential for both planned and unplanned costs into a total cost of ownership evaluation of DI technology.

The cost in software licensing alone can make DI significantly more expensive than it needs to be. In 2008, Gartner estimated that companies could save $250,000 or more per year by consolidating redundant solutions. The savings stem from a variety of factors, including reduced software licensing and management costs. However, relying on disparate systems also increases the chance for unexpected spending to emerge during a DI project.
Gartner noted the value of focusing on simplicity, which allows organizations without a large pool of expertise to use their DI tools more effectively. Consolidation is also likely to make it easier to see where data is stored and how it moves through the company's IT environment.

**Perception That Data Integration Is Overwhelming**

Technology has streamlined processes and lifted a great deal of burden by facilitating automation, but it's important to remember the human element. Negative perceptions may not be as easy to track as hours clocked or help tickets addressed, but they can still affect productivity. Seeing DI as a single, overwhelming task also makes it difficult to plan jobs and effectively manage workflows.

It is important to set realistic goals and identify specific areas for improvement. If DI is taking an excessive amount of time, it is likely due to one of the following operational issues:

- Developers often have to write custom code to integrate disparate platforms
- The business is using many different solutions
- IT must rely on manual processes either for DI or delivering business intelligence reports
- The existing DI tool is too complex to use efficiently

If the company is facing the majority of these problems, then it may be worth changing the data integration paradigm. An operational shift is also likely to necessitate new technology investment to deliver the functionality needed to improve visibility and increase efficiency. DI tools come in all shapes and sizes, so it is helpful to go into the planning phase with a baseline set of features to look for.

**Critical DI Software Features**

Most DI tools come with some form of job scheduler, but the devil is in the details. Asking questions such as whether job scheduling software supports cross-platform job tasks and centralizes the management of numerous systems can save a significant amount of time and money. A report from Bloor Research explored the issue of DI by breaking projects down by the vendor used.

Bloor's analysts noted that Informatica PowerCenter® users rely mostly on a single vendor for their DI needs while those using tools from other providers often utilized multiple, disparate products. The average number of person-weeks that PowerCenter users spent on DI projects ranged between 11 and 17. The variance was explained by the fact that certain resource migrations were more complex than others, with SaaS projects taking the most time.

One of PowerCenter's features that researchers highlighted was its reusability across multiple scenarios. This had a significant impact in determining the overall value of the solution, and in maintaining a relatively low cost per project per end point of $2,080.

For those considering investing in a new DI tool, it may be worth taking stock of the company's existing IT ecosystem. For example, businesses that frequently deal with complex sets of information from disparate sources would need to place a premium on robust functionality to streamline the DI process and provide greater visibility over the operation. Going through this process in the planning phase can help to identify the must-haves for the business.
For most organizations, centralized logging and visibility will likely be critical considerations in their DI tool selection. This can result in a significant amount of savings. In its analysis of the DI vendor landscape, InfoTech Research identified several key considerations for evaluating these tools:

- Real-time integration
- Data cleansing functionality
- Post-failure integration recovery
- Performance monitoring
- Middleware capability
- Data semantics
- Synchronization

Using clear frameworks to evaluate existing solutions will likely produce advantages for a number of businesses. InfoTech's data also revealed a significant amount of dissatisfaction with the functionality of existing DI solutions with 40 percent of respondents saying their tools aren't robust enough.

This doesn't necessarily mean that IT leaders should trade in their existing DI software for an entirely new platform. This would likely create additional challenges, given that employees already familiar with a tool may be reluctant to start using something completely new. While a thorough investigation of already implemented solutions and practices is likely to reveal numerous areas for improvement, many of the common frustrations associated with DI can be overcome by better incorporating DI into the rest of IT operations.

**Why Consider Enterprise Job Scheduling Software?**

Built-in DI tools work for the application they are designed for, but integration initiatives are frequently dependent on more than a single application. Informatica PowerCenter comes with robust functionality and can address a large range of business needs, but IT professionals are still struggling with their integration efforts. The problem is that DI tasks frequently depend on external applications, or system files events. It's simply not enough to just bring large volumes of data together—someone wants to make use of the information, and employees need data in a comprehensible format to do that. As a result, the overall DI process is often the starting point for additional tasks that depend on timeliness. Business intelligence and DI, for example, are linked because the best insights are frequently gleaned by connecting data that comes from a wide range of sources.

Similarly, relying solely on a DI tool for both the process and reporting leaves a lot of room for error and delays. There are several benefits to utilizing an enterprise job scheduler, including:

- Centralized visibility and management over IT operations
- More exceptions to scheduled tasks
- Better workload management

Although the business technology world is increasingly moving toward automation, and achieving impressive efficiency gains as a result, it's important to remember the people who ultimately need to use the information brought together by a DI effort. Data reporting needs to run on a real-world schedule, which is usually complicated to work into a DI tool. An enterprise job scheduler is better designed to incorporate exceptions based on events like holidays or stagger tasks for different time zones.
It’s also important to consider functionality for event-based scheduling. For instance, if a report is dependent on a particular file that hasn’t arrived yet, it wouldn’t make sense to run an integration. Using a more comprehensive job scheduling tool would allow IT to set the integration task to start when the file arrives.

Centralizing job management in this way reduces the risk of overtaxing hardware resources. The common pain point of DI being too resource-intensive likely stems from the fact that most DI tools simply don't provide an easy way to schedule around what else is happening on the system. Bringing DI tasks together with other IT jobs creates an environment for easier workflow planning as it becomes much clearer when there will be little running also on the system. Although all of this is likely to make life easier for the IT department, the business gains the benefit of its technical staff being able to focus on other tasks and completing DI projects more efficiently. In addition, a job scheduling tool will also make compliance less of a hassle.

**Job Scheduling: A Critical Part Of The Toolset**

The push for real-time analytics and data-driven insight has created a sense of urgency for integration. Business users increasingly expect to be equipped with timely, accurate, and relevant information that can be used to improve the organization. Because data relevance is now so heavily dependent on the speed at which it is delivered, it is increasingly important to have dynamic control over the start and continuation of DI tasks. For example, an IT administrator may need another application to begin immediately following a PowerCenter workflow. Relying on a DI platform alone, IT would likely need to use manual input, resulting in potential delays, or write custom scripts to facilitate the process.

While there is no substitute for a robust DI tool like PowerCenter, enterprise job schedulers such as Automate Schedule™ have become a critical component of the DI toolset. Centralizing and integrating workflow schedules into the rest of IT operations not only lessens the burden on IT staff, but significantly mitigates the risk of error while improving time-to-delivery. Automate’s advanced job scheduler offers the advantage of already including support for Informatica PowerCenter as well as numerous Windows applications, SAP® and Oracle® E-Business Suite NetWeaver. This means it is easy to incorporate the job scheduling software into many IT ecosystems, so that businesses can improve their DI processes sooner rather than later.