



# 25

## CRITICAL CHECKPOINTS FOR PAIN-FREE

# ETL

Key checkpoints, gleaned from  
700+ man-years' experience,  
that can help your ETL success.

If you're looking for ways to reduce the pain of your ETL projects, you'll find these 25 critical checkpoints can help. Data Junction's 700+ man-years' experience in integration is the source for these checkpoints. Benefit from others' hard-won wisdom and make these checkpoints part of your next integration proof-of-concept or RFP.

Expedite ETL projects by knowing when to use intermediate data formats.

*Checkpoints:*

- 1** Does your project involve legacy data or unsupported applications or data formats?
- 2** Are the original architects no longer available?
- 3** Is there poor documentation or no information about the architecture?

If you answered “yes” to any of these checkpoints, you'll likely expedite your project by exporting data to an intermediate data format. It may in fact save you considerable time and resources to use intermediate data formats, rather than attempt to force a more “streamlined” project flow.

Enhance web and/or net enablement involving ETL processes by assessing needs for recoverability versus accessibility.

*Checkpoints:*

- 4 Are you using FTP in web or net ETL processes?
- 5 Are you evaluating whether to extract and load to HTTP or SMTP?
- 6 Have you considered whether you need URI support?

If you answered “yes” to any of these checkpoints, here are some tips. If you’re using FTP in a web or net process, you’ll likely save time and resources by considering a switch to HTTP or SMTP. When evaluating which to use, bear in mind that HTTP may offer better remote and end-user accessibility, while SMTP can offer faster recoverability. URI support in an ETL tool can enhance your web or net-enabled ETL processes by providing alternatives like these to fit your needs.

Reduce future ETL pain by “mapping” the variety of proprietary and open systems and formats that may touch your project.

*Checkpoints:*

- 7 Are proprietary systems and formats predominant?
- 8 Are systems as well as data formats heterogeneous?
- 9 What other business units (and thus what other systems and formats) may be added after successful deployment of your ETL project?

In the rush to deploy ETL projects, it's critical to take time to "map touch-points". Heterogeneous or proprietary systems can complicate even simple ETL projects. Though you may be unable to plan for an unforeseeable future, using modular tools and solutions can reduce future ETL pain by enabling more nimble approaches.

## Speed time-to-market by avoiding these ETL speed traps.

### *Checkpoints:*

- 10** Are there other implementations and/or applications that may be introduced during your ETL project deployment?
- 11** Are the data schema and structure clearly defined, and are they static or dynamic?

It's common for time-to-market pressure to drive ETL projects into hidden "speed traps" such as the above instances. Clearly define the target database data and metadata models and the load phase of your ETL project should go faster, preserving your efforts to speed the extract and transformation phases. A simple survey of implementations and applications that may be introduced can also optimize your time-to-market speed.

**Download your  
FREE trial today!**

[www.datajunction.com](http://www.datajunction.com)



## Plan as if ETL project scalability and reusability were linked.

### *Checkpoints:*

- 12** Will the ETL process execute on “live” databases?
- 13** Have you defined what scalability means for your ETL project – desktops vs. servers, ETL projects expanding to run across business units or divisions, web or net enablement, etc.?
- 14** Have you defined what reusability means for your ETL project – metadata exchanges, central map repositories, planning for global changes that may affect business rules, etc.?

Reusability and scalability impact each other in ways that can affect the success of your ETL project. Consider them linked while planning. For example, a change in scalability – adding a remote business division needing web enablement for live processes – may drive changes in planning for central map repositories. Plan intelligently and reduce your ETL pain.

## Automation can enhance ETL efficiency – but make sure your proof-of-concept includes as full a test of automation as possible.

### *Checkpoints:*

- 15** Do you need to buy a scheduler, or will a scheduler in an OS suffice?
- 16** Do you have a true simulation prepared for automation testing, rather than a test based on sample data?

Too often ETL automation is tested on partial or sample data. Make sure automation will meet business needs by developing a thorough, full assessment for your proof-of-concept.

Different kinds of ETL flexibility give you a variety of ways to plan for reduced ETL pain.

*Checkpoints:*

**17** Flexibility within the transformation itself: can it accept changes in data structure, can it easily integrate with new applications, can it be configured differently for disparate use environments?

**18** Flexibility of the development team: can they adapt to the changing needs of end users, can they quickly deploy ETL projects to accommodate changes in business needs?

**19** Flexibility of the end users: what are their skill sets and business needs?

By building your RFP in consideration of the different kinds of ETL flexibility noted above, you'll find ways to reduce pain specific to the conditions affecting your ETL project.

Improve ETL performance by evaluating user ramp-up time as well as performance benchmarks.

*Checkpoints:*

**20** Will potential vendors develop realistic performance benchmarks based on your project conditions?

**21** Do benchmarks include transactions per thread (if needed) as well as transactions per second?

**22** Have you considered training and ramp-up time for users when assessing ETL performance?



Benchmarks are best developed in the context of a thorough proof-of-concept, rather than in an abstract performance environment that doesn't reflect your ETL project parameters. Remember to factor in user ramp-up time – fast throughput can be deceptive if the time-to-deploy and mean-time-to-repair are painfully slow.

## Evaluate recoverability when planning for reduced ETL project pain.

### Checkpoints:

- 23** Have potential vendors discussed and developed recoverability as part of your proof-of-concept?
- 24** Does your ETL project need transactional controls over the databases and the processes to enhance recoverability?
- 25** If you're using MOMs, will any built-in recoverability need to be augmented?

Message-Oriented Middleware (MOMs) usually have built-in recoverability and transaction controls. Evaluate both built-in and added-on recoverability as part of a thorough proof-of-concept for reduced ETL pain.



*You don't have to suffer painful ETL — contact Data Junction for a proof-of-concept or RFP today!*

Analyst groups note that corporations spend billions of dollars writing custom code to support ETL projects. Custom coded solutions may give you the illusion of control, but the reality is they leave you vulnerable in many ways.

- What if the architects leave or do a poor job of recording their work?
- What if typically tightly-coupled solutions can't fit new needs?
- What if your time-to-market is accelerated or your resources are cut?

Reduce the pain on your next ETL project with help from [Data Junction Corporation](#). Our award-winning, widely-adopted products and solutions have helped IT professionals quickly deploy projects when other options would take too much time or cost too much money.

If you have upcoming or ongoing projects involving the following, we have a proven record of success and experience that can help you.

- [ETL or legacy data migration](#)
- [e-business integration](#)
- [XML integration](#)  
(EDI to XML, SOAP, and similar)
- [Data marts](#)
- [Data quality or data cleansing](#)
- [Much more](#)

Please [contact us](#) today for a proof-of-concept or RFP and prove it to yourself – you can reduce the pain of your ETL projects. Download a [FREE trial](#) if you'd like to start evaluating immediately.



Integrating the  
Interconnected  
World®

2201 Northland Drive  
Austin TX 78756

800-580-4411

[www.datajunction.com](http://www.datajunction.com)