

S Q L

P O W E R

Utilizing ETL Tools

Can you afford not to?

S Q L
P O W E R

- The Data Integration development backlog
- Why is custom-coding a non-option?
- What is an ETL Tool?
- Benefits of utilizing an ETL Tool
- What should I look for in an ETL Tool?
- What's out there?
- Are they affordable?
- Cost justifying the purchase of an ETL Tool
- How to negotiate with ETL tool vendors?
- Summary

Company Overview

S Q L
P O W E R

Who is SQL Power?

Toronto based software firm specializing in Data Warehousing & Performance Management (since 1988)

Products

A complete suite of DW & Performance Management tools

- Web based KPI Tracking tool – the **Power*Dashboard**
- Web based Balanced Scorecard tool – the **Power*Scorecard**
- Data Migration (ETL) tool – the **Power*Loader**
- Data Cleansing/De-duping tool – the **Power*MatchMaker**
- Data Aggregation tool – the **Power*Summarizer**

Professional Services

Business Intelligence and Performance Measurement Consulting:

- High Level Management Consulting
- Project Scoping
- Architecture, Design, Development and Implementation.
- Decision support consulting
(turning data into timely information)

ETL Tools .. Can we afford them?

S Q L
P O W E R

The Development Backlog

Since the Business Focus is constantly changing, and along with it the information required to support the decision making process, IT seems to always be scrambling trying to meet the growing demand:

- Additional work, need to hire more full-time staff
- Can't find good full-time staff, so they hire and train consultants
- Can't afford indefinite consulting engagements, need to replace consultants with full-time staff and knowledge transfer
- Highly skilled full-timer just got a better offer, and is walking out the door along with the specific program knowledge
- Hiring freeze, can't replace departed full-timer, need consultant
- Custom-coding of interfaces by high-priced consultants usually results in missed dead-lines and cost-overruns
- Just bought a new company, interface work-load just tripled due to required one-time conversions, data integration and the need to build DW

It's an all too familiar story played out in most North American organizations pegging Business Users against IT's ability to deliver timely/quality data!

ETL Tools .. Can we afford them?

S Q L
P O W E R

The Development Backlog

The rate of increase in demand for new information has rendered IT's best planning efforts insufficient and futile – IT needs to opt out of vicious cycle of : needed it yesterday, will try to give it to you next quarter

- Need to lessen dependency on highly-skilled developers and consultants
- Need to simplify process of interface building and assoc. maintenance
- Need to reduce the skill-set required to deliver interfaces
- Need to deliver interfaces more efficiently (with < 30% of the effort)
- Need to standardize interface development, so that all development resources are interchangeable
- Need to make the development environment more interesting and more state-of-the-art for full-time employees (with valued business/data knowledge)
- Need to restore IT's credibility to deliver data integration projects (on time and under budget)
- Need some elbow-room (buffer) in data integration project plans

Need to find a more efficient way to deliver IT's data integration tasks!

Custom-Coding has been rendered a non-option for most organizations, here are some of the reasons:

- Custom-Coding of interfaces is analogous with re-inventing the wheel, and takes 3 – 5 times longer than using a productivity (ETL) tool
- Difficult to enforce standards and a standard development style
- Custom-coded interfaces are too difficult to read – especially for non-programmers (managers and architects)
- Too difficult to maintain other people's interfaces
- Tight-deadlines and cost over-runs tend to force developers to cut corners sacrificing:
 - Development standards
 - Exhaustive exception handling
 - Creation of Log files, capturing of run-time statistics
 - Testing and Documentation
 - Tuning

All of the above has resulted in organizations turning to ETL tools for help!

S Q L

P O W E R

An ETL tool is a Meta Data driven productivity tool that automates and simplifies every aspect of:

- Data **E**xtraction
- Data **T**ransformation
- Data **L**oading

ETL Tools .. Can we afford them?

S Q L
P O W E R

When would I use an ETL tool ?

Typically ETL tools are used to:

- Migrate data from a legacy system to a new operational system
- Build interfaces between two operational systems
- Migrate data from operational systems to a Data Warehouse or Data Mart

In a nutshell ETL tools are used to move data from any system/database to any other system and/or database.

S Q L
P O W E R

You would use an ETL tool to:

- Simplify the process of data migration
- Standardize the method of data migration
- Store all data transformation logic/rules as Meta data
- Enable users, managers and architects to understand, review, and modify the various interfaces
- Reduce cost and effort associated with building interfaces (custom coding in conventional languages could cost 3 - 5 times as much)
- Expedite the development of interfaces
- Help meet the increasing demand for data integration in a timely fashion
- Self-document the various data migration/integration processes

S Q L
P O W E R

Other useful features of good ETL tools:

- Simplify and automate the process of data aggregation
- Simplify and automate the process of address correction
- Simplify and automate the process of data cleansing and de-duping
- Standardize the method of data cleansing and de-duping
- Store all data aggregation/cleansing logic/rules as Meta data
- Reduce cost and effort associated with data cleansing, address correction, duplicate removal, and data aggregation
- Enable users, managers and architects to understand, review, modify, and validate the various data cleansing and duplicate removal steps
- Help meet the increasing demand for data cleansing, aggregation and duplicate removal
- Self-document the various data cleansing efforts

Since ETL tools come in all sorts of shapes, sizes and prices, here are the main things to look for:

- Engine based, not a source code generator
- Portability, ETL tool should:
 - run on all Windows and Unix environments
 - Be able to read/write to all your database platforms
- Should be able to group interfaces into logical units of work
- Should generate log, error and bad files and capture run statistics
- Should be able import/export, backup/recover Meta Data
- Should be able to migrate Meta Data releases from Dev -> Test -> Prod
- Should provide ability to aggregate data
- Should have User/Group, System/Object level security
- Should be able to extract/load all required data within available window
- Ease of use – ETL tool should be learned in < 3 days by non-techies

Other nice to have features:

- Should be able to identify/generate deltas
- Should provide ability to create/alter target table structures
- Should provide ability to Debug and Tune interfaces
- Should provide data cleansing capabilities (incl. Address correction)
- Should provide duplicate identification/removal capabilities
- Should have a scheduling component
- Should have Web-based Meta Data Reports

Although there are over 100 ETL tools on the market, there are only a dozen or so serious contenders, including:

- Ab Initio
- Acta (Business Objects)
- Computer Associates
- DataStage (Ascential)
- Data Mirror
- Decision Stream (Cognos)
- Embarcadero
- ETI
- Genio (Hummingbird)
- Informatica
- Sagent
- Power*Loader (SQL Power)

ETL Tools .. Can we afford them?

S Q L
P O W E R

*Are ETL Tools
affordable?*

Full-featured ETL tool prices are generally all over the map, they range from approx. \$50,000 to \$750,000 CDN. In a recent RFP process for the Alberta government, the following ETL tools prices were submitted (for the exact same environment and licensing terms)

Needless to say there is tons of room for negotiation!

Cost Justifying an ETL Tool

How do I cost justify the purchase of an ETL Tool ?

S Q L
POWER

In order to justify the cost, we must first quantify the savings. Assuming that the monthly cost of a development resource is \$12,000/Month or \$3,000/week (blended rate of \$8,000/month for Employees and \$16,000/month for Consultants) :

Example 1 – Subscriber feed conversion at a Telecom customer

- Initial development of Subscriber interface loading basic subscriber info into the Data Warehouse took approx. 6 weeks
- Of those 6 weeks, 1 week was spent gathering requirements and designing the feed while the other 5 weeks were spent on developing and testing the feed using C and Pro*C at a cost of 5 wks x \$3,000/wk or \$15,000
- The same subscriber feed was re-developed and tested using the Power*Loader in 2 Days at a cost of 0.4 wks x \$3,000/wk or \$1,200
- For a saving of **\$13,800** (or 92%) and **23 person/days** for a single interface

Example 2 – Conversion of entire Data Warehouse at a Pharmaceutical client

- Initial design and development of the Data Warehouse took approx. 24 months (4 dedicated full-time resources)
- Of the 24 months, 6 months were spent gathering requirements and designing the required 120 interfaces; while the rest of the time (approx. 18 months) was spent developing and testing those 120 interfaces at a cost of 4 resources x 18 months x \$12,000 or **\$864,000**
- In May 1999, Roche purchased the Power*Loader Suite and embarked on converting the Data Warehouse (and the 120 Interfaces) to utilize Oracle and the Power*Loader Suite
- The entire Data Warehouse and the 120 interfaces were converted by the same 4 resources using the Power*Loader Suite in less than 4 months at a cost of 4 resources x 4 months x \$12,000 or **\$192,000**
- For a total saving of **\$ 672,000** (or 78%) and **56 person/mths** of effort

Cost justifying the purchase of an ETL tool is easier than you think :

In the previous examples we've demonstrated that utilizing our ETL tool can save you anywhere between 70% and 90% of the development effort, but erring on the conservative side, let's assume that the ETL tool will only save 50% of the effort.

On a given 6 months of a project

- Assume that your company is undertaking a 6 months data integration project (with a team of 4 resources) - the project requires 2 months for analysis/design and 4 months for development and testing.
- Given that an ETL tool saves 50% of the development effort, then utilizing an ETL tool will save half of the development time (or 2 months of effort) **for a cost saving of** (4 resources x 2.0 month x \$ 12,000) or **\$96,000** for just one project in just 6 months.

A good ETL tool will pay for itself in the first 6 months!

S Q L

P O W E R

The increased demand for new data, and the reduced cost of ownership for ETL tools, has rendered custom-coding a non-option in the 21st century!

In this very competitive ETL tool market, there are huge savings to be had, if we negotiate aggressively with the ETL tool vendors – here are some insider tips!

- Always make sure that the ETL tool vendor knows that you are price conscious and that you are talking to other vendors
- If you are spending a large sum of money with a B.I. Vendor – they may be willing to throw in their ETL tool
- ETL Vendors can usually be talked into providing you with:
 - unlimited development seats
 - free development and test licenses (irrespective of the # of CPU(s))
 - free on-site installation

If your ETL vendor is not willing to heavily discount their product, move on to the next vendor on your list- you'll get a virtually identical product for less!

What makes SQL Power the ideal BI partner:

- Provide our clients with custom solutions not just products
- Provide Data Warehousing and Business Intelligence experts at affordable prices
- Supplement our clients' software investment with our own DW productivity tools, thus reducing the time, cost and effort required to successfully deliver BI projects
- Exclusive TDWI (official DW) trainer in Ontario
- Offer unique loyalty program that rewards our clients for their patronage with consulting and training credits

We have the products, expertise and local presence required to successfully execute your Business Intelligence and Performance Measurement plans!

S Q L
P O W E R

*We look forward to the
opportunity to work
with you!*

SQL Power Contact Info ...

SQL Power Contact Info:

SQL Power Group Inc.

5650 Yonge St., Suite # 1204

Toronto, Ontario, M2M 4G3

Tel #: (416) 221-4220

Fax #: (416) 221-5898

www.sqlpower.ca