

Proof of Concept for ETL Vendors

Informatica Response

Course: Evaluating ETL Tools and Technologies, afternoon session
ETL Vendors in Action

Table of Contents

Proof of Concept Overview	2
Scenario Overview	2
1. Product Architecture Overview and ETL Perspective.....	4
2. Extract Scenario 1: Customer Dimension Incremental Extract	7
3. Extract Scenario 2: Shipments Fact Table Extract	8
4. Extraction Scenario 3: Open Case	9
5. Extraction Scenario 4: Time Dimension.....	10
6. Maintenance Features	11
7. Operations and Deployment	12
8. Pricing	13
9. Performance Features.....	15

Proof of Concept Overview

The scenarios for the proof of concept are all based on a wholesale business that supplies specialty products to retailers. The scenarios are based on the items that one might consider important when evaluating an ETL solution for a single data warehouse.

The examples are all built around a wholesale shipments schema, with a set of source tables loaded with data and a set of target tables to be populated by the tools. The extract rules for the schema are simple, but should be enough to demonstrate basic and some advanced capabilities in the products.

The afternoon will be a mix of discussion and demo, with the emphasis on showing how the products are used to accomplish specific tasks. While the focus is on extraction, some of the scenarios or presentation topics involve showing other features like metadata management, data profiling or monitoring job execution.

Because there's no way to show the entire set of ETL for three vendors in the time allotted, we'll be using different elements to show different features. For the scenarios listed we expect to see the features used to accomplish the task live. It isn't expected that we can see the entire extract constructed for each scenario in the time given. However, a complete set of extracts is required in order to show how dependencies, scheduling and monitoring work.

Demo time is limited so there are topics/scenarios labeled "time permitted" which we may not be able to show. They are included in case we have extra time at the end of the class.

Scenario Overview

In a proof of concept you provide to vendors all the source and target table definitions, extract rules and source data. Since this is meant to reflect the real ETL you'll be doing, it's a good idea to select both simple extracts and complex extracts or extracts that have problem data. When you provide this information, it should be formally documented so the vendor understands in detail what they are supposed to show.

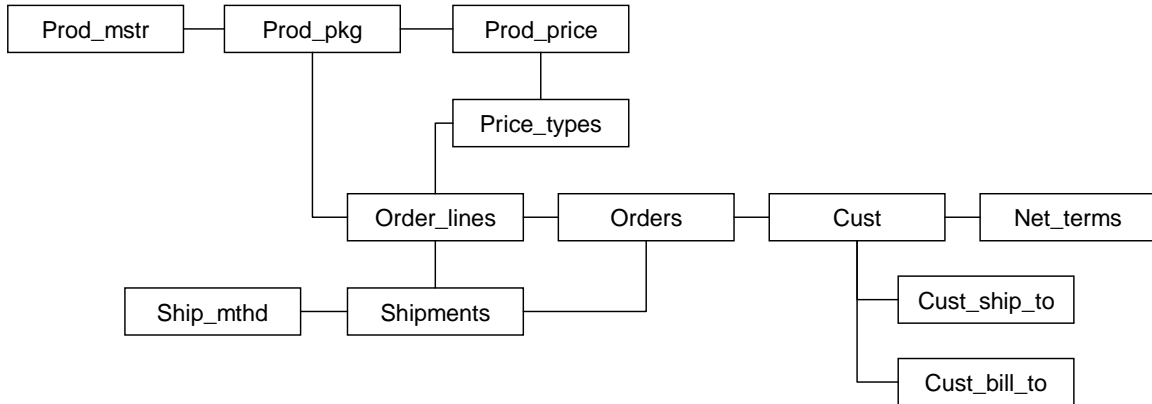
Part of the reason for selecting source data with quality problems is that this will show how a developer is expected to work within the tool. If all the extracts are based on ideal tables and data, as with standard vendor demos, then you won't see what a developer really has to face when dealing with data exceptions.

As a rule, you should have imperfect data, tables with relationship problems like different types on join or lookup columns, and you should always require the use of relational database in the proof of concept.

Using a database is important because it will show you how the tool interacts with a database. Many vendor demos you see are based on text files as input and output, which

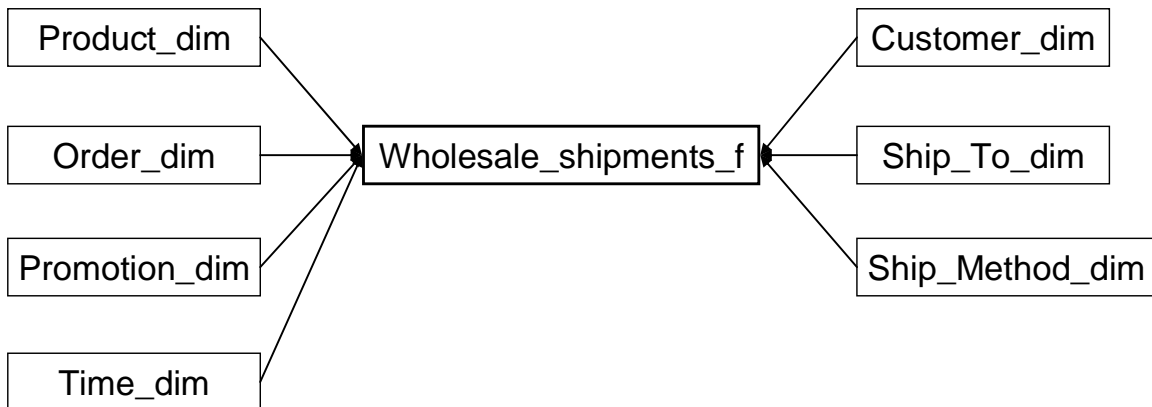
can avoid some of the difficulties when dealing with SQL since all the work is done directly in the ETL engine.

For our scenarios, we will be using the following source and target schemas. The source schema consists of 12 tables with some of the common design problems found in source databases.



Among the problems are a mix of second and third normal form, multi-part keys, invalid foreign key constraints, and multiple join paths to the same data elements. In addition to the above tables there are two change-data-capture tables for the shipment and customer data. These are used for incremental extract examples.

The target schema has seven dimension tables and one fact table. The fact table contains all shipments of order lines by day to customer addresses in the ship_to dimension.



There are several things in this target schema that complicate the extracts. The time dimension is based on a fiscal calendar using 4 and 5 week periods rather than a simple date-based dimension. There is no source for this data, so it must be constructed by the ETL job. A non-data-driven extract is a challenge for some ETL products. The ship_method dimension has unique rows based on a multi-part key which can cause trouble for some ETL tools' lookup functions. The specific details about extract rules and data are available at the end of this document.

1. Product Architecture Overview and ETL Perspective

This is a short (5 minutes or less) presentation about the company, its products and its view of the ETL and data integration market. You should expect to hear answers to the following types of questions:

- The product name or components used for this demo, whether all components are included in the base product and / or how it's packaged.
- What is the core ETL product offering?

Informatica's Product Platform spans 3 distinct but tightly integrated product lines, Informatica PowerCenter®, Informatica PowerExchange™, and Informatica Data Quality. PowerCenter and PowerExchange will be used throughout the demo scenarios.

PowerCenter is available in two editions and serves as the underlying infrastructure used to develop, deploy and manage data integration solutions. PowerCenter packaging is modular in nature where customers start with a base edition and add-on additional capabilities through options to support their data integration requirements.

PowerExchange provides access to virtually all electronic data, whatever format, wherever it resides.

This includes the following formats:

- Legacy – mainframe and midrange data systems/files, databases
- ERP – Oracle, PeopleSoft, Siebel, SAP, SAS
- Standards – Web Services, ODBC, LDAP, POP3, IMAP, HTTP
- Messaging – IBM MQ Series, JMS, MSMQ, Tibco, WebMethods
- Semi-structured – XML, industry standards such as HIPAA, ACORD, SWIFT
- Complex unstructured – PDF, Microsoft Office, HL7, email, and other standards

PowerExchange packaging is based on data/source type and in some cases the capacity of the host data system.

- What optional components are provided?

Informatica PowerCenter Options extend the enterprise data integration platform's core capabilities. These options are available with either PowerCenter Standard Edition or PowerCenter Advanced Edition.

Data Cleanse and Match Option provides powerful, integrated cleansing and matching capabilities. Corrects and removes duplicate customer data to maximize the value of an organization's information assets.

Data Federation Option provides virtual data federation services or Enterprise Information Integration (EII) capabilities. Combining traditional physical and virtual data integration approaches in a single platform, this option creates a powerful tool for delivering holistic data quickly, easily, and cost-effectively.

Data Profiling Option provides comprehensive, accurate information about the content, quality, and structure of data in virtually any operational system.

Enterprise Grid Option provides scalability within a grid computing environment. This option reduces the administrative overhead of supporting a grid. It also delivers optimal performance by automatically load balancing in response to runtime changes in data volumes or node utilization rates.

High Availability Option provides high availability and seamless failover and recovery of all PowerCenter components. This option minimizes service interruption in the event of a hardware and/or software outage and reduces costs associated with data downtime.

Mapping Generation Option provides the ability to automatically generate PowerCenter data integration mappings from best practice templates, as well as the ability to reverse-engineer existing mappings into reusable template form. This option increases developer productivity, reduces time-to-results, and simplifies the data integration lifecycle.

Metadata Exchange Options coordinate technical and business metadata from data modeling tools, business intelligence tools, source and target database catalogs, and PowerCenter repositories. This family of options helps organizations leverage the time and effort already invested in defining data structures.

Partitioning Option executes optimal parallel sessions by dividing data processing into subsets which run in parallel and are spread among available CPUs in a multi-processor system. This option helps organizations maximize their technology investments by enabling hardware and software to jointly scale to handle large volumes of data and users.

Pushdown Optimization Option enables data transformation processing, where appropriate, to be “pushed down” into any relational database. This option saves hardware costs by making better use of existing database assets and helps organizations cost-effectively scale to meet increased data demand.

Real Time Option extends PowerCenter's capabilities to transform and process operational data in real time. This option helps organizations integrate real-time operational data with historical information stored in enterprise data warehouses, powering business processes and accelerating decision-making.

Unstructured Data Option expands PowerCenter's data access capabilities to include unstructured and semi-structured data formats. This option provides organizations with

virtually unlimited access to all enterprise data formats, creating a powerful tool to help organizations achieve a holistic view of data, increase IT productivity, and achieve regulatory compliance.

- How are components / editions bundled?

PowerCenter is available for initial purchase through two editions.

PowerCenter Standard Edition (SE) is a single, unified enterprise data integration platform that consists of a high-performance, highly available, and secure data server, a global metadata infrastructure, and GUI-based development and administration tools.

PowerCenter Advanced Edition (AE) expands the breadth of PowerCenter SE with powerful metadata analysis, team-based development, and Web-based reporting capabilities that are ideally suited for the development of Integration Competency Centers and/or the broader use of PowerCenter across a wide spectrum of data integration initiatives.

- What requirements are there for server installs, client installs, etc.?

Server system with sufficient disk space and memory (1G, 2G recommended minimum)
Windows-based client system with sufficient disk space and memory (1G, 2G recommended minimum)

- What external dependencies are there? For example, a license for a database to be used as a metadata repository.

Installation requires an RDBMS user id and connection to a wide variety of database types (DB2, Oracle, Sybase, SQLServer...).

- What do you consider your sweet spot in applications or in industries?

Informatica's customers span all industries and are increasingly using Informatica technology for "Broader Data Integration" in areas including Data Warehousing, Data Migration, Master Data Management, Data Synchronization, and Data Quality.

Informatica has specific expertise with customer/organizational deployments in the Financial Services, Banking, Insurance, Health Care & Life Sciences, High Tech, Government & Public Sector, Telecommunications, Manufacturing, Retail & Services, Transportation & Distribution, and Energy & Utilities.

- What are your product's strongest points?
 - Open, platform neutral architecture
 - Proven enterprise scalability and performance
 - Universal data access from a single vendor including patented CDC and Unstructured types

- Metadata-driven separation of design and run-time enabling multi-project re-use and adaptability
 - Role-based tools and access orientation supporting the entire Data Integration lifecycle
 - Proven customer success
 - Support for any required data latency (real-time, change-only, batch, federated)
- Why would a customer choose your product over others?
 - Acknowledged market leadership through proven customer success
 - A singular focus on Data Integration
 - Open, platform neutral architecture
 - Enterprise scalability and performance
 - Universal data access
 - Partner ecosystem and the availability of certified Informatica practitioners

2. Extract Scenario 1: Customer Dimension Incremental Extract

This scenario demonstrates the extract logic and facilities for a slowly changing dimension where history is preserved in the rows using start and end dates.

Scenario

Change rows are stored in a change data capture table. This is a table that mirrors the source table, but has additional columns that indicate whether a row is an insert, update or delete, and a timestamp for the activity. This is what you might see as the result of change replication using database replication facilities. The expectation is that the change table provides the activity, while the source table shows the current state of the data.

To properly test the extract, there are four rows for three different conditions. The test cases are as follows:

- Insert of a new customer
- Update of an existing customer, changing address line 2.
- Same day insert and later update of a row (within the same batch).

The goal of this scenario is to show how one would address the issues associated with slowly-changing dimension tables where a changed-data capture table is used as input. The intent is to also show some of the features available for looking at data values before and after update from within the developer's interface. The demonstration should answer the types of questions outlined below.

- How do you perform a simple lookup from another table?
 - PowerCenter contains a tried and true lookup facility used to perform lookups to a wide variety of types. Within a mapping, a lookup transformation is included that specifies connection and lookup parameters/logic. Underneath this transformation is a scalable lookup

subsystem comprised of memory/cache management and lookup execution.

- We can do lookup both from relational as well as non relation sources i.e. flat files, SAP, Mainframe by using PowerExchange options.
- Are there any features that automate or simplify the maintenance of slowly Changing dimensions, or must you do this work yourself?
 - Wizards are included as part of the PowerCenter Designer that facilitate rapid creation of type1, type2 or hybrid type3 dimensions.
- How do you deal with both inserts and updates against a target table without requiring separate logic?
 - One of the supplied transformations, the Update Strategy Transformation supports both insert and update to the same table in one mapping.
- Do you have any problems with locking or staging when retrieving and updating rows in the target table as part of the extract?
 - No problems, concurrent operations are managed within PowerCenter
- How do you preview source data from within the design interface?
 - Data preview is an inherent facility throughout the PowerCenter Designer and within the Informatica Data Profiling and Data Quality products. On a number of design panes, you can use the preview sample data option for *every* type of source data supported.
- Can you look at both the input and output data from within the developer interface to see how updates were applied?
 - A fully functional, step-based debugger is included as part of the PowerCenter Designer. The debugger allows for preview of before and after transformation changes and enables control of the mapping debugging process.
- In the case where you do not preserve history, what steps do you have to take for a destructive load?
 - We have the truncate table option which we can set at the time we convert the logical mapping into a physical workflow.
- Is there any support for managing hierarchies in dimensions?
 - Hierarchies can be created/managed within the PowerCenter “Target Designer” but generally, we rely on data modeling tools to provide hierarchy information for us. We support Metadata Exchange of hierarchies from a wide variety of modeling tools and to a wide variety of Business Intelligence tools.

3. Extract Scenario 2: Shipments Fact Table Extract

The purpose of this scenario is to show a more complex extract involving conditional logic, calculations, many source tables, multiple lookups, and available facilities for dealing with surrogate key matching and validation for dimensional fact tables.

Scenario

The goal of this scenario is to show how one would address the more complex issues associated with building a fact table. Added complications in the data include common problems faced by developers. This includes dealing with missing data (forcing outer joins or conditional lookups), varchar and char data type conversions and comparisons, and missing values. The demonstration should provide answers to the following types of questions.

- How do you do a lookup based on a multi-column input key? (ship_method_key uses both ship_cd and carrier)
 - In the lookup transformation we define both the input key and map it to the lookup key. We can even define priority by arranging the order of the lookup keys.
- How would you accomplish the equivalent of an outer join between two tables? (the shipments table has some bad product values that won't join to the product tables, there is a non-existent customer number, and the ship_to address is missing for a shipment)
 - We can define the outer join within the source qualifier, transformation that controls the 'extraction data set' specification. Changes to the default, generated SQL are managed within the PowerCenter repository.
- Assume the developer forgot to include a table in the metadata so it isn't present but is needed at design time. How does the developer import the metadata so the table can be used in this extract?
 - We can add a new table at any time during the design phase – the developer can simply import the table from within the design interface to have it included.
- How do you do calculations involving values from multiple tables, or aggregation? (unit calculations are an example, and missing rows in one table can cause difficulties with these calculations)
 - Supplied as part of the transformation library are a variety of transformations to performing cross-table calculations. The aggregator transformation as well as the expression transformation can be used in this manner.
- What features does the tool have to automate common tasks like keying fact tables?
 - PowerCenter supports metadata exchange from a wide variety of modeling tools used to develop and manage physical and logical data models. You can also import definitions directly from the database catalogs or as a last resort, input/manage the column definitions from within the PowerCenter Target Designer, one of the tools within the design environment.

4. Extraction Scenario 3: Open Case

This is an open case. The vendors have been asked to demonstrate something that is unique or compelling about their tool. The specific features they show aren't known in advance so there is nothing to include in this portion of the document.

Trade Reconciliation Demo

Scenario: Banks exchange trading information via transactional systems and also rely on an email-based reconciliation process to verify daily exchanges with trading partners. In this particular scenario, a “host bank” is receiving email trade statements from a pair of “partner banks”. Each partner bank provides an email with trade information in a PDF or Word document.

The objective of this demonstration is to highlight the real-time and unstructured and semi-structured data processing capabilities of the Informatica platform.

Key Capabilities Highlighted:

PowerCenter Real-time Engine
Unstructured Data Support
Email Data Support

This demonstration starts by leveraging the PowerExchange for Email to process email attachments sent in by trading partners of a custodian bank. These email attachments contain trade data (aka interest swaps) that counterparties have transacted over the past month. The custodian bank needs to load this information in order to reconcile it with the real-time trades as they happened during the month (the demo does not yet address that portion).

The challenge faced by the custodian bank is that it must accommodate hundreds of customers (and growing), each one providing trade data in different file types (PDF, Excel, Word etc.) and varied structural formats. Prior to Informatica, this was a labor intensive process whereby operators would manually convert these documents into the system of record, an error prone process responsible for the high degree of downstream data inaccuracies.

Using PowerCenter we are able to read customer provided emails in real time, extract the attachment regardless of its format or file type and based on its originating source and file format route the email to the appropriate parser which in turn extracts the trade and then loads it into a header and detail table in Oracle.

5. Extraction Scenario 4: Time Dimension

This scenario involves building a derived table where there is no data source. The job must construct a time dimension based on a fiscal (4-4-5) calendar with a number of derived date attributes. The vendors have been asked to explain and demonstrate how one would construct a derived table like this where there is no data source and the data must be created by the ETL program.

This sort of extract also shows how vendors deal with looping constructs and conditional logic, both of which are needed to work out problems like leap years and fiscal years crossing into new calendar years.

- What facilities are there to generate data?
 - PowerCenter provides the ability to leverage any of the data generation facilities supplied within the underlying databases. PowerCenter also provides a “Data Masking Option” used to generate test data from existing data sources. .
- Ease of addressing conditional logic and loops.
 - Conditional logic is supported within a variety of transformation objects. Lookups, expressions, aggregator, SQL to name a few. Conditional logic is also inherent in the underlying expression language that enables if,then-

else type processing and a number of other constructs. Looping is primarily supported through the use of the Java transformation. A number of looping examples are available from the “Informatica Developer Network” or IDN, where Informatica users contribute examples of commonly encountered scenarios.

- Date management and attribute functionality.
 - These capabilities are supported throughout the wide variety of supplied transformations.
- Does the product recognize time as a specific type of dimension?
 - Not currently.

6. Maintenance Features

Post-deployment maintenance is an important aspect of the ETL process. There is no specific scenario. Instead, the vendors have been asked to describe and demonstrate features available to developers that address the types of questions outlined below.

- Assume there is a change to the data type of the order_nbr column in the Orders dimension. What is the process of tracing the impact of this change and how would a developer change the affected extracts?
 - Change-impact analysis is supported through a number of PowerCenter facilities. Within the PowerCenter Designer, “link path” enables the developer to trace forward and backward within a mapping as to where a specific object is used. At the repository level, shared objects have a facility for tracing where they are used throughout various mappings. Standard reports are also provided that document the scope and touch points to transformation objects.
 - PowerCenter Advanced Edition Metadata Manager takes what if change analysis to a higher level, broadening the analysis capabilities across the entire data integration solution ‘stack’, from modeling tool to source to PowerCenter to BI tool, change analysis can be performed through the lineage of an object.
- What source control or version control features are available?
 - We offer a full range of version control – we can check in, check out versions and also be able to use deployment groups.
- What facilities are available for documenting or annotating extracts?
 - We can document and annotate every thing from within the design interface. We also provide ability to do metadata extensions whereby we can do detailed documentation which can then be stored in the metadata repository. Through metadata exchange, documentation from modeling and BI tools can also be imported and managed within the PowerCenter repository.
- How does a developer compare and find differences between an extract that has been deployed in production and one in the repository / development environment?

- When versioning is turned on – the developer has the ability to compare multiple extracts (could be mappings, source, targets or transformations) with each other to see the changes
- How does a developer address production deployment and rollback of jobs?
 - We have the features within our versioned repository whereby jobs can be promoted from development to production and can then be roll backed if needed as we maintain all the different versions of the job in the repository unless explicitly deleted.

7. Operations and Deployment

To show features available for scheduling and monitoring, we are using the complete set of extracts for the dimension and fact tables. The vendors have been asked to show how dependencies between the various extracts are configured, how a schedule for execution is created, and how the extract jobs are monitored. The goal is to show product features related to the types of questions outlined below.

- What is the executable unit of work in the product?
 - Informatica's run-time design and monitoring architecture provides the ability to create **workflow** that consists of interdependent **sessions**. PowerCenter Workflow Manager and Workflow Monitor are role-based tools that support design and decision-based execution and monitoring of simple to complex data integration scenarios.
- Demonstrate how execution dependencies between jobs are set up (e.g. fact table doesn't load until dimension tables have loaded).
 - Within a workflow, dependencies can be set up and alternative routing/alerts based on execution.
- How do you make a change to the dependencies?
 - Visually through Workflow Manager.
- How do schedule-based initiation and event-based initiation of jobs work?
 - Included with PowerCenter is a complete, comprehensive scheduling subsystem that allows for scheduled or event-driven execution of workflow. Event driven can be as simple as waiting for a file to arrive in a directory or scripted, command-line invocation based on an event. Event driven can also leverage Informatica's Real-time or Web Services capabilities for use in a dynamic, Data Services architecture.
- How can execution be monitored by a developer within the development environment, and by an administrator outside the development environment?

- PowerCenter Workflow Monitor is available on a permissions-basis to developers and administrators based on access privileges. In addition, at design time, a comprehensive debugger is available for running mappings in “design/debug” mode.
- Explain the mechanisms available for monitoring execution and sending alerts if there are problems.
 - See above, Workflow Manager and Monitor combine to enable alerting functionality within workflow and/or visual cues of run-time problems.

8. Pricing

Since pricing is so variable we asked for information about the licensing policies and how costs are calculated.

For ETL , there is the added complication of sizing. Depending on the tools, you may need a large, small, or no ETL server. You may also need to expand the warehouse or source servers to accommodate the workload. How you configure your environment can affect the pricing of the software.

Basic criteria:

Is there a per seat cost for developers?

No costs

Is there a per seat cost for administrators?

No Costs

Is there a price per server by CPU? Per core?

Per CPU or Core depending on Server architecture

Is the price different for different server operating systems?

No difference with exceptions for Z/Linux and Z/OS

Are there per source or per target server / instance charges?

Two unlimited (source/target) types are provided as part of the base license

Source capacity and instance count are used for a small number of source types related to mainframe and/or CDC.

Are there additional source / target connector charges?

See above

Is there a charge for development or test environments? If so, is it the same cost?

Yes there is a charge but it is based on a flat charge per dev/test environment, not based on production costs and replicates production licensed footprint.

How is maintenance charged? What is the range if it is a percent of some set of costs?

18% to 22% based on desired support level

How many different editions or bundles are offered?

See above

Are there additional charges for team-based development, e.g. source control, project-level security, role-based access?

See Options discussion in initial introduction. Team-based development (Versioning, Project promotion...) is a standard part of PowerCenter Advanced Edition and an option for Standard Edition.

Please provide an estimated list price and support cost for these two scenarios:

Scenario 1: Department / project level ETL

A single data warehouse, with one ETL project

3 ETL developers, 1 administrator / operations role

2 different (non-legacy) database source types, and some file-based extracts

1 target database server

One production environment

One test and development environment

Small volumes of raw data moved through standard star-schema style batch extract, with the total target warehouse size of 180 GB of data (60GB of data loaded per year).

Standard next-business-day support

Informatica PowerCenter pricing starts at \$140K and is based on customer needs.

Scenario 2: Enterprise ETL

Multiple data warehouses/marts with several different ETL projects

10 ETL developers, 3 administrator / operations roles

3 different (non-legacy) database source types, file-based extracts, one SAP source system, requirement to consume XML formatted data

3 target database servers for warehouses / marts

3 production environments for the three projects (but infrastructure, repositories, etc. is centrally managed)

3 half-scale development environments

3 half-scale test environments

Moderate volumes of raw data moved through standard star-schema style batch extracts for two projects, with the total target sizes of each at 500 GB of data (~160 GB of data loaded per year).

One larger project environment with 2 TB of data (~650GB of data loaded per year), through more complex rules in batch mode, plus small amounts of data streaming in through your choice of either message queues / ESB / event publishing and then processed through your choice of either on-demand or in mini-batches. Specify the preferred option for streaming data consumption for the purpose of this pricing exercise.

Note: this environment requires team-based development support and project-level security and roles.

Enterprise level (same-day response) support including off-hours and weekends

Informatica PowerCenter pricing starts at \$140K and is based on customer needs.

9. Performance Features

There is no easy way in a class setting to demonstrate performance. Instead, the vendors have been asked to describe features in the product that specifically address performance needs and answer common performance questions like the following.

Informatica is the acknowledged leader in data integration performance, flexibility and scalability. Base platform capabilities, flexible packaging and performance enhancing options have been added to provide assistance with large volumes of batch and message-oriented data.

- What features are available to deal with large volumes of source data?

PowerCenter SE and PowerCenter AE provide an inherent thread-based architecture that parallelizes the “read/transform/write” process for increasing out of the box parallel processing of data within sessions. Subsequent options are added based on customer requirements for parallelization and run-time architecture.

Partitioning Option executes optimal parallel sessions by dividing data processing into subsets which run in parallel and are spread among available CPUs in a multi-processor system. This option helps organizations maximize their technology investments by enabling hardware and software to jointly scale to handle large volumes of data.

Pushdown Optimization Option enables data transformation processing, where appropriate, to be “pushed down” into any relational database. This option saves

hardware costs by making better use of existing database assets and helps organizations cost-effectively scale to meet increased data demand.

Enterprise Grid Option provides scalability within a grid computing environment. This option reduces the administrative overhead of supporting a grid. It also delivers optimal performance by automatically load balancing in response to runtime changes in data volumes or node utilization rates.

In addition, PowerExchange offers patented, **CDC** or **Change Data Capture** for mainframe, mid-range and relational databases and systems that can greatly reduce the volume of data required for processing through capturing and moving only the changed data.

- How does a developer set up or create an extract that can execute in parallel? Does it require special transforms or is it the same ETL logic regardless of parallel degree? Does it require changes to the database like table partitioning?

Informatica's unique separation between design time and run-time allows for simplified design of parallelized execution. No change in mapping logic is required. PowerCenter Workflow Manager is used to visually design an appropriate execution strategy and leverages database partitioning where appropriate and/or specific partition-able points within a mapping.

- What features are available for caching of source data or lookup values, and what are the limitations?

Caching is used extensively when sorting, aggregation, for lookups and has been optimized for large volume, dynamic use.

- How can the product be configured to run extracts on more than one physical server, if this capability is available? What must the developer do differently to take advantage of these features?

Informatica PowerCenter is based on a "**flex CPU**" packaging model where the run-time footprint can be spread across as many servers as desired for the appropriately licensed CPU count. A distributed configuration across servers is created using the web-based Administration Console. Workflow Manager can then leverage the available "distributed capacity" for various workflows. All setup, design and management of a "distributed configuration" is at run-time, not design time.

If scalability of an **Individual Session** is required, the Enterprise Grid Option is used. All set up is performed at run-time design rather than mapping design time.

- Can individual extracts be parallelized to run across servers?

See Enterprise Grid Option above. All set up is performed at run-time design rather than mapping design time.

- Are there grid / concurrent processing capabilities? If so, how do they work?

See Enterprise Grid Option above. All set up is performed at run-time design rather than mapping design time.