Case Study: ETL Platform Migration to Informatica

Transforming ETL Platform Migration using Software Enabled Services

SAP BODS MIGRATION: AN ETL FACTORY CASE STUDY

In 2014, ETL Factory worked with a specialist SAP BI vendor to enable them to better support customers who use Informatica PowerCenter as their Data Integration platform. This involved the migration of the ETL layer that extracts data from SAP using processes developed in SAP BODS to PowerCenter. The migration was completed in 70% less time and at 50% less cost of the estimated effort. These time and cost savings were realised by using ETL Factory's pattern-driven technology to rebuild the functionality in Informatica's PowerCenter platform.

Background:

The specialist SAP BI vendor had developed a solution to tackle the complexity of SAP analytics from a different angle to provide business with powerful BI, customer benefits that include the ability to rapidly integrate SAP data into an existing warehouse to answer new business questions and enhance business insight using out-of-the-box technology for rapid deployment. A specialist SAP BI vendor recognised the strategic advantage of being able to operate their ETL layer on the Informatica platform and commissioned ETL Factory to reproduce the functionality contained within their existing ETL layer, which had been implemented using BODS, in Informatica's PowerCenter platform.

Challenge:

A specialist SAP BI vendor had an immediate requirement to install their software at a global manufacturing organisation therefore the migration of PowerCenter had to be completed in a short timeframe and with 100% accuracy. In addition, the extract process required complex ETL logic to combine the required elements of the SAP data structures.

The ETL Factory architect worked closely with the engineers at the vendor to define their data integration processes as ETL patterns. There were four stages in the data extraction process and ETL patterns were defined for each stage as follows:

- SAP to SAP Pre-staging
- SAP to Staging
- Staging to ODS
- ODS to target Data Warehouse structures: defined as two separate patterns

In the first, pattern data was extracted from SAP by business classification: Customer, Company, etc. The second pattern, SAP to Staging, applied data cleansing and standardisation rules to the data extracted in stage one. The stage three pattern, Staging to ODS, performed a variety of calculations and applied various business rules based upon the context of the data and certain parameter values generated within the workflows themselves. The final stage, from ODS to the target Fact and Dimensions tables were implemented as two separate patterns: one to build and maintain the Warehouse dimension tables and a second to build and maintain the Fact tables.

Over two hundred separate PowerCenter mappings and sessions were required to reproduce the functionality of the application's ETL layer in PowerCenter. The generated code was then tested using a side-by-side comparison with the existing application. Once testing was complete then the PowerCenter layer was implemented as part of the product installation at the global manufacturing organisation.

The Pattern Definition Process

The process to define the ETL patterns was driven by the functional design of the code base. In this, the extract process is complex due to how SAP stores data that relates to a single entity, such as 'Customer' across a number of source tables. After extraction, data standardisation rules are applied and finally the data is loaded to the target data warehouse structures. The patterns that ETL factory defined followed the same four-phase process.

ETL Factory developed five patterns in total. At each stage the relevant part of the code base was analysed and the different ways in which source data was joined or referenced was catalogued and built into the relevant pattern, the same process was repeated for the other functional stages by identifying the standardisation and cleansing logic needed for step two, the calculations and other rules for stage three and finally the proprietary rules used by the vendor to build their target data warehouse structures.

Side-by-side comparison of the output from each stage of the existing process and the new process enabled mismatches to be identified and then applied to the relevant pattern. The code was then regenerated and re-tested on an iterative basis until each stage of the old and new processes matched.

Business Benefit:

Prior to engaging ETL Factory, the SAP BI vendors had obtained an estimate using industry standard metrics on the effort required to build and test the process in PowerCenter. The estimated effort was 355 days based upon the following metrics:

- Mapping development, 110 low complexity mappings: 50 days of effort
- Mapping development 95 medium complexity mappings: 100 days of effort
- Testing and remediation: 50 days of effort
- Productionisation: 5 days of effort

Using ETL Factory, the same delivery was completed in less than 100 days and required a significantly smaller team size than manual conversion would have required. Resource usage broke down as follows:

- Pattern development and code generation – 30 man days of effort
- Expert input from The specialist SAP BI vendor staff – 20 man days of effort
- Testing and remediation – 30 man days of effort
- Productionisation – 5 man days of effort

Cost savings of 45% were realised from the use of ETL Factory's software-as-services model over the traditional offshore delivery model that is the alternative approach to platform migration projects.

Summary:

By focusing on a software solution utilizing pattern matching and template generation the ETL Factory team cut the effort required in the above example to less than 30% of the man days effort and less than 50% of the estimated costs of offshore effort to deliver the same solution.

About ETL Factory

ETL Factory has helped our customers to significantly reduce developments costs and delivery times by providing innovative software that accelerates the development of data integration processes on the Informatica platform.

Our software operates by enabling the user to define a common underlying pattern for an ETL process and then to generate the code base that is needed by combining the source and target metadata with the ETL pattern.

The result is high-quality and standardised application code that enables data to be delivered to the business in days rather than weeks or months. When requirements change then the code base is regenerated by applying changes to the pattern rather than to individual code objects to enable rapid turn around and truly agile delivery.