Introducing PDI's Parsers & Mapping Generators

The Drawbacks of Python

Python is quickly growing to overtake Java as the most popular coding language, as Python has a wide range of uses and helps developers build strong code quickly. But as companies look to migrate to the cloud, they are encountering problems familiar since the 90's, when developers began writing custom PL/SQL and COBOL code to move data. No matter the language, the fact remains that every developer has a unique style. The individual who wrote the original code will always understand it better than anyone who steps in to replace them.

This is the issue companies are beginning to grapple with today, and it is a problem they will continue to face as developers move on from job to job. When a Python developer leaves a company, it's hard to keep track of their custom scripts and manual coding - some information is always lost during the knowledge transfer, no matter how skilled the new developers are.

This struggle is the very essence of why ETL tools were created, so companies do not have to write code. Instead, they have a visual, graphical interface that can be leveraged across developers with no learning curve.

PDI's Python Parser

Strong ETL tools, such as the tools that Informatica offers, are becoming essential to companies embarking on the journey to cloud. These tools have been made popular due to a myriad of benefits, including:

- Ease of Use
- Operational Resilience
- Advanced Data Cleansing

- High ROI
- Visual Flow

The list of benefits are long, but the arduous task of migrating legacy python codes to Informatica Cloud can make a company hesitate to take that step forward. Creating so many mappings takes a huge investment of man-hours, but not with Pacific Data Integrator's Python Parser to INFA Cloud!

To mitigate the rising struggle of converting legacy python metadata, PDI has developed a parser that reads Python scripts and creates INFA Cloud mappings as a streamlined and cost effective way to gain control of the data integration design pattern necessary for migrating on-premise data to the cloud. Any company leveraging the cloud today needs a robust data integration pipeline from on-premise to the cloud and vice versa, using a next-generation tool such as Informatica Intelligent Cloud Services (IICS). In short, with PDI's parser a company's legacy data pipelines built on python will be converted into Informatica's cloud ETL.

Now that PDI's Python Parser can read an organization's python scripts and create the Informatica Cloud mappings, companies will be able to migrate to Informatica Cloud, to mitigate the challenges of using developers to personalize code, and to convert legacy scripts with ease!

PDI's EDC Python Parser

Informatica's Enterprise Data Catalog (EDC) is a robust data cataloging tool that gives companies an unparalleled understanding of their data lineages. The extensive benefits of EDC, such as user-friendly interfaces, are why EDC is Informatica's fastest growing product ever.

However, a companies' understanding of their data movement is incomplete as long as the intelligence present in the company's legacy python code is collecting dust. Only when this metadata is parsed and exposed in EDC, will business users have access to an accurate view of their business's metadata and data lineages in EDC's visual, user-

friendly dashboards.

There is a plethora of data lineage information locked away in a company's python code, but this data is often ignored because migrating legacy python code is an arduous task. That's no longer the case with Pacific Data Integrator's EDC Python Connector!

To unlock the metadata stored in company's python code, our tool parses the python code, extracts the metadata, and exposes it on the EDC framework via a custom connector. Our EDC Python Connector combines EDC' out-of-the-box connectors with our modern microservices architecture. Our micro web framework leverages ANTLR, React Gatsby, Python Flask and is fully compliant with microservices' architecture.

In short, the EDC Python Connector is a streamlined, cost-effective way to leverage the metadata trapped in legacy python code in EDC. Your corporate data movement will remain incomplete until you have access to this information!

The Limits of the Out-of-the-Box Connector

Informatica offers an out-of-the-box SAP connector for their Enterprise Data Catalog (EDC). This connector will bring the metadata for a company's SAP tables into EDC, but the drawback of this general SAP to EDC connector is that it pulls all existing SAP table metadata into the Enterprise Data Catalog.

The downside is that this process can easily result in tens of thousands of tables! This mass migration of data crowds EDC with unneeded information and swamps it with thousands of inessential tables.

This is why Pacific Data Integrators (PDI) developed a specialized EDC Connector for SAP, to offer a more streamlined and targeted approach. Rather than pulling all SAP metadata, our tool specifically targets Informatica metadata.

The PDI EDC Connector for SAP parses Informatica PowerExchange metadata for SAP and ONLY brings the SAP metadata for tables that are being used in the data warehouse into EDC. The result? All relevant data is parsed, without the clutter of unneeded SAP data filling up a company's Enterprise Data Catalog.

PowerExchange to EDC Scanner

Leverage PowerExchange Metadata in EDC

Pacific Data Integrators (PDI) is proud to announce our new PowerExchange to EDC Scanner. This powerful custom scanner parses Informatica PowerExchange metadata, empowering users to add the full data lineage visibility of a source system into Informatica Enterprise Data Catalog (EDC) – without having to connect the source system.

Connecting a source system to EDC extends project timelines, as architects have to get permissions, ensure connectivity, and work with dependency on other systems. This is an arduous process, but with our PowerExchange to EDC Scanner, users can now bring data from their source system into EDC *without* pouring time and money into connecting the source system.

With this custom scanner parsing your Informatica repositories, your company can bring the necessary data lineage for analytics into Enterprise Data Catalog and expose your data lineage in a rapid way.

Pacific Data Integrators has a wide range of connectors and parsers that are all built on the latest cloud technologies using modern microservices architecture. This micro web framework leverages ANTLR, React Gatsby, Python Flask and is fully compliant with microservices' architecture. For more information, visit www.pacificdataintegrators.com for more information.

PDI Mapping Generator

It is essential that IT keep pace with fast-moving business environments, which means IT must operate with efficiency, speed and agility.

IT projects are frequently viewed as a significant cost, yet at the same time they are known to be crucially important to delivering technology that transforms how business operates. Keeping this in mind, it becomes readily apparent that cost reduction has to begin within IT methodology.

PDI has created tools that use existing APIs to automate redundant and repetitive tasks, which create significant timeline reductions and significant cost saving. For example, our team has used this automation technology to migrate from legacy IBM MDM to Informatica MDM.

This automation technology has been used to migrate from legacy IBM MDM to INFA MDM.

Product Description

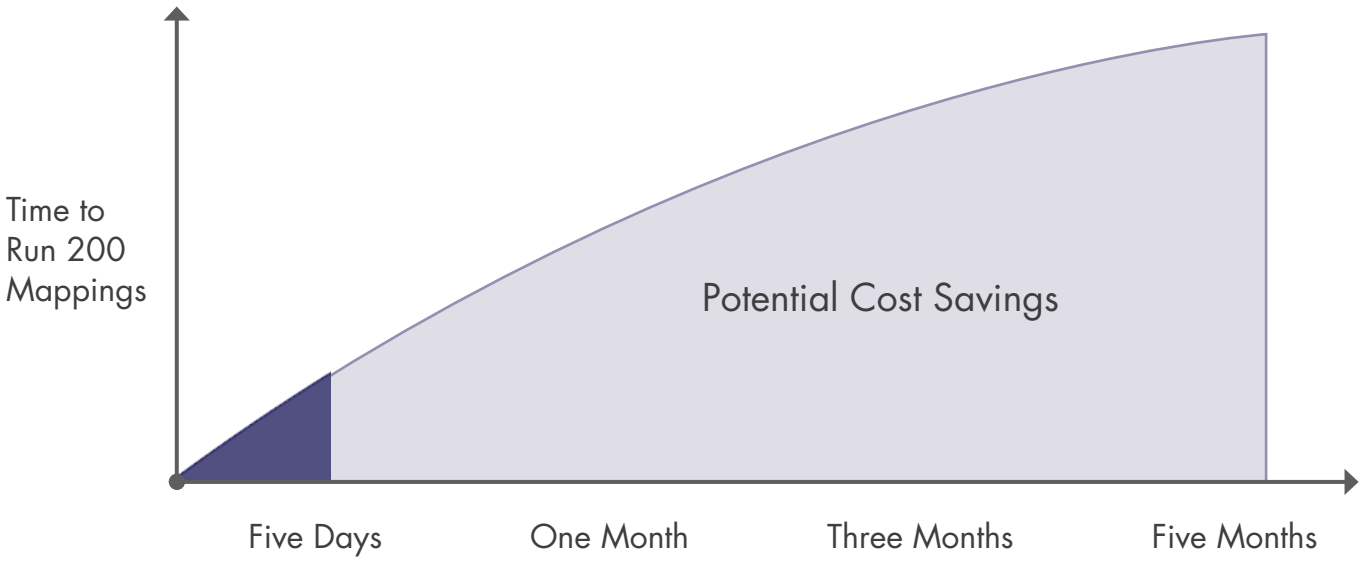
PDI's Mapping Generator uses Informatica Java APIs to generate a Metadata Object file. This Metadata Object file can be then used inside Informatica via the PowerCenter Designer/Workflow Manager or Repository.

In order to generate a Metadata Object file, a user would first need to launch the Mapping Generator.

Once launched, the user would connect to a source database and a target database. The tool will then list all available tables in each database Next, the user would manually create a connection between the source and target tables. Once all connections have been crafted, the user could create an Informatica Metadata Object file that contains information about the source and target database.

After the connection has been made, the Mapping generator can make any mapping between source and target or session and a workflow.

This means that the tool can generate nearly any sort of arbitrary mapping.



The typical five month timeline that 200 mappings would take is cut down to five days when using Pacific Data Integrators' Mapping Generator.

Convert any ETL to Informatica

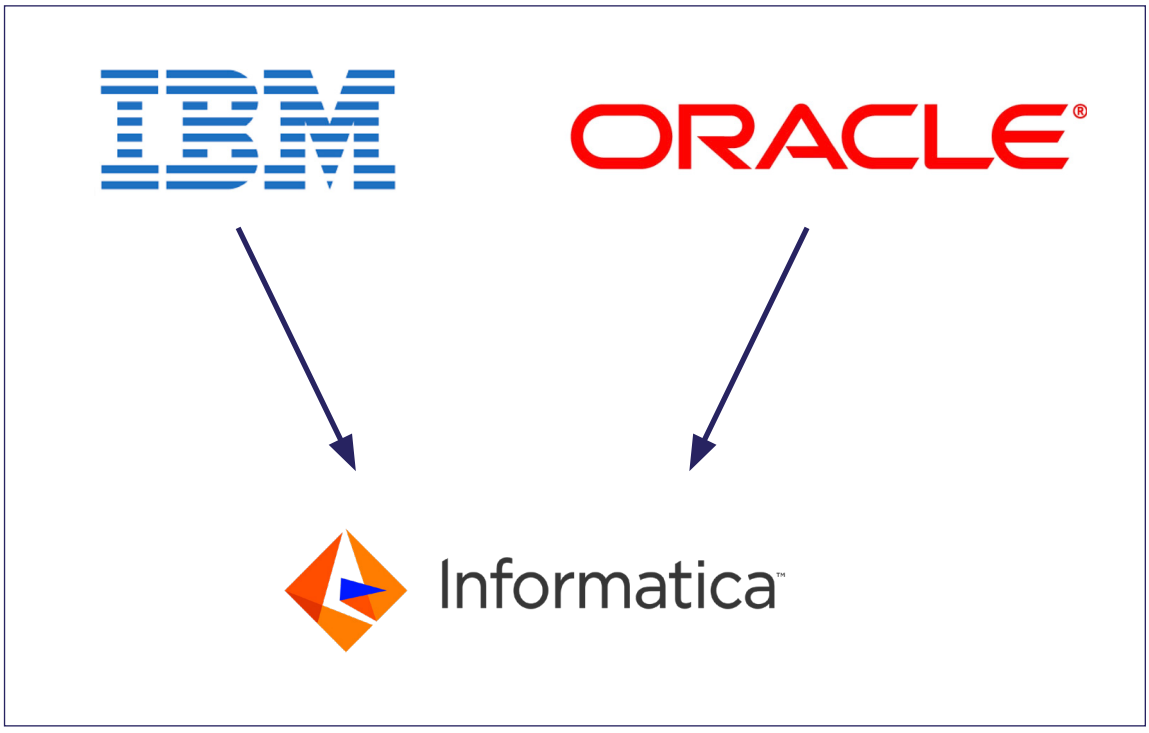
Our robust Mapping Generator is able to convert any ETL into Informatica Cloud in a significantly shorter time than a manual ETL conversion project would require. It is modular enough to read any metadata repository and convert it using our pre-buld code blocks using Informatica API.

We have successfully used our Mapping Generator to convert platforms from:

- IBM Cognos ETL, on-premise only
- Oracle OWB ETL, on-premise only

Complete your own data conversions in record time and ensure your projects are successful, rather than attempting to tackle time-consuming conversions manually.

Our team has successfully leveraged our mapping generators to deliver projects, including a project where we read IBM ETL and converted it to Informatica for a healthcare company out of Canada.



Our Mapping Generator converts metadata from IBM and Oracle.

Custom Scanners

In addition to reading and converting IBM ETL, our team has created scanners for several PowerCenter sources. Informatica's Enterprise Data Catalog (EDC) has out-of-the-box scanners for many of Informatica PowerCenter's sources, but for the PowerCenter sources that EDC does not scan, our team can create custom scanners that scan the metadata in the system.

Additional Supported Templates

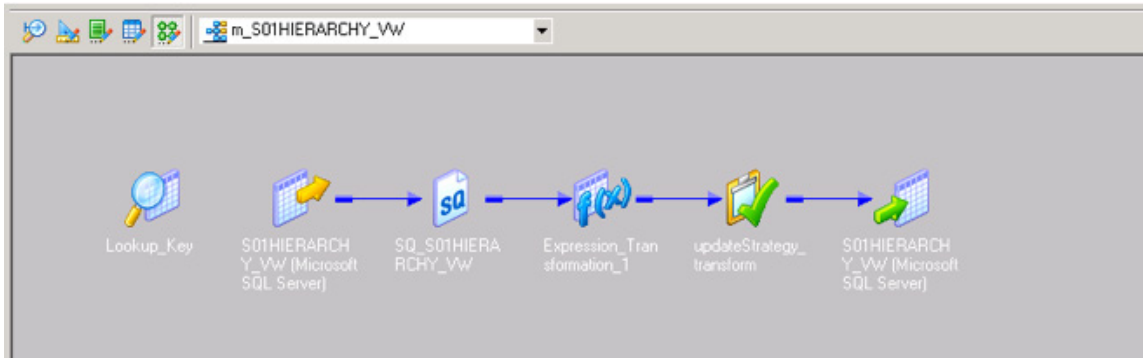
We took our advanced Mapping Generator a step further by creating several templates that can generate Informatica mappings in minutes; a process that could take a developer several days to build and test.

These additional functionalities leverage a template-based approach using Java API's. Templates are available out of the box in an easy-to-use excel file. For each template, our Mapping Generator dramatically reduces the time it takes to map various functions.

For instance, our Mapping Generator can produce an average Type 2 Dimension Load in fifteen minutes, where creating it from scratch would average one hour. Compounded by one hundred mappings, our clients have saved an average of 75 hours with our automation tools.

Type 1 Dimension Load

This type of mapping needs to update existing rows. Note that a Type 1 Dimension Load does not maintain a history when updating rows with changes; it overwrites the existing rows. Our tool generates code that allows a customer to quickly and easily create a Type 1 Dimension Mapping for several dimension tables in a matter of minutes

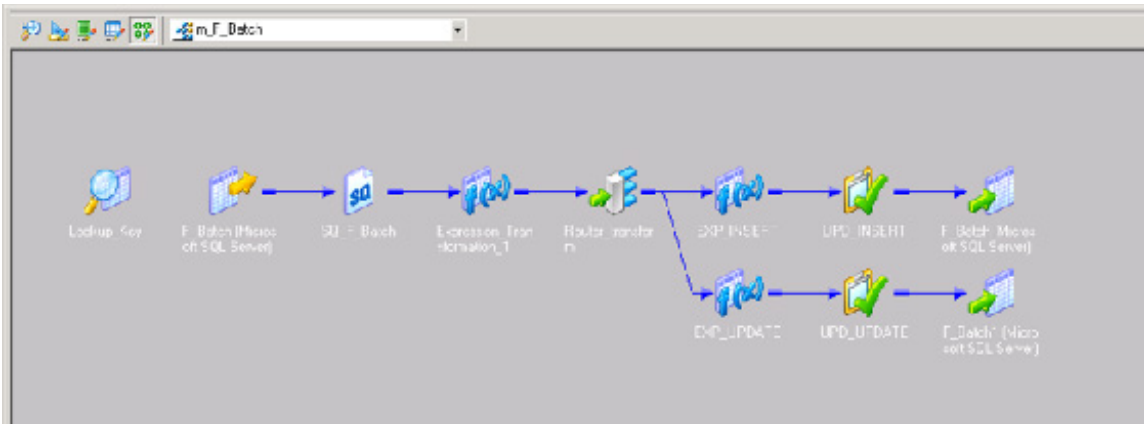


Example of a Type 1 Dimension Load Mapping

Type 2 Dimension Load

A Type 2 Dimension Mapping also has the ability to create new rows and update existing rows. Unlike a Type 1 Mapping, a Type 2 Dimension Load maintains a history when updating rows with changes.

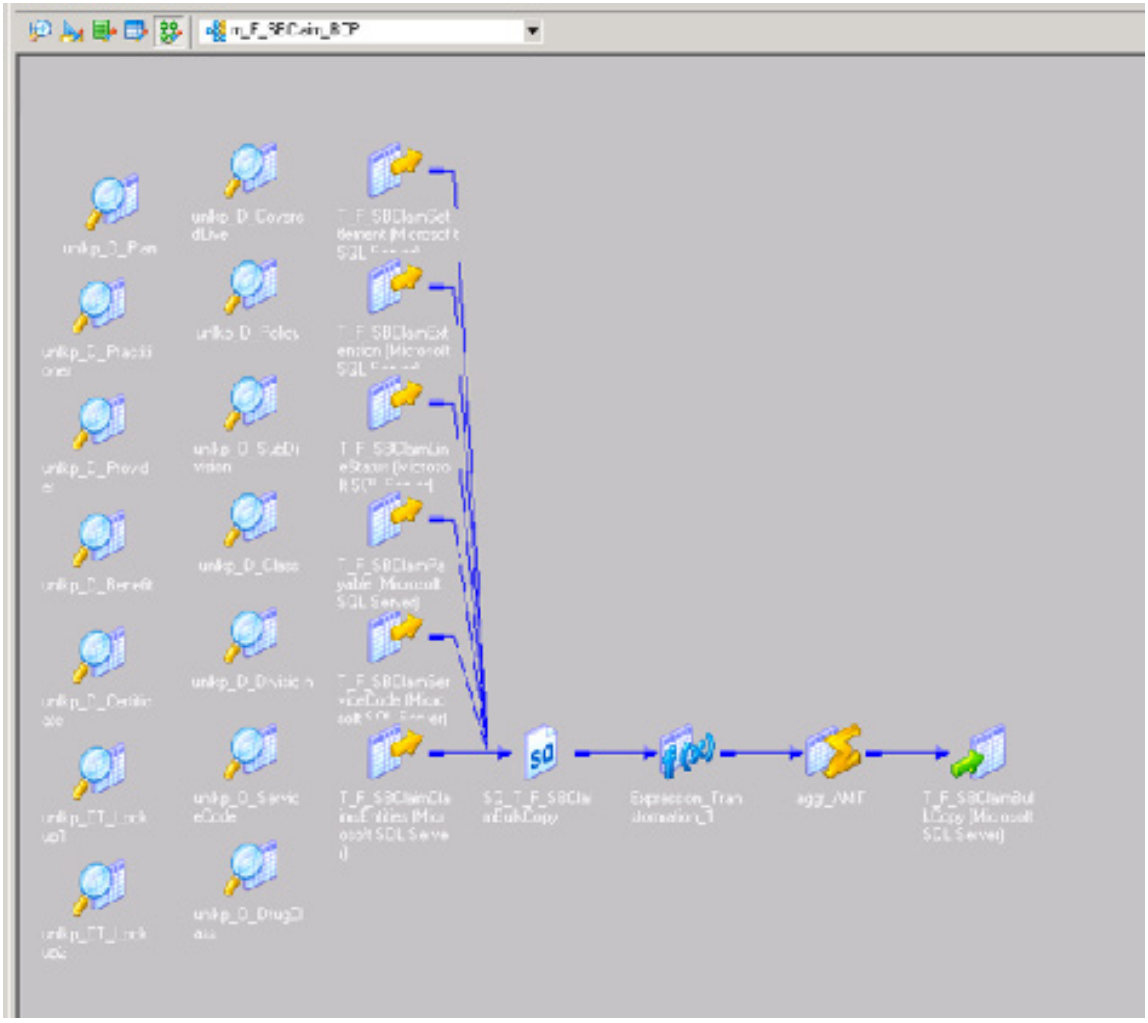
The tool can detect whether a dimension table is Type 1 or Type 2 and create mappings accordingly, based on the input configuration file. In addition, the tool has the intelligence to decide when a dimension has one column as part of the business key, or several columns.



Example of a Type 2 Dimension Load Mapping

Aggregate and Summary Table Loads

Our Mapping Generator also has a template to support aggregate and summary table loads.



An example of Aggregate and Summary Table Loads