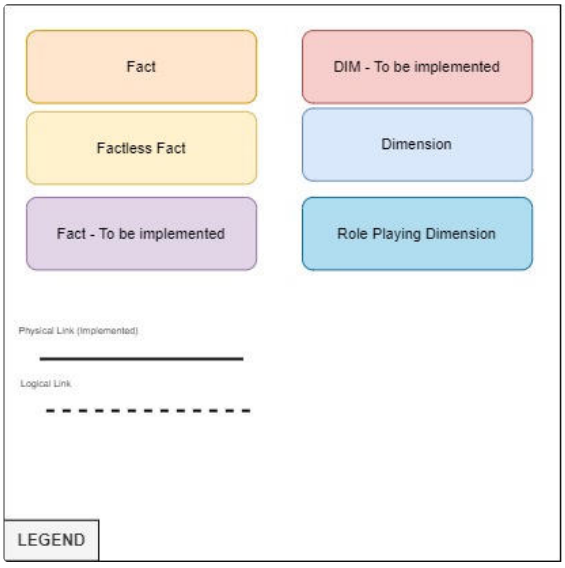
DWH Core Concepts

What is a data warehouse (DWH)?

A data warehouse (DWH) is a special kind of database designed to collect and store data from different sources in one place. Imagine it as a large, organized library where information from different areas of a business—like logistics, finance, and HR—is brought together and stored for easy access. This data is cleaned, organized, and structured so that it can be used for analysis, reporting, and decision-making.

Legend for the Data Model

Please use the following as a legend over the entire data model:



Legend to be used across the documentation

Entities in the Data Warehouse

The data model of the DW provides a collection of **simple but rich** entities (tables) coming from operational systems (Fin4u, Medex, HR, IM, etc). Those entities supports business oriented scenarios. For example, loading purchase requests, budget, and account entities into a dashboard allows end-users to aggregate data on the primary areas of financial spending allocation (i.e., an executive board member was interested in knowing the percentage of PRs by account number, and by loading the aforementioned entities, the executive realized that 30% of the allocated by budget is spent over purchasing materials (construction materials, renting, etc.).

> **In the following, we will describe some of the terminologies that identify a type of entity.**

Dimension Tables

In short, a table is defined as dimension (DIM) in a the data model, when it provide additional information and context around a business process. For instance, a transaction table (Actual) has the following fields: account, org unit, project, award, amount, amount (USD), date, etc. The information around the amount are considered as a dimensions where users can load to add meaning to the values tracked across the organization.

Point-in-Time Dimensions

Certain dimension entities loaded into the model are designated as Point-in-Time (PIT) tables, allowing users to track historical changes within dimensions. For example, the currency rates table is identified as a PIT table because daily exchange rates are loaded as they are updated from the source. Each record in these entities includes a "LOAD_DT" column, indicating the date and time the record was added to the model.

When working with PIT tables, users should be mindful that they may sometimes require only the most recent data. In such cases, they should use a designated function (*max(LOAD_DT)*) to retrieve the latest record based on specific criteria. More on that in the how-to section.

Role-Playing Dimensions

Some dimensions in a data model can represent different roles depending on where they are used. For example, the “**Procurement Date**” dimension can be used as a “**Shipment Date**”, “**Delivery Date**”, “**PR Submission Date**”, or even an “**Actual Delivery Date**”.

Role-playing dimensions are structured to consolidate redundant data into a single table rows. By loading various entities into one dimension table and labeling them appropriately, the data becomes flexible and easy to filter.

For instance, if a user is reporting on Purchase Requests (PR), and they want to look for items adjusted today: Created, updated, receive, or paid today; they can, using a single filter, to find all items related to a specific date.

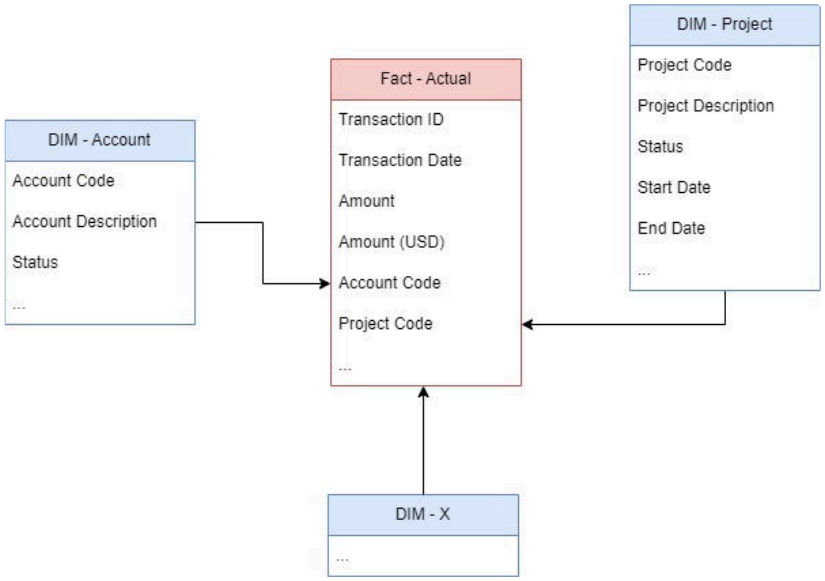
Here's the current list of role-playing dimensions:

- **procurement_dates**
- **procurement_users** (owner, buyer, etc.).
- **procurement_sites** (requesting site, destination_site, etc.).

Fact Tables

A fact table is a central table in a data model that stores the key details about an organization's business process. Each row in fact table represent a numeric value (i.e., transaction amount) with keys that links it to dimensions. Let's have an example with the actuals table as it's considered technically a fact table. To define it, each row in the table represents an amount spent over a project, award, location, account, etc, per date.

Below, an example of a model with a central fact table and dimensions around it:

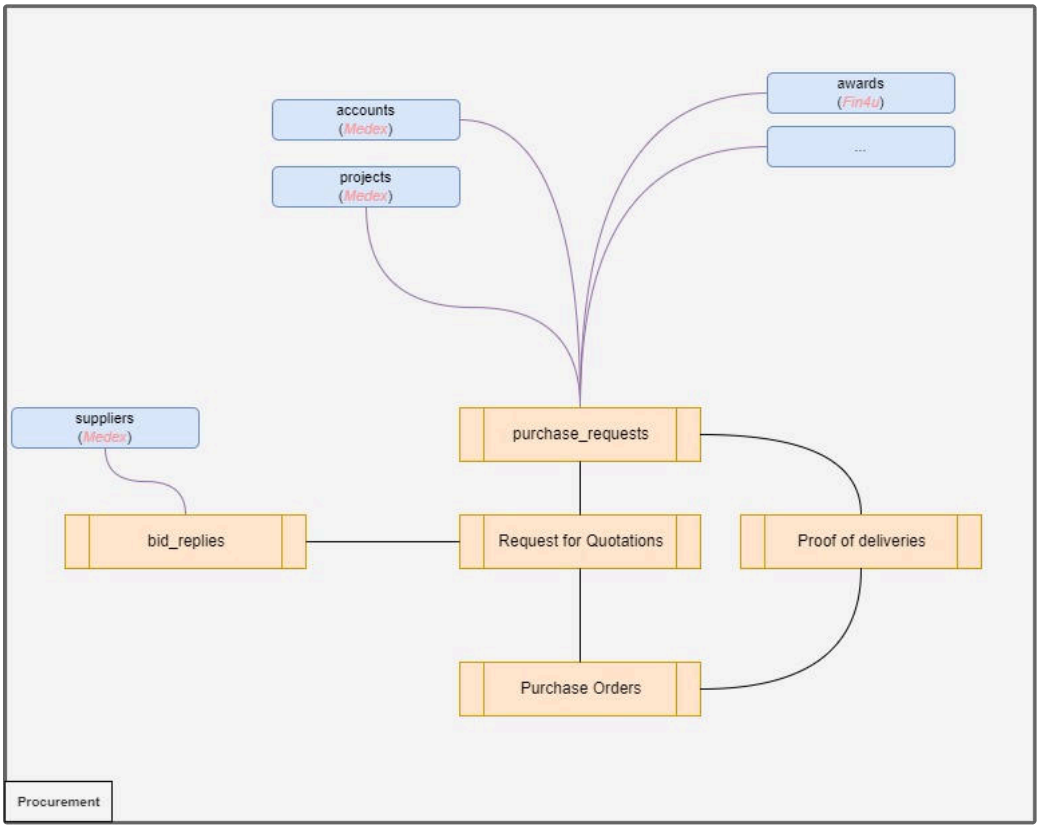


Example of a data model in DWH

Factless-Fact Tables

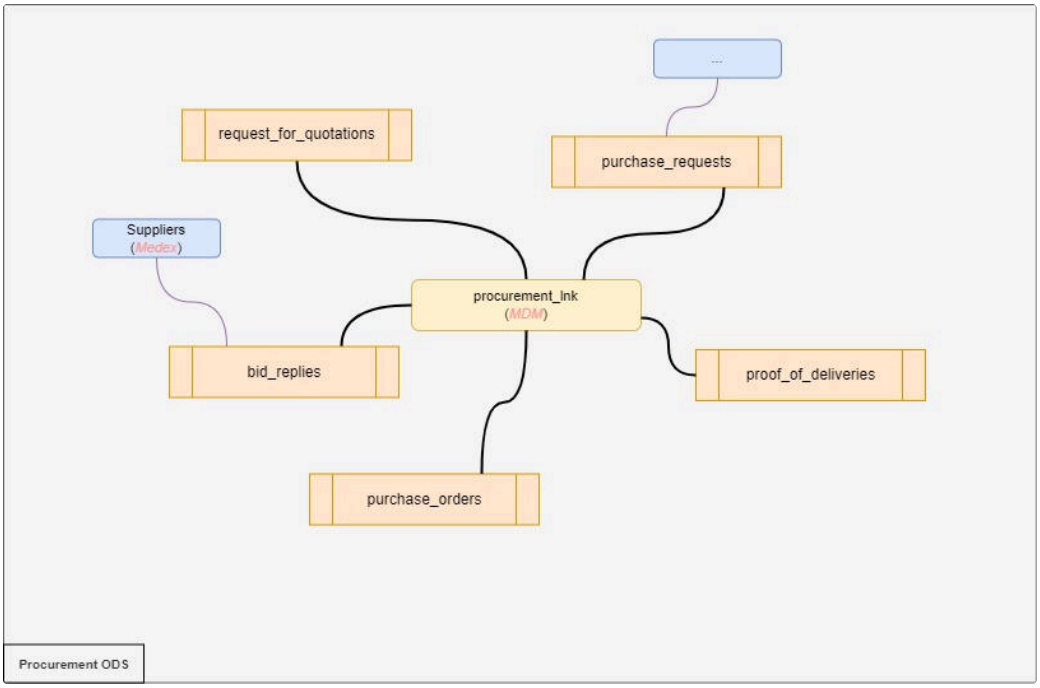
A factless fact table is a type of table in a data model that doesn't contain any measurable data, like actuals or purchase orders, but instead tracks events or conditions. It's used to record that something happened without any numeric values.

For instance, in Medex, the procurement contains at least 5 main modules with complex linking as shown in the image below:



Procurement model in Medex

However, with a complex linking between various modules we had to implement a different data model structure for the Medex part of the DWH to provide flexible, non confusing model for the end-users to work with, which resulted in the model below:



Implemented Medex's procurement in the DWH

As illustrated in the image above, we have introduced a "Procurement Link" table, which functions as a **factless** fact table. This table serves as a connector between other fact tables, containing a combination of keys from those entities but no numeric values that represent measurable business processes. Essentially, this **factless** table is designed to capture non-numeric events.

This approach eliminates duplication and offers a simple, flexible, and comprehensive data model for end-users to interact with effectively.

For example, when a new purchase request (PR) is created and linked to a purchase order (PO), even at an early planning stage, a record linking the two is added to the link table. As the process progresses, the same PR can be associated with additional keys, such as bid responses, purchase orders, and proofs of delivery, all of which will be tracked within the link table.