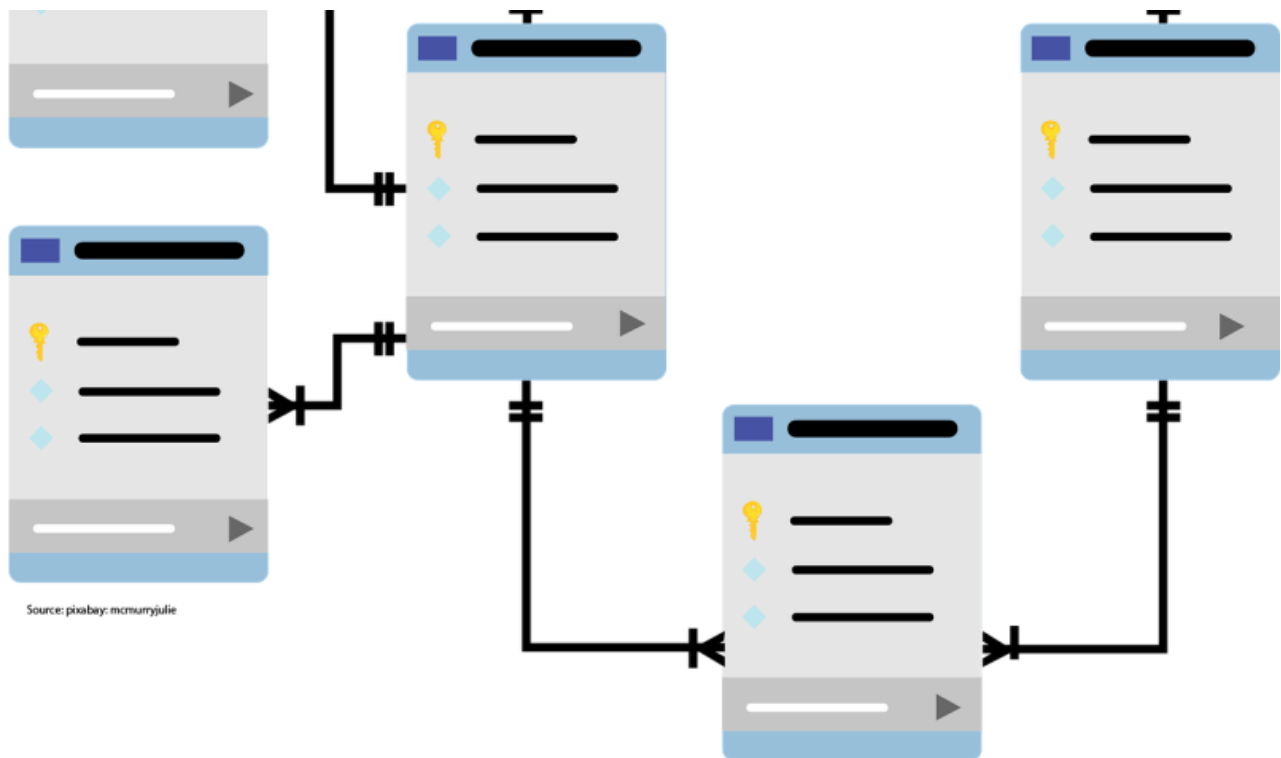


Dimensional Modeling – My Cheatsheet

 johnxiong.com/2020/04/10/dimensional-modeling-my-cheatsheet/

August 28, 2021



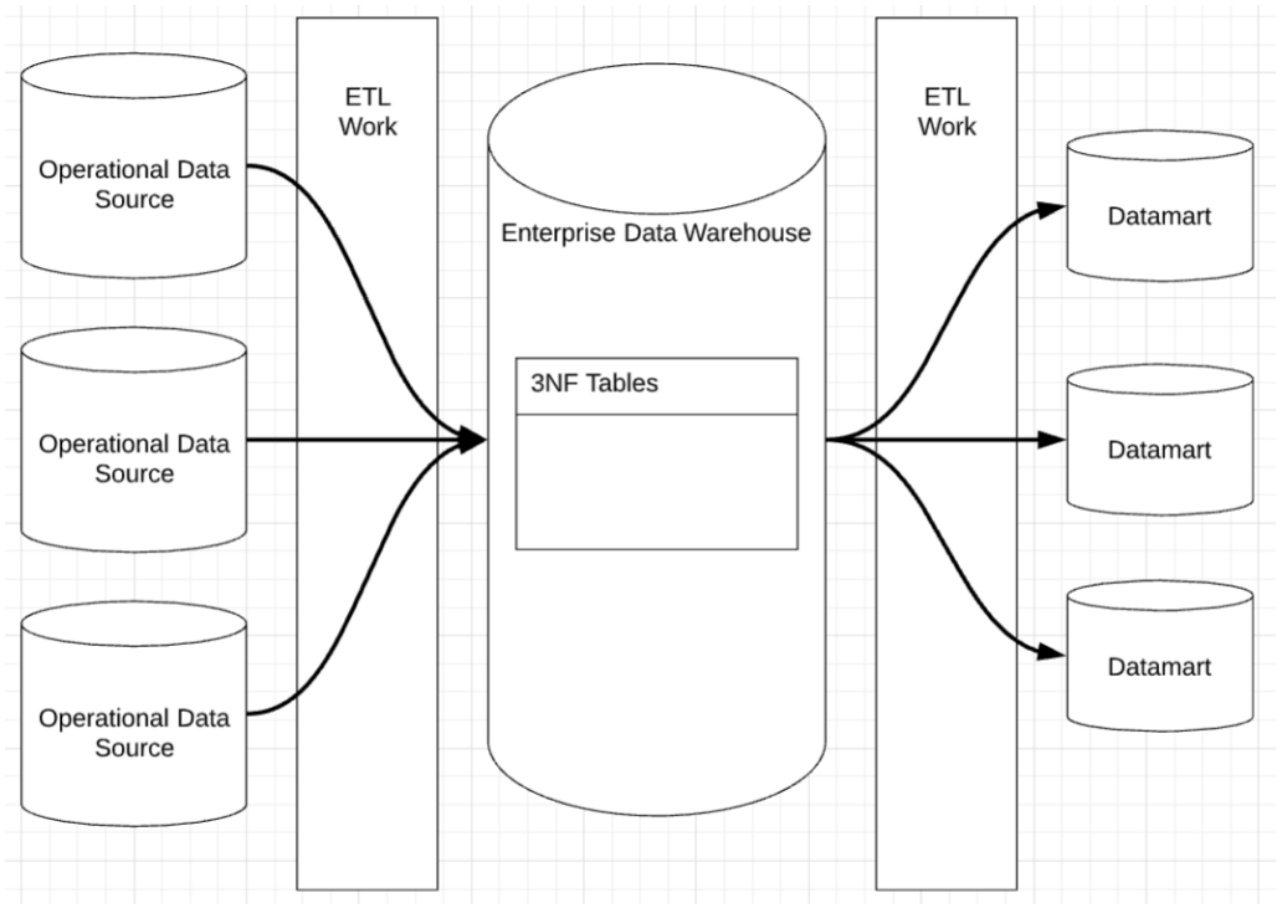
Note – This post is used as a cheatsheet that gives a high-level overview of the Kimball approach to dimensional modeling. If you seek further information, there are plenty of resources such as <https://www.kimballgroup.com/> to answer your question; which is basically where most of this information is from anyway.

What is Dimensional Modeling?

Dimensional modeling is a way to help businesses gather their data in a conformed way and use it analytical purposes. When the data is grouped, that is often referred to as a data warehouse. There are various ways of creating a data warehouse, however, two of the more popular architectures stem from the Inmon vs Kimball debate. Each has their pros and cons that I will go over at a high-level, but I will be more focused on the Kimball approach for the latter part of this post.

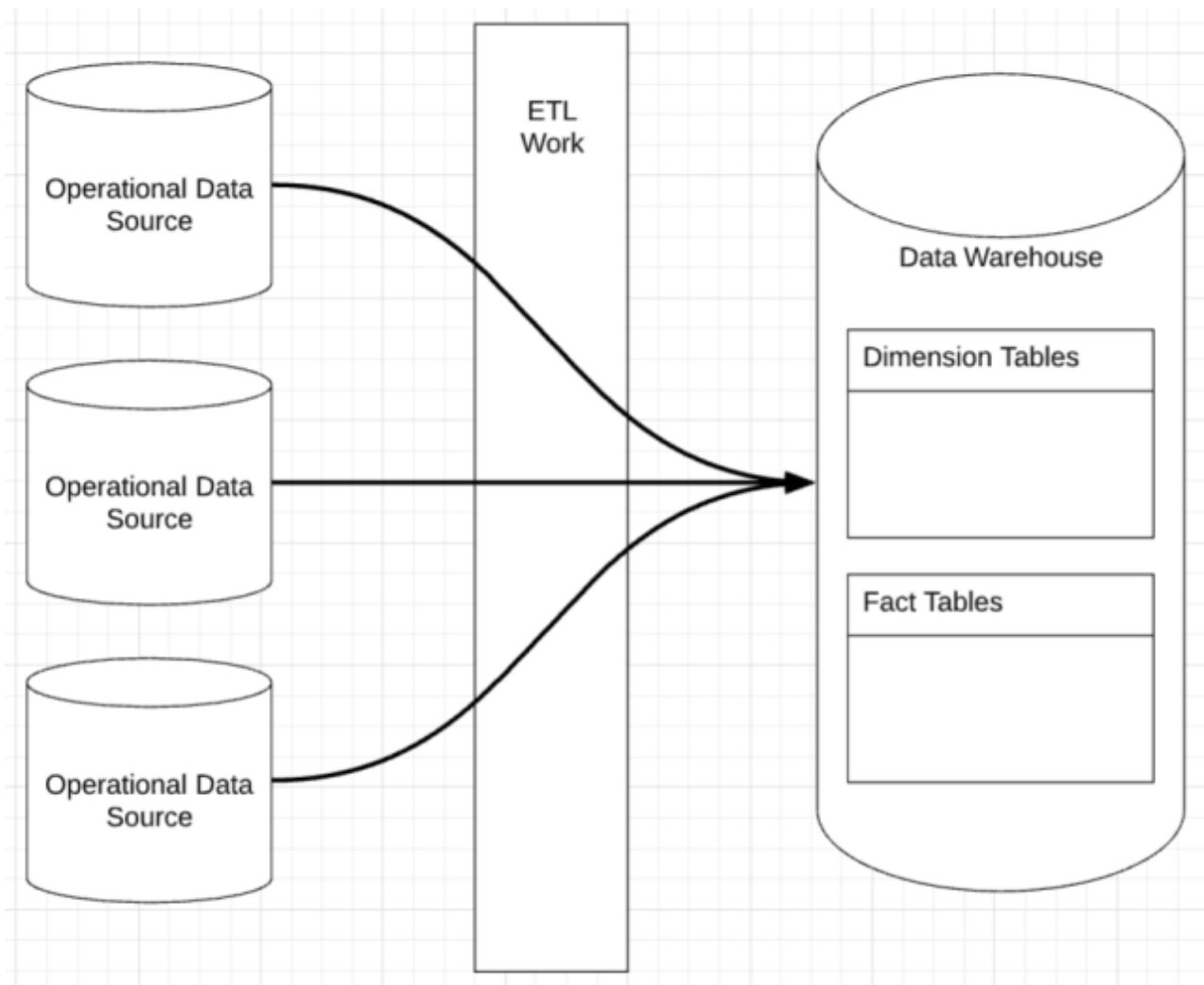
Inmon vs Kimball

Inmon Approach



The Inmon approach, also known as hub-and-spoke corporate information factory, extracts operational data and loads it into a 3NF database. This location is considered the enterprise data warehouse. Data is then delivered into data marts where data is often summarized and departmentalized.

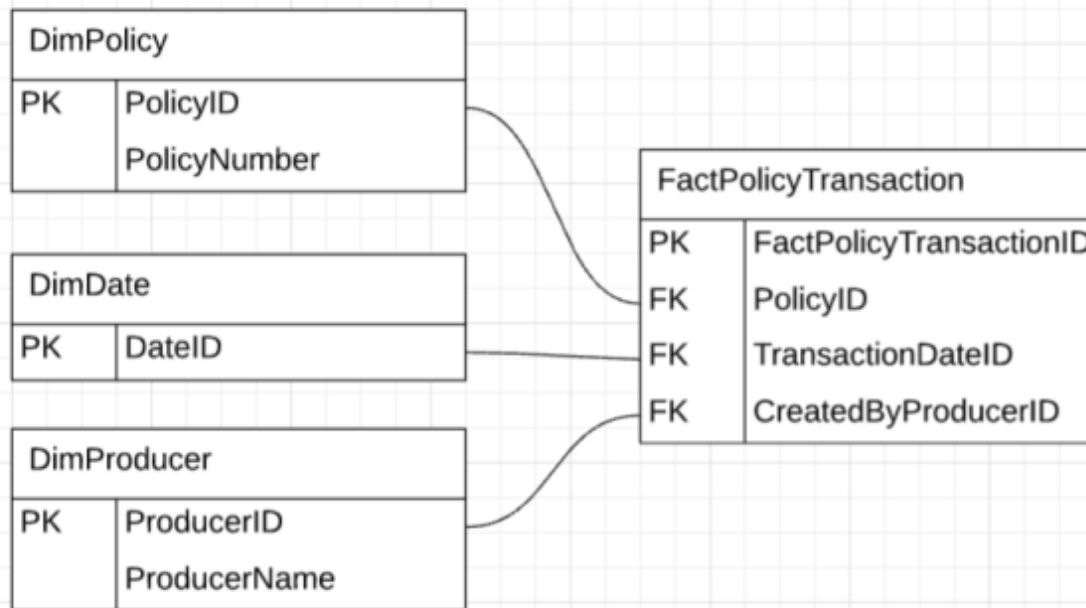
Kimball Approach



The Kimball approach extracts operational data, transforms it in staging tables, and loads it into a dimensional model; where there are conformed dimension tables and fact tables for measurements. This dimensional model is the data warehouse where business users can access the data.

Basics of Kimball

As stated earlier, data is extracted, transformed, and loaded into a dimensional model. This dimensional model consists of dimension tables and fact tables. Dimension tables store textual context that associates with business processes. For instance, a personal auto insurance policy would be a dimension when that policy is issued. Fact tables store performance measurements based on business processes. Extending on the personal auto policy example, an example of fact table in this context would be recording written premium based on a policy transaction. In the image below, we show the relation between the policy transaction fact table and a few other dimension tables:



Dimension Tables

Depending on how data is used and architected, there may be a need for different types of dimension tables:

- **Degenerate** – Usually stores primary keys and is often used to grouping related rows in a fact table. Example: grouping claimed items to a policy.
- **Multiple Hierarchy** – When a dimension naturally has a hierarchy within its data. Example: dates can be day-to-week, week-to-month, or fiscal period-to-year.
- **Junk** – Used to avoid creating multiple dimensions and/or reduce number of columns in a fact table for low-cardinality information by combining them together. Example: Unique combinations of available vehicle colors and body type.
- **Outrigger** – A dimension that is created and used by another dimension to represent more details. Example: A new dimension to determine other information when a policy was written referenced in the policy dimension.

Slowly Changing Dimensions (SCD)

There are times when dimension table data need to be altered or stored in various ways (or not at all); making the table a slowly changing dimension. There are 7 ways that handle most scenarios:

- **Type 0:** Retain Original – Attributes will never change.
- **Type 1:** Overwrite – Overwrite an attribute if it has changed, reflecting the most recent version of the record.
- **Type 2:** Add New Row – Expire previous row and add a new row with the new information. Set as the current record.
- **Type 3:** Add New Attribute – Move old attribute value into the “old_value” column and update the main value column with incoming value.

- **Type 4:** Add Mini-Dimension – When a group of attributes in a dimension are rapidly changing is split into it's own dimension; aka Rapidly Changing Monster Dimension.
- **Type 5:** Add Mini-Dimension and Type 1 Outtrigger – Preserve historical attribute values and report historical facts according to current attributes.
- **Type 6:** Add Type 1 Attributes to Type 2 Dimension
- **Type 7:** Dual Type 1 and Type 2 Dimension – Support as-was and as-is reporting

Fact Tables

Fact tables represent an event and store attributes and measurements of that event. Based on rules, there are different types of fact tables and different types of numeric measures.

Types of Fact Tables

- **Periodic Snapshot** – Summarizes measurement events over a period of time, such as daily, weekly, or monthly.
- **Accumulating Snapshot** – Summarizes events occurring in predictable steps. Example: A row with dates to determine when a policy was quoted, bound, and written.
- **Factless** – Records when a set of dimensions come together at a point in time. Example: A potential customer meeting with various agents.
- **Aggregate** – Numeric rollups of other fact tables to increase query performance.
- **Consolidated** – Combining other facts into a single consolidated fact table.
- **Transaction** – Corresponds to an event in space and time.

Types of Numeric Measures

- **Additive** – When a column can be summed across all dimensions.
- **Semi-Additive** – When a column can be summed across some dimensions.
- **Non-Additive** – When a column can't be summed at all.

4 Step Process

So how does this all come together when creating a dimensional model? There is a 4-step process that should be followed that incorporates mostly everything above:

1. **Select the business process** – What metrics are you looking to gain insight on based on a business process or operational activity? Example: Filing a claim transaction.
2. **Declare the grain** – What is the lowest level of should the data be captured? Example: In a claim transaction, should the replacement cost be loaded individually or aggregated for all scheduled items?
3. **Identify the dimensions** – The who, what, where, when, why, and how of the business process. Example: The producer, claimant, location, date and time, reasons.
4. **Identify the facts** (measurements) – What information to we want to measure? Example: How much the replacement cost is for each scheduled item.