Learning Data Vault is like learning how to make Beer!

Cognitive load[1] theory, developed in the 1980s by John Sweller, is an important concept in learning enforcement and the theory is applied in school education[2]. Cognitive load refers to the used amount of *working* (human) memory resources – the limited capacity that can hold information temporarily. To quote, "the human brain can only process a small amount of new information at once, but it can process very large amounts of stored information. The average person can only hold about four 'chunks' of information in their working memory at one time."[3]

Rather than bombard the student with all information all at one time, introducing a new concept is performed by limiting the new information into digestible 'chunks'. Further… "information is stored in the long-term memory in 'schemas', which provide a system for organising and storing knowledge."

When information is very complex or new, it is important that teachers reduce the load on students' working memory as much as possible to maximise learning

When information is easy for students to understand, teachers can gradually increase the complexity of the lesson to maximise students' learning
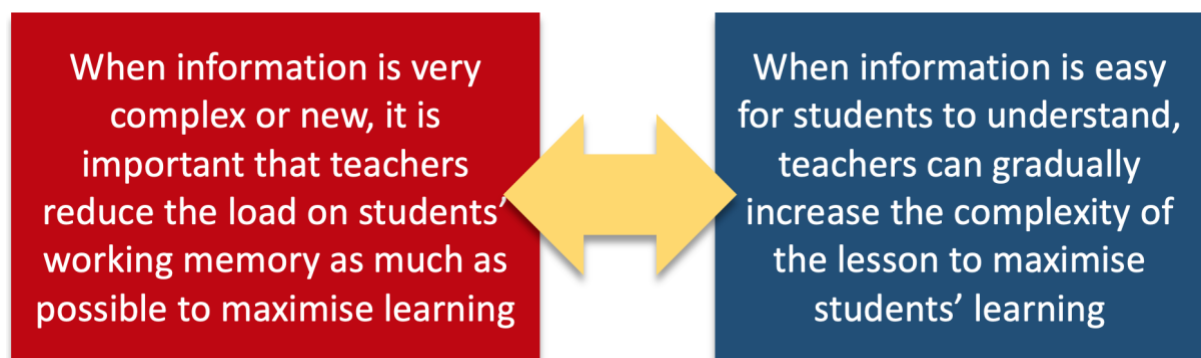
Figure 0-1 lifted from Department of Education, Cognitive Load Theory in Practice (NSW Government)

Once new information is learnt the information is reinforced and as the student is taught new information that information already has a base to build from. Gradually as the student becomes more proficient in the subject matter independent problem-solving is introduced.

Let's illustrate by way of an example…by making beer!

"Brewers make wort. Yeast makes beer"- Dave Miller[4]

---

[1] Cognitive load, bit.ly/3gvmkwC
[2] Cognitive load theory in practice, bit.ly/2Rkmhd9
[3] Cognitive load theory: Research that teachers really need to understand, bit.ly/3jcw14x
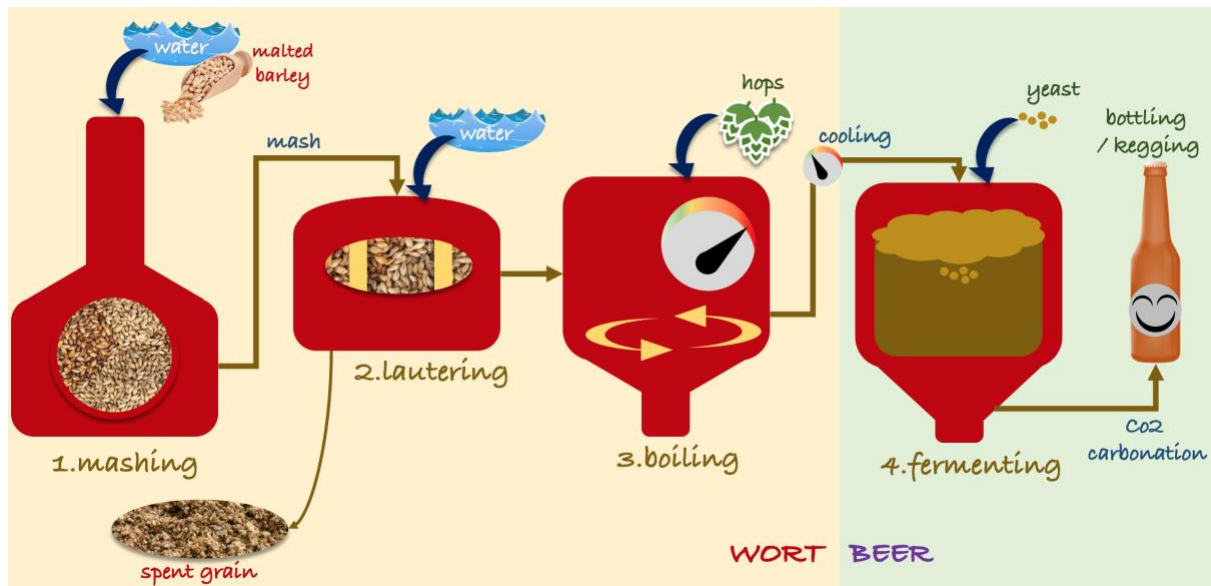[4] Dave Miller's Homebrewing Guide: Everything You Need to Know to Make Great-Tasting Beer amzn.to/3mD7lED

Figure 0-2 Making wort for healthy yeast to make beer[5]

Although the *science* of brewing involves a few keys steps the *art* of making beer is the variance on those key steps that experienced brewers learn and experiment with over time. Every beer from a dark bitter stout to crispy golden nectar of a lager involves four main steps, these are the sequential of chunks in making beer.

1. Mashing – the process of mixing malted grains (the grain bill) into a boil. Malted grains contain the starch and enzymes that through boiling is encouraged to depart from the rest of the grain, the starch is needed for the hungry yeast

2. Lautering & Sparging - rinsing of the mash grain bed to extract as much of the starchy sugars from the grain as possible. The additional water used to rinse the half-spent grain is added to the wort and the spent grain itself is discarded as fertilizer and animal feed.

3. Boiling – at different temperatures (typically around 55 Celsius or 131 Fahrenheit) and lengths of boils (60-90 minutes), bittering, flavouring and aroma hops are added to the wort which at the end of the boil becomes a syrup ready for the hungry yeast – the wort.

4. Fermenting – once the boil has completed the wort is rapidly cooled to a temperature to suite the type of yeast you will use that will eat the starch and produce alcohol as a by-product. 11-15 Celsius (51.8 – 59 Fahrenheit) for lager yeast and 17-25 Celsius (62.6 – 77 Fahrenheit) for ale yeasts. The wort is left alone in an airtight tub for 3 to 10 days.

- Kegging or bottling – after the yeast has performed its job what you are left with is the beer but at this point it is a dull liquid devoid of any bubbles. As a part of kegging or bottling the liquid is carbonated for 1 to 3 weeks before they are ready for distribution and consumption.
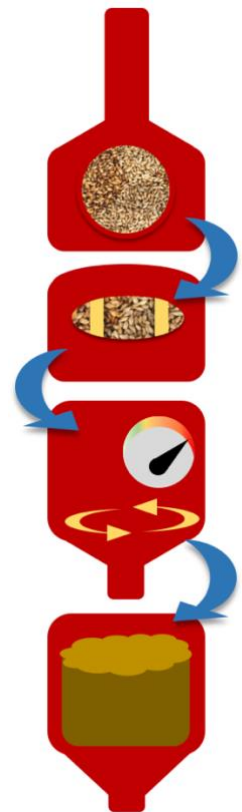
    That's it!

Variance in the steps, ingredients, temperature and time is what produces different beers, for instance a Munich Dunkel will use lager yeast, some roasted malts and Hallertau hops,

---

[5] HOW TO MAKE BEER, ACCORDING TO A BREWER, bit.ly/3jQ1Lgb

whereas an Indian Pale Ale (IPA) can use Belgian ale yeast, Pilsner malts and Saaz hops. There are more things you can do tamper with the flavour, aroma, colour and clarity of the beer however the core four steps remain the same.

This is not too different to learning data vault for the first time (or any other skill for that matter); first we establish the core 'chunks', the initial learnings are reinforced and new information is built upon those learnings and the cycle goes on gained through coaching and experience. Let's start with the first chunks,

1. When modelling a data vault we start with understanding what are the business entities that matter to the business. Like conceptual modelling we learn how the business sees their enterprise. These entities are modelled as tables called "**Hub** tables" and contain a unique list of business entity identifiers that represent a thing, entity or object. The identifier values are the immutable unique **key** that for the life of the entity in the business process does not change. That key will forever tie to that business entity.

2. Naturally after identifying the things, objects or entities we wish to model it is important to understand how they relate to each other; this is the **unit of work** and in the data vault is represented in a "**Link** table". A Link table captures the business processes present in the enterprise as the relationships between business entities is engrained in these Link tables. One-to-one, one-to-many and many-to-many relationships are stored in a link table structure, this could be a relationship between two or more entities but may be between one or more hubs as is the case of hierarchical relationships or multiple business keys representing the same entity!

3. Finally, when it comes to the descriptive content about the business entity or relationship we store that content as "**Satellite** tables". A satellite table may describe a single business entity, object or thing and another satellite may describe the relationship between two or more entities.
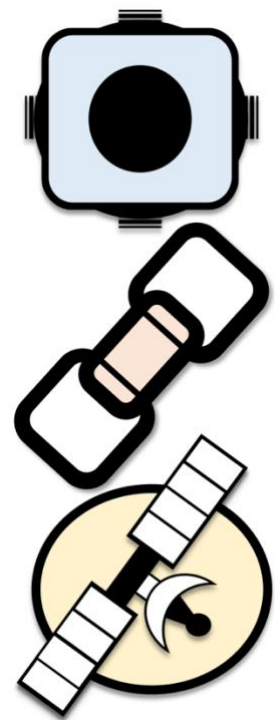
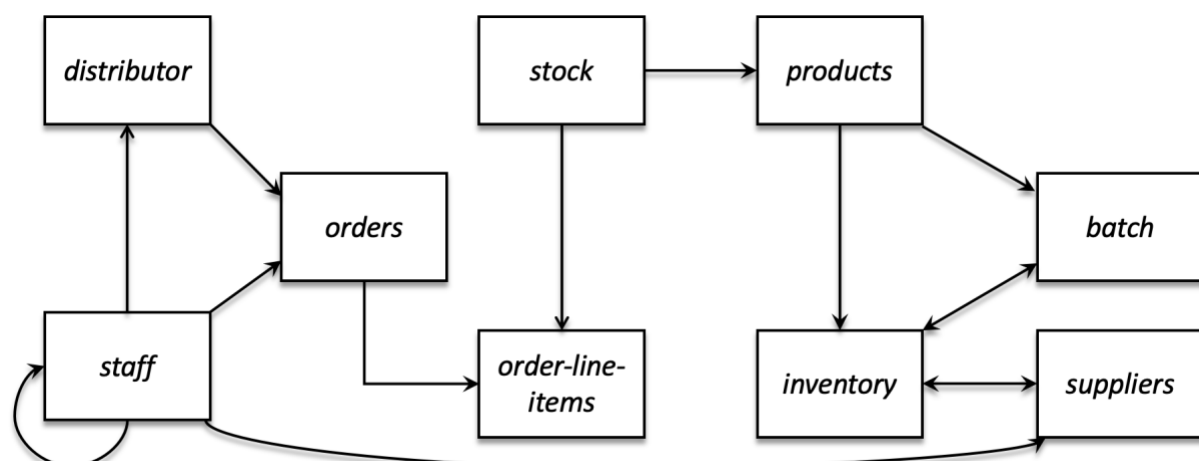We will take our beer example further by introducing a logical data model for beer production and distribution[6].

Figure 0-3 model for the fictitious brewery

---

[6] Understanding the Beer Game, bit.ly/3kRvAhM

Staff take orders from distributors and each order may have multiple line items and each line item is describing the quantity of products. Products are sold in one of two types of units, either as bottles or as kegs. For bottles the minimum spend are eight cases (24 bottles in each case) or four kegs. Each product has ingredients which are common across most of the products, the ingredients used in a batch are used to produce products and we also track who supplies the ingredients, the frequency and in what quantities. Of course the overall model would be much larger but we're not trying to boil the ocean in our use case, we want to track the history and efficiency of our supply chains to beer output. On to our modelling process!

1. Business entities

- Staff? Yes, we track name, contact details, address and so on. Details about the staff can change and change often. In the enterprise ontology we have salary details, performance and so on that tie to a staff id (business key).
- Distributor? Yes, these details can change and we track address about them too and they are tied to a distributor id (business key)
- Order? Yes, each order is tracked until fulfillment against an order number (business key)
- Order-line-item? No, a line item has no history, and order details are tracked to the order id (can be a dependent-child key)
- Product? Yes, we have stock id (business key) and track how much <u>stock</u> we have, but products themselves are referenced from a reference table by product code.
- Inventory/ingredients? Yes, each item is tracked with smart key with parts of the key denoting what type of item it is.
- Batch? Yes, each batch has a batch id (business key) and we track where the batch is in the brewing process (mash, lauter, boil, fermentation and bottling/kegging).
- Supplier? Yes, where we get ingredients, gas and parts from we track with our own supplier id.

hub_distributor   hub_order   hub_stock   hub_batch

hub_staff   hub_inventory   hub_supplier
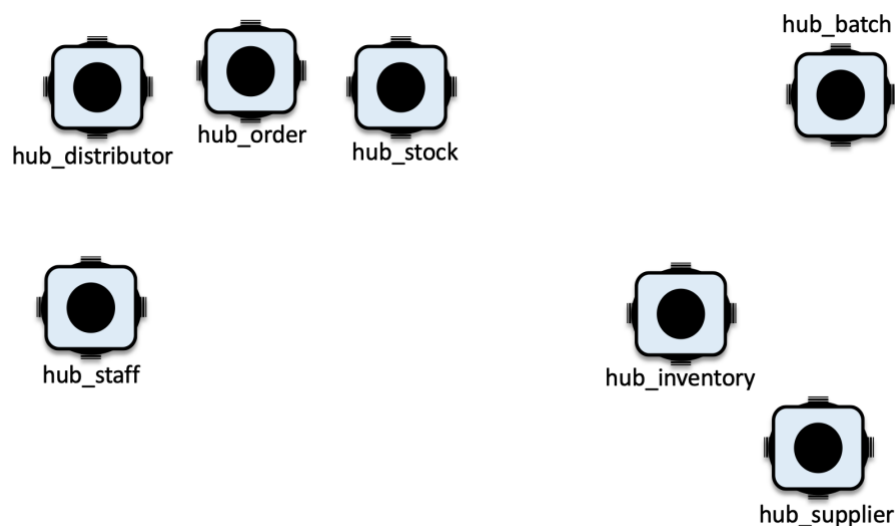
2. Relationships or unit of work

- An order is placed by a staff member for a distributor, an order has line items with quantities of each product. Although the line-item itself is an id for a different order the same line item id will more than likely be describing a different product and quantity, it is a weak key (dependent-child id)
- Staff is hierarchical in the sense that we have junior sales staff that report to account managers who in turn report to a sales director. Staff is not limited to sales either, we have staff who are involved in the brewing process as well as security staff.

- Brewing staff are involved in placing orders for ingredients and supplies for operation and maintenance, including water and electricity to run the brewing process.
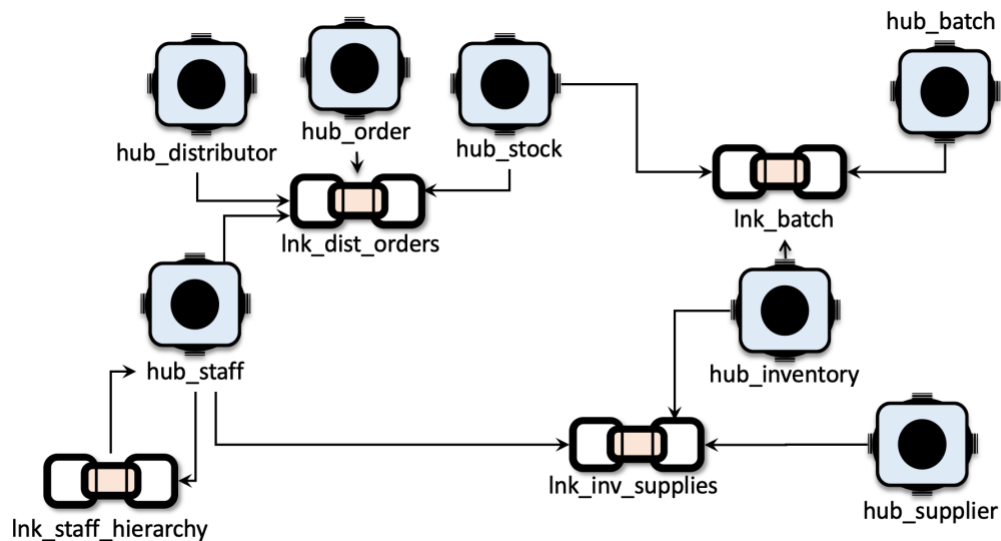
3. Descriptive content
- Staff content must be split because we capture personally identifiable information (PII) and confidential data.
- Unit of work descriptors are tracked for making orders and for order fulfillment, purchasing supplies for inventory, inventory used in a batch to produce products
- Descriptive content is also tracked on individual business entities such as contact details, stock movement and so on
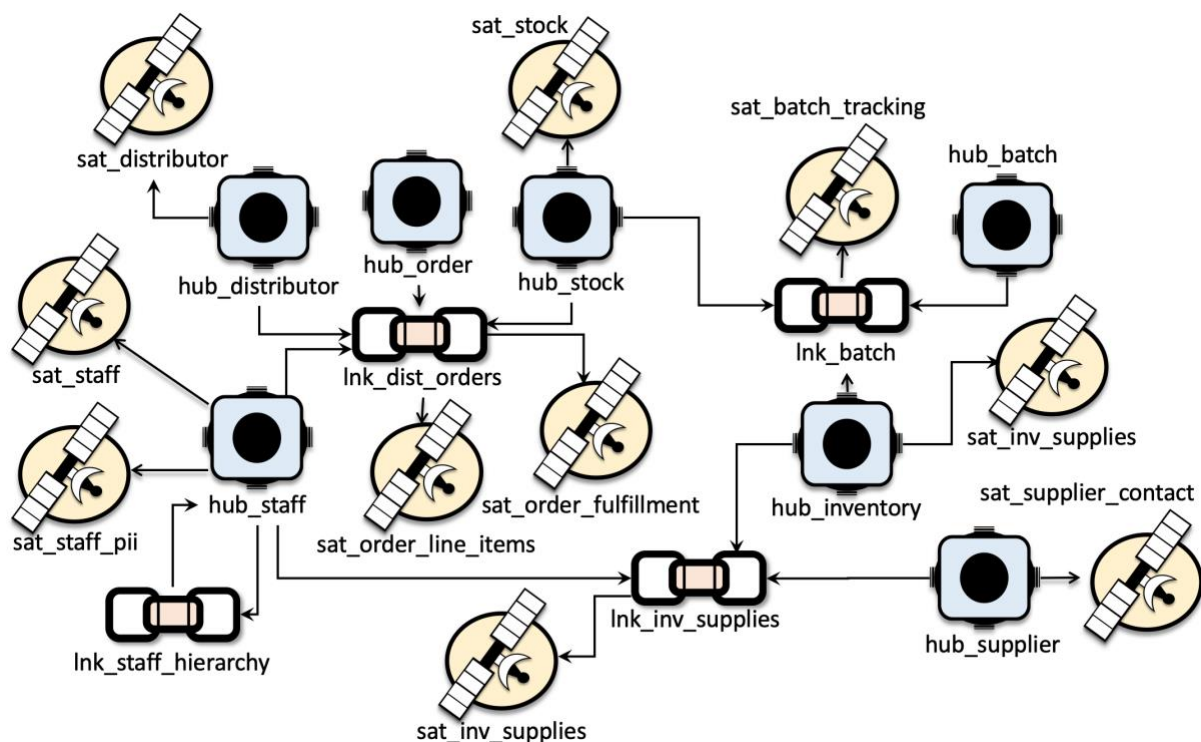
That's it! Modelling in data vault is down to three steps, all other considerations to the data vault model build upon those three core model types, hubs, links and satellites. Data

Vault 2.0 adds to the modelling of data vault by discussing the three pillars of data vault, architecture, methodology and modelling.

To summarize

- Learn the Standards and you can apply them to any jurisdiction or data platform and you can be safe in the knowledge that the data vault model will flex as the enterprise *and* industry flexes
- Agile principles apply, although there are so many variances (ingredients) to delivering a data vault; remembering how to design and build the base artefacts will get you going. You are capturing the enterprise ontology in a data vault therefore having a business analyst and subject matter experts to help you develop the data vault model is vital.
- To gain the confidence on delivering a data vault the options are available in the market
  - Get training through a certified training provider, and/or
  - Get onsite coaching, and/or
  - Get data vault review

> ***Fun Fact***: the old medieval name for yeast (because its properties for making bread and beer were shrouded in mystery) was known as ***Godisgoode***.