Data Vault Pitch

Ever find yourself having to explain Data Vault to a customer? You could take the generic approach but you know not everyone one can relate to discussions on hubs, links and satellites! Therefore, here is a list of the things I would say to a customer based on his or her role in the organization.

1. **Business executive**. Data Vault brings the proven experience of integrating your business' business process landscape into repeatable automation patterns in technology and agile practices that not only speed up delivery and business value, but are predictable, easy to change and flex and grow as your business grows while providing full audit history of your data. See: bit.ly/3dlVDJn

2. **Data modeller**. Data Vault does not replace a dimensional model, instead a data vault is best suited to model the data warehouse as it maps the enterprise landscape to the enterprise ontology. Dimensional models still have their place but to best serve data presentation that often conforms to Business Intelligence tools or other reporting or dashboard requirements. Data Vault provides the audit history in a non-conformed way whereas your dimensional model will conform to present the data to the business and in that respect you can always recreate the source application at any point in time. See: bit.ly/2JJ3ozH and bit.ly/3rfV7V3

3. **Enterprise architect**. The Data Vault chiefly maps your enterprise ontology to the Data Warehouse, making your enterprise ontology data driven. We identify what those business entities are whether that be customer, accounts, products etc and the Data Vault will map those as **Hubs**, the relationships between those business entities are recorded in **Links** and the descriptive details (enterprise history as business process outcomes) is tracked in **Satellite** tables around these hubs and links. The Data Vault model represents what we know about the business' application and data integration landscape and is agnostic to the platform it is delivered on, all in all it really represents what the business architecture desire is and ingrains data governance best practices and contains the necessary components to address data privacy concerns. See: bit.ly/2V6TlYf

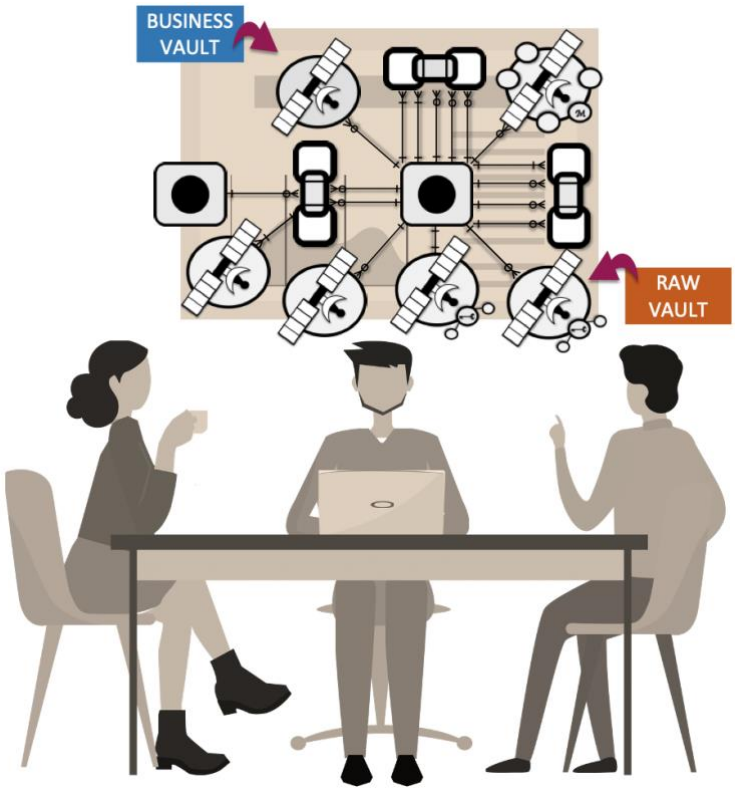
4. **Product owner**. Data Vault makes the development of enterprise-tied analytics ***accelerated*** because it gives you the ***confidence*** that what you are building is relevant to the business by using repeatable established patterns and a proven track record of integrating hundreds to thousands of data sources into a ***single source of the facts***. We can show this confidence by showing you a ***steel thread*** of an implementation, from landed file to analytic value through repeatable patterns that

after establishing the standards the ***cadence*** in delivery ***increases*** as we model more of the data in. See: bit.ly/2CqZTKJ

5. **Scrum master**. Data Vault provides the same loading patterns for any use case needed to be modelled and loaded to an enterprise data warehouse, we identify what we want to model, split into chunks of the model that we model out into sprints that will themselves follow a repeatable pattern. See: bit.ly/2NiceTN

6. **Business analyst**. Think of the Data Vault as a place where we record the history of the business processes as it pertains to the relevant parties and entities to the business, in here you only need to understand what the business entities are and how they relate to each to be able to deliver data analytics as it relates to the business. When we add more data sources into the data vault there is no change to the existing data vault artefacts, only (if needed) the addition of new artefacts if there are data and metrics not yet available in the data vault. See: bit.ly/31bx6UR

7. **Solution architect**. Data Vault differs to Kimball or Inmon in that we establish the loading patterns and rules for hubs, links and satellites (our raw tables) that comes from our source applications and load to ***raw vault*** and fill the technical and application gaps in ***business vault*** to establish a full enterprise view of the data landscape of the enterprise. From there with the data warehouse populated by these three simple table types, the information marts are delivered on top of the data vault, the data vault provides the enterprise audit history which makes our information marts disposable. See: bit.ly/371PykS and bit.ly/3ccwMcq.



8. **Data(Ops) engineer**. Once the loading patterns are established the non-modelling point of view of Data Vault is the ***repeatability*** of those loading patterns; once the landed files are staged (adding hash-keys, record source columns, applied date, load date) that staged file will load one or more hub tables, zero or more hub-satellites, zero or more link tables or zero or more link-satellites. That's it! Just load ***three table types*** and the rest of the tasks around managing a data vault is around scheduling, task management and so on. The data vault modellers you support are left to model new data into the data vault using those repeatable loading patterns and because they are repeatable loading patterns we have repeatable testing patterns as well. See: bit.ly/35xhVGL

9. **Data scientist**. When looking holistically at the data vault the data model resembles a mathematical hyper-graph with ***nodes*** and ***edges***; nodes are our business entities represented as ***business entity hubs*** and edges are the ***multi-node relationships*** represented as links. This gives us the business reasons, rules and processes mapped to the data warehouse to support the business service. In this enterprise ontology associated data modelling structure we attach descriptive historised content about nodes and edges which can open the opportunity to learn about what we do well and what we can improve by designing new ***neural networks*** over the established business application landscape. See: bit.ly/3ezZ6Wh

10. **Source-system subject matter expert**. We are interested in the business keys, their relationships and the grain of the source data, we will never boil the ocean and model your entire platform, only the tables and entities that are a part of the business case we're modelling.

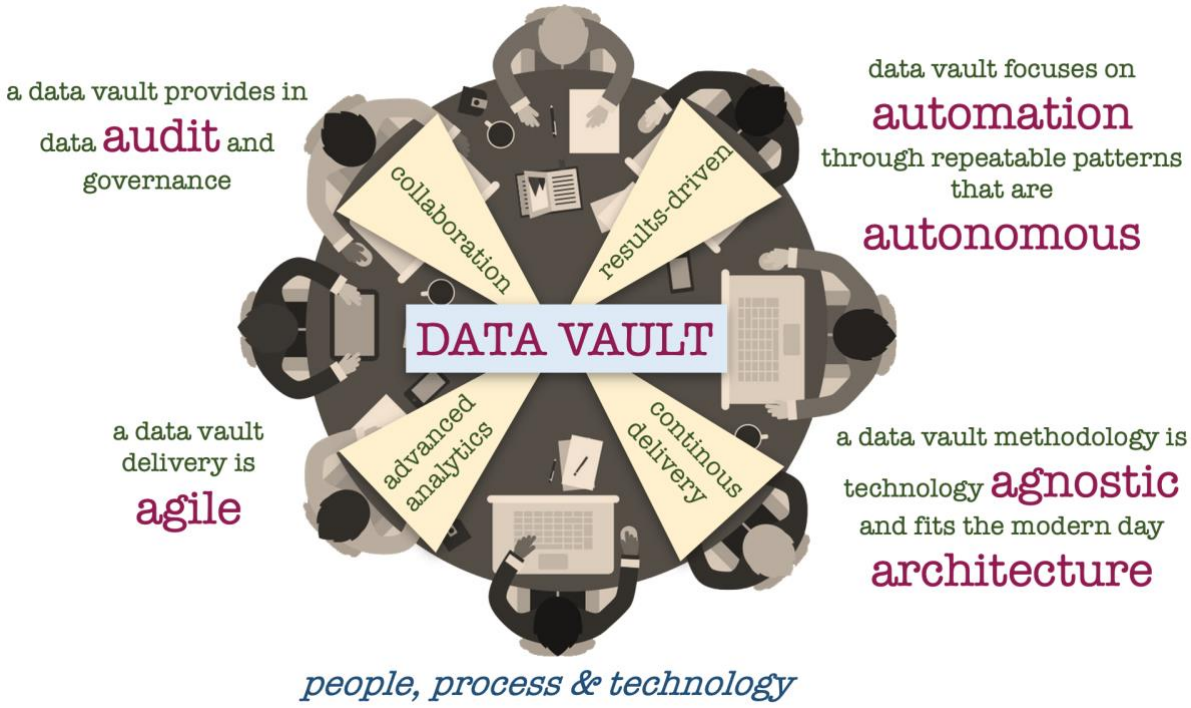
In addition to few sentences we have discussed above we can also mention Data Vault is:

- **Agile** – an enterprise model that grows ***horizontally*** as you integrate more data sources to it and grows ***vertically*** as those data sources push or pull data into it. Yes changes are ***non-destructive*** to the existing data vault model and multiple data squads can work on the same model and therefore centralizing the source of the ***facts***
- **Automated** – embraces ***DataOps***, only three table loader types and as data is loaded it is tested and this happens at any time of day as data arrives it is loaded and therefore your enterprise data vault model is in constant growth ingesting everything important about your business in an ***eventually consistent*** data vault model.
- **Auditable** – ***all the data all the time***, dimensional marts are conformed and disposable and the data vault is not! Conformed to ***data retention*** rules by jurisdiction of course, that data vault is your ***corporate history***!

Data Vault also has patterns for ***data-driven timeline correction***! Yes, as the data vault captures business entity changes and issue might arrive when data is loaded out of order. With this pattern this is no longer the case, see: bit.ly/3dlVDJn
In addition Data Vault also has patterns for tracking the business entities themselves that sometimes are not supplied by the source system explicitly, ***last seen date***, ***status*** of an entity key or relationship and the ***driver key*** to non-driver-key effectivity!

To learn more about Data Vault check out these additional resources:

- Data Vault Mysteries... Zero Keys & Ghost Records, url
- Seven Deadly Sins of Fake Vault, url
- Data Vault Mysteries... Business Vault, bit.ly/3rfV7V3
- The Data Vault Guru, amzn.to/3d7LsJV, amzn.to/3nsqTfR, amzn.to/30IxOYF
- GitHub: thedatamustflow, bit.ly/31bx6UR
- Data Vault Alliance, datavaultalliance.com



In closing… with collaboration and an effective business sponsor you can get the ball rolling by delivering a ***steel thread*** of a data vault implementation… a steel thread is a ***prototype*** that is the minimum size required to completely implement the main elements of a design. In other words, it is a vertical prototype that dives deep into the technical implementation of the design but offers only a ***thin horizontal slice*** of functionality.
How do we do a steel thread for a data vault?

1. Establish a council to oversee the standards
- data vault table and column names
- surrogate-hash or ***natural key*** vault
- hash collision strategies and ***namespace***
- ***business rule*** management
- ***data quality*** framework

2. Design your infrastructure
- Platform hardware on cloud, strengths & weaknesses and data warehouse/lake
- ***Build or buy*** ELT/ETL/DV2.0 automation tool
- Separation of concern, database, schema, roles, secure/non-secure content

3. Get ***training***/***coaching***/skilled people
- on the automation tool

- on Data Vault 2.0

4. Decide on a ***Use Case***
5. Identify Mob Modelling participants

- identify business entities and unit of work for raw + business vault
- extract/information mart/report/dashboard/analytics

6. Set the ***key performance indicators (KPI)*** to measure cadence

Steps 1-6 establishes the steel thread, building out the enterprise data vault only needs the selection of the next use cases (steps 4 & 5) that can be executed in parallel