

Link Applications

Summary ▾

- In general, the different application of links have a very similar structure or definition, what changes the most is the purpose or the concept.
- One link can be more than one type at a time. For example an exploration link can be a same-as link at the same time.

⚠ Do not create link-to-link references >

Link-to-link relationships are forbidden in Data Vault modeling because it is not scalable and requires a redesign of the model y a new requirement associated to the relationships comes up.

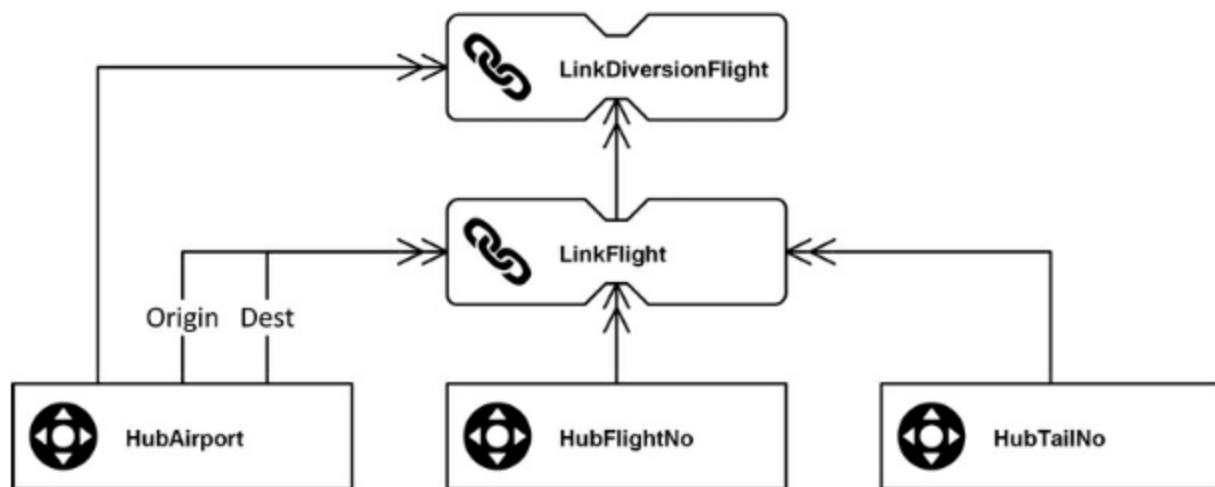


FIGURE 5.4 Link-on-link Data Vault model (logical design).

The recommended procedure is to create independent links...

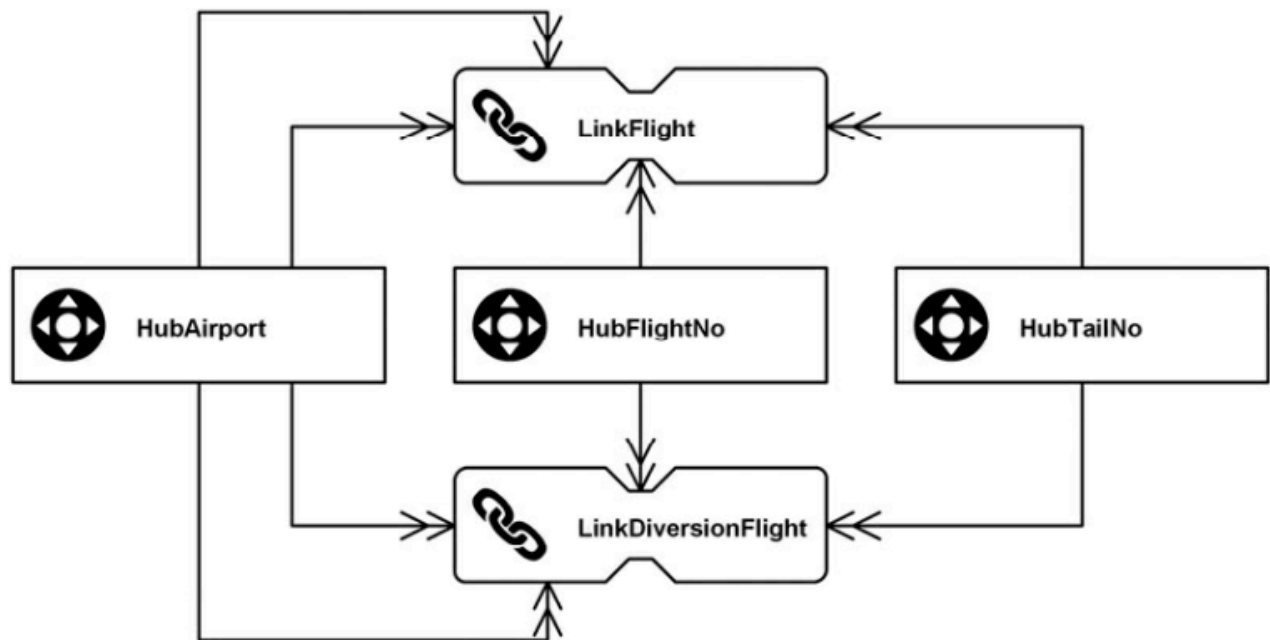


FIGURE 5.5 Independent links (logical design).

Same-as Links (**SAL**)

- Used to indicate duplicate business keys within a hub entity.
- Required when the same business objects are identified by more than one BK.
- The business source has to provide a mapping table showing the 1:1 relationship between the primary keys of the two different source systems.
- The hub entity will store at least two records per real entity so the link is capable of creating the mapping of the source systems.

	Same-as Link	
PK	SALHashKey	Same as in Link definition
	LoadDate	Same as in Link definition
	RecordSource	Same as in Link definition
FK	<ref_hub>MasterHashKey	Reference to the PK of the hub record associated to the primary source system.
FK	<ref_hub>DuplicateHashKey	Reference to the PK of the hub record associated to the secondary source system.

Hierarchical Links

- Used to represent hierarchies within the same hub.

- The structure of the link is essentially the same as the Same-as Link, what varies the most is the concept.
- The hub stores all the parent and child components, and the link creates the hierarchy.
- Example. The components of an airplane: the engine is ensemble of many smaller elements, same as the propeller but all the components are `mechanicalParts` of the airplane with similar characteristics (meaning, same hub).

	Hierarchical Link	
PK	<code>HashKey</code>	Same as in Link definition
	<code>LoadDate</code>	Same as in Link definition
	<code>RecordSource</code>	Same as in Link definition
FK	<code><ref_hub>ParentHashKey</code>	Same as in Link Applications > Same-as Links (`SAL`) The difference is that now we are storing a parent instead of a master.
FK	<code><ref_hub>ChildHashKey</code>	Same as in Link Applications > Same-as Links (`SAL`) The difference is that now we are storing a child instead of a Duplicate.

Non-historized links

- Cannot edit or remove records.
- If a record has to be edited or removed, a new reversing or editing record has to be created.
- Cannot have satellites attached to it in logical design but there are ways to add satellites in the physical implementation. The attributes that are part of the record can be added directly in the non-historized link.
- These links are characterized by **entities that cannot be altered due to its business nature**. For example, an airplane ticket will always have the same attributes, if I wish to edit the date of my flight the airline will cancel my current ticket and generate a completely new ticket.
- There are two ways to physically model a non-historized link:
 1. Standard link entity + satellite without a `LoadEndDate` attribute. It is not possible to insert new versions of the records in the satellite because it is not possible to end-date the record and place a new version.
 2. Add the attributes of the transaction directly to the link structure and abstain from using a satellite structure at all. **This option should be avoided in most cases.** It can

be used in cases where performance is an extremely high priority (loading data in milliseconds).

Non-descriptive links

- Links characterized by being focused only on the connection it is creating.
- These links do not require to describe the relationship, hence no satellite is needed.

Computed aggregate links. #developing

- These kind of links exist when a link has data that is calculated or aggregated.
- There are two approaches to model these links. The chosen one is defined by the answer to the following question: ***Is the link aggregated data calculated from the incoming data of the source system or is it calculated from the raw data using business logic?***
 1. It is calculated from the source system. Then, the data is modeled as a raw link with an attached computed satellite (containing the aggregated data).
 2. It is calculated from the raw data, meaning the link is a Business Vault entity. The data can be stored directly as a link attribute.

Exploration links.

- They are born for business reasons only. For example, when the business wishes to explore the data coming from a relationship that it does not exist, currently; so they decide to create an exploration link for exploration purposes.
- It has the standard link structure. What is unique, is the concept.
- Some other reasons to create exploration links:
 - Determine relationships between business entities (hubs) that are not in the source system.
 - Represent relationships that are only indirect in the current model.
 - Consolidate links between business objects if one of the referenced hubs contains duplicate entries (same-as links).
 - Identify clusters of similar entries within hubs (using same-as links).
 - Automate discovery of patterns, for example in fraud detection.