

Data Vault Rules v1.0.9 Cheat Sheet

Hub

A **Hub** is a list of unique business keys with low propensity to change

Required fields

- **Key 1: At least one (composite) Business key**
- DataVault Load date/time
- Record source

Optional fields

- **Key 2: (Hub) Sequence ID**
- (Manual) Update user, Update date/time
- Source system extract date
- Last Seen in DataVault date

Rules

- No composite set of business keys (*)
- Should have at least 1 satellite
- Business keys SHOULD be true ID's of business entities (*license plate, SSN, ordernumber, etc.*)
- Business keys CAN be synthetic source-system created keys
- Business keys MUST stand alone

(*) This is not the same as "no composite keys"

Link

A **Link** is a unique list of (n to n) relationships between business keys

Required fields

- **Key 1: Two or more Hub or Link keys**
- DataVault Load date/time
- Record source

Optional fields

- **Key 2: (Link) Sequence ID**
- (Manual) Update user, Update date/time
- Source system extract date
- Last Seen in DataVault date

Rules

- A link indicates (timeless) existence of a relationship
- The imported keys composite must be unique
- We can end-date links using an 'exists'-attribute in a satellite
- Satellites are optional, except for hierarchical links
- Links exist at the lowest level of detail for the imported entities

Additional rules for hierarchies

- Hierarchical relationships are implemented using a link between the keys of ONE hub, migrated in as both Parent and Child
- End-dating links in a Satellite is mandatory
- The Child-key determines the end-of-life of the relationship

Satellite

A **Satellite** contains the historical data associated with the Hubs and Links

Required fields

- **Key 1a: One Link or Hub key (or sequence ID)**
- **Key 1b: DataVault load date/time**
- Record source
- DataVault Load **End** date/time
- At least one descriptive field about the Key

Optional fields

- (Manual) Update user, Update date/time
- Source system extract date
- Last Seen in DataVault date
- **Key 1c: Sequence ID** for uniqueness in the same key

Rules

- Contains all non-key data
- Attached to exactly one Hub or Link
- Has at least one descriptive field
- Data is grouped into separate satellites according to type, rate of change and source
- References (not foreign keys) to **stand-alone Reference** tables or **single Hub/Satellite combinations** are allowed
- Insert 'default' satellite rows for new hub/link keys without data to avoid outer joins

Reference

A **Reference table** is an a-historical lookup table

Required fields

- **Key: Reference(d) code**
- One or more attributes

Rules

- Ref-tables are stand-alone (sets of) tables that are used throughout the model for lookups
- Ref-tables can have any form: DV, 3NF, Kimball, etc.
- Stand-alone tables that contain history should be modelled as Hub/Link/Satellite structures

Naming

Recommended naming conventions

Entities

- Hubs: HUB or H prefix or suffix
- Links: LINK or L prefix or suffix
- Hierarchical links: HLNK or HIER or HL prefix or suffix
- Same-As links: SLNK or SAL or SA prefix or suffix
- Satellites: SAT or S prefix or suffix
- Reference tables: REF or R prefix or suffix

(Prefix to sort on table type. Suffix to sort on business concept)

Fields

- Record source: REC_SRC or RECORD_SOURCE or prefix/suffix with RCSRC or RSRC
- Sequence ID's: SEQ_ID or SEQUENCE_ID or prefix/suffix with SQN
- Date/time stamps: prefix or suffix with DTS
- Date stamps: prefix or suffix with DT
- Time stamps: prefix or suffix with TM
- Load Date/Time stamps: prefix or suffix with LDDTS
- Load End Date/Time stamps: prefix or suffix with LEDTS
- User watch fields: prefix or suffix with USR