

Data Vault Rules v1.0.8 Cheat Sheet

Hub

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

Required fields

- **PK: (Hub) Sequence ID**
- DataVault Load Date/Time
- Record source
- At least one Business key

Optional fields

- Manual update user
- Manual update timestamp
- Sourcesystem extract date
- Last seen in DataVault date

Rules

- No composite set of business keys (*)
- 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 list of (n to n) relationships between business keys

Required fields

- **PK: (Link) Sequence ID**
- DataVault Load Date/Time
- Record source
- At least two "imported" Hub or Link Sequence ID's

Optional fields

- Manual update user
- Manual update timestamp
- Sourcesystem extract date
- Last seen in DataVault date

Rules

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

Additional rules for hierarchies

- Hierarchical relationships are implemented using a link between exactly two hubs: the Child and Parent Hubs
- End-dating links 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

- **PK: Link or Hub Sequence ID**
- **PK: DataVault Load Date/Time**
- Record source
- DataVault Load **End** Date/Time

Optional fields

- Manual update user
- Manual update timestamp
- Sourcesystem extract date
- Last seen in DataVault date
- **PK: Sequence ID** for uniqueness

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

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

Rules

- Ref-tables are stand-alone tables
- Ref-tables do not store historical data
- Ref-tables are used for performance improvement or lookups on the way to the reporting layer/data mart

Naming

Recommended naming conventions

Entities

- Hubs: HUB or H prefix or suffix
- Links: LINK or L prefix or suffix
- Hierarchical links: HLNK or HIER 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