



Data Warehouse & Business Intelligence Fundamentals

Todor Kichukov

todor.kichukov@bipartner.biz

<https://www.facebook.com/groups/SUDWBI2022/>

Faculty of Mathematics and Informatics

Sofia University

2022

Data Warehouse & Business Intelligence Fundamentals

Course Scope

- DW Concept
- DW Architecture
- DW Data Modeling
- Data Integration
- Gathering and Analyzing Requirements
- Business Intelligence
- Deployment, Support and Maintenance

Data Warehouse Data Modeling Part III

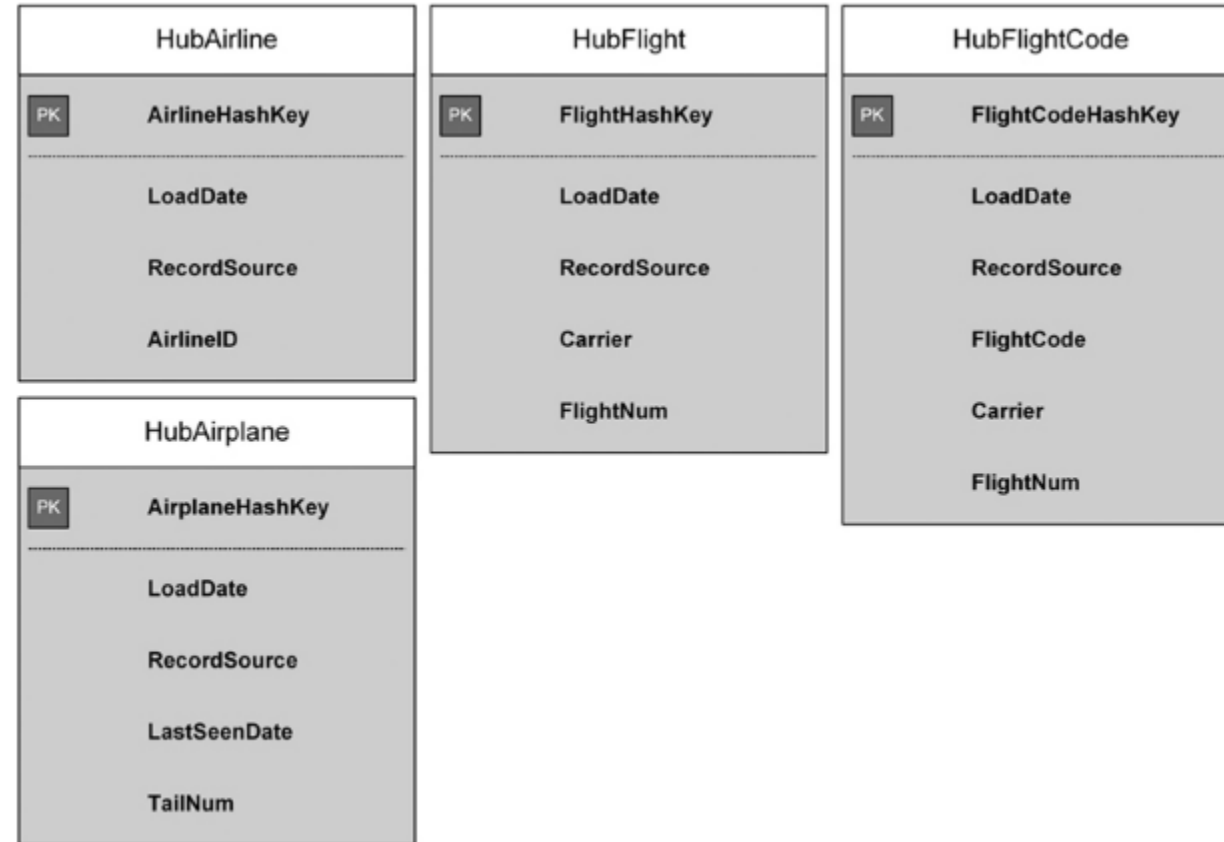
- Data Vault Entity Types
- Hub, Link, Satellite in details
- Usage Specifics
- Reference Tables
- Terminology
- Project Work

Data Vault Basic Entity Types

- **Hub** – a unique list of business keys.
- **Link** – a unique list of relationships (intersections) between two or more business keys (hubs). Responsible for modeling transactions, associations, hierarchies; doesn't capture timelines
- **Satellite** – keeps the content at a given time or over a time period - the attributes of the business keys or relationships (hubs or links).

Hub

- **HashKey** – PK, support linking to other data sources (NoSQL)
- **LoadDate** – Timestamp (same for all data in the same batch)
- **RecordSource** – SYSTEM.MODULE.OBJECT
- **BusinessKey(s)** – CompositeKey, SmartKey



Hub (examples)

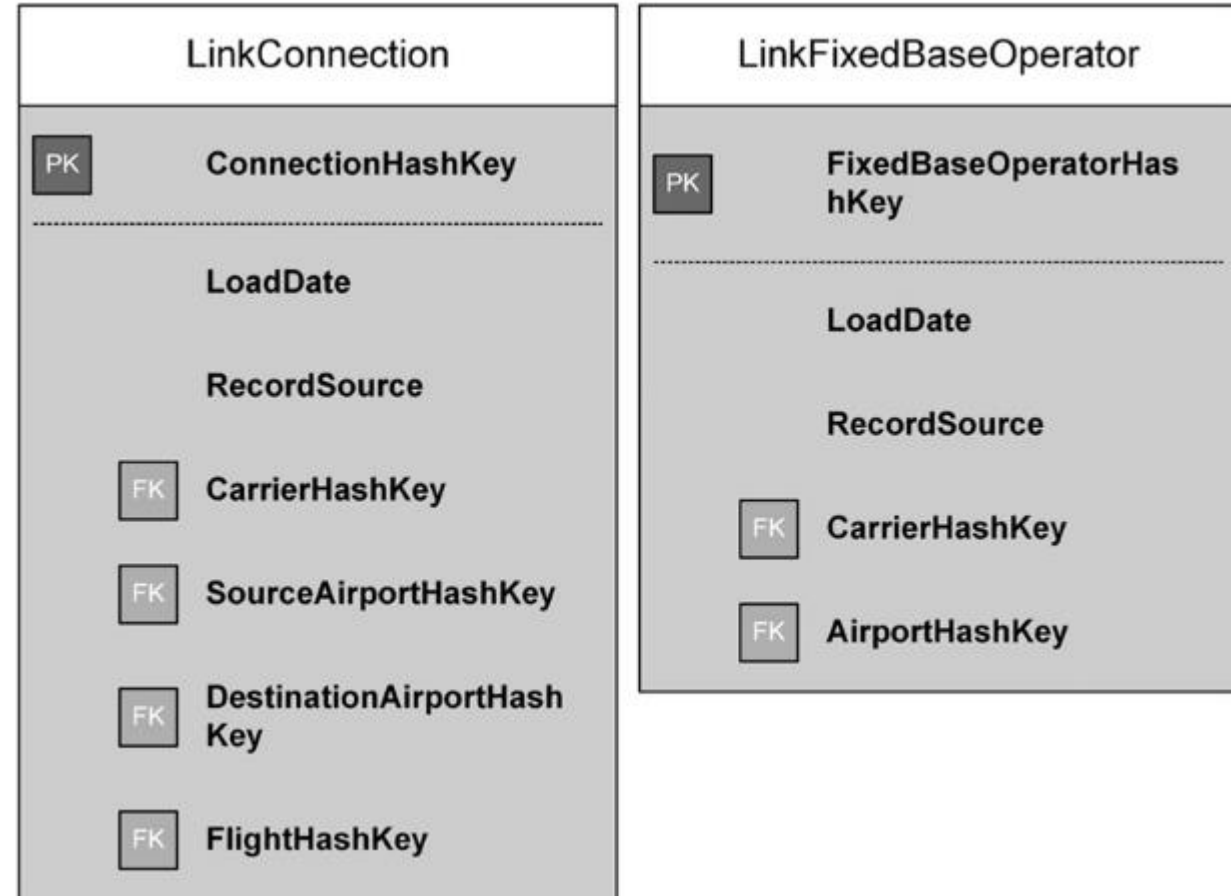
H_R_ACCOUNT ▲	
P *	R_ACCOUNT_HASH_KEY
	LOAD_DATE
	RECORD_SOURCE
*	IBAN

H_DOCUMENT ▲	
P *	DOCUMENT_HASH_KEY
	LOAD_DATE
	RECORD_SOURCE
*	DOCUMENT_ID

H_CUSTOMER ▲	
P *	CUSTOMER_HASH_KEY
	LOAD_DATE
	RECORD_SOURCE
*	CUSTOMER_NUMBER

Link

- **HashKey** - PK
- **LoadDate**
- **RecordSource**
- **Entity1HashKey** – FK
- ...
- **EntityNHashKey** - FK
- **(DependentChildKey)** - Degenerate Field



Link (examples)

L_R_ACCOUNT_CUSTOMER ▲	
P *	R_ACCOUNT_CUSTOMER_HASH_KEY
	LOAD_DATE
	RECORD_SOURCE
F *	R_ACCOUNT_HASH_KEY
F *	CUSTOMER_HASH_KEY

L_ORGANIZATION_HIERARCHY ▲	
P *	ORGANIZATION_HIERARCHY_HASH_KEY
	LOAD_DATE
	RECORD_SOURCE
F *	PARENT_ORGANIZATION_HASH_KEY
F *	CHILD_ORGANIZATION_HASH_KEY

L_CONTACT_PREFERENCE ▲	
P *	CONTACT_PREFERENCE_HASH_KEY
	LOAD_DATE
	RECORD_SOURCE
F *	PARTY_HASH_KEY
F *	CONTACT_POINT_HASH_KEY
	CONTACT_PREFERENCE_ID

Satellite

- **ParentHashKey** – PK, FK
- **LoadDate** – PK - Timestamp
- **LoadEndDate** (*not used after DV 2.0.1!*)
- **RecordSource**
- **(HashDifference)**
- **(ExtractDate)** – for reference when load from flat files, etc.
- **Attribute1**
- **AttributeN**

SatAirport	SatAirportTZ	SatConnection
<div><div>PKFK</div>AirportHashKey</div>	<div><div>PKFK</div>AirportHashKey</div>	<div><div>PKFK</div>ConnectionHashKey</div>
<div>PK</div> LoadDate	<div>PK</div> LoadDate	<div>PK</div> LoadDate
LoadEndDate	LoadEndDate	LoadEndDate
RecordSource	RecordSource	RecordSource
HashDiff	GMTOffset	CRSDepTime
Name		DepTime
RunwayLength		DepDelay
RunwayElevation		TaxiOut
Longitude		WheelsOff
Latitude		WheelsOn
Telephone		TaxiIn
Fax		CRSArrTime
Email		ArrTime
Website		ArrDelay

Satellite (examples)

S_R_ACCOUNT_MID_MAIN ▲
PF * R_ACCOUNT_HASH_KEY
P * LOAD_DATE
RECORD_SOURCE
HASH_DIFFERENCE
RETAIL_ACC_NUM
ACC_SEQ_NUMBER
DATE_ACC_CLOSED
DATE_ACC_OPENED
* BRANCH_CODE
SWIFT_CCY

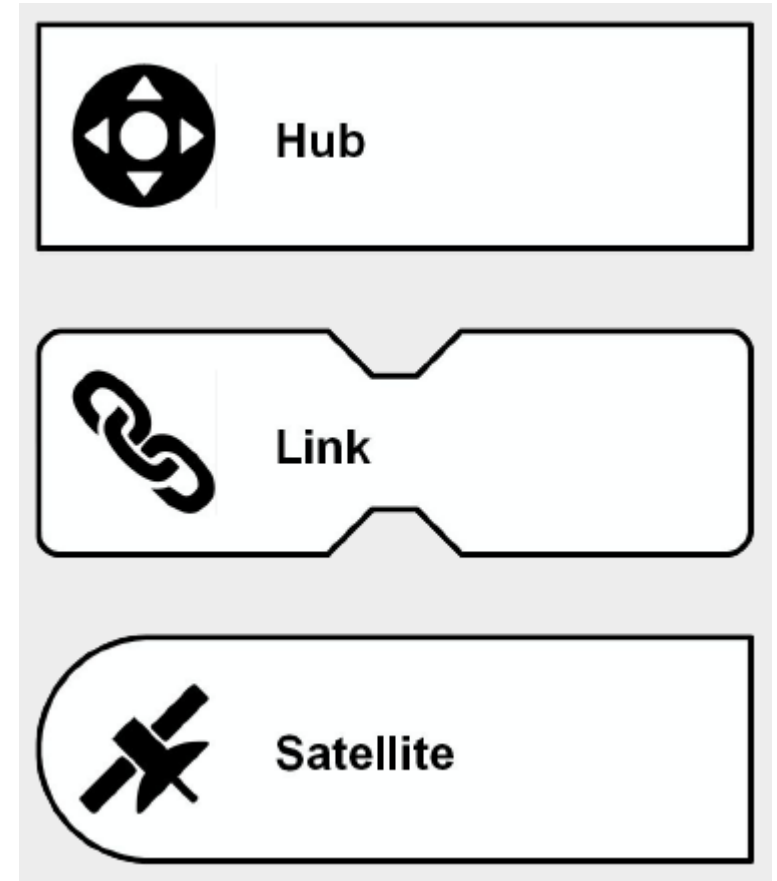
S_R_ACCOUNT_MID_CALC ▲
PF * R_ACCOUNT_HASH_KEY
P * LOAD_DATE
RECORD_SOURCE
HASH_DIFFERENCE
MIN_BALANCE_LC
OD_LIMIT_LC
DISPO_BALANCE_OC
DISPO_BALANCE_LC
CLEARED_BALANCE_LC
LEDGER_BALANCE_LC
FLAG_ELBA_INTERNATIONAL
FLAG_UNATHOURIZED_EXCHANGE
FLAG_SWEEPING_ORDER
FLAG_INTEREST_COMPENSATION
FLAG_STANDING_ORDER

S_R_ACCOUNT_MID_ADDITIONAL
PF * R_ACCOUNT_HASH_KEY
P * LOAD_DATE
RECORD_SOURCE
HASH_DIFFERENCE
ADDRESS_REFF
ACC_NAME
CT_RATE_CHANGE_DATE
CT_CAPITAL_DAY
CT_INT_BASE_TYPE
CT_INT_CALC_BASIS
CT_INT_FREQUENCY
CT_INT_RATE_SPREAD
DT_RATE_CHANGE_DATE
DT_CAPITAL_DAY
DT_INT_BASE_TYPE
DT_INT_CALC_BASIS
DT_INT_FREQUENCY
DT_INT_RATE_SPREAD
HELD_AMOUNT
LAST_CHANGE_DATE
MIN_BALANCE_OC
NEXT_CT_INT_CAPT_DATE
NEXT_DT_INT_CAPT_DATE
NEXT_STMENT_DATE
OD_LIMIT_EXPIRY_DATE
OD_LIMIT_OC
OD_SECURED
CT_RATE_CHG_VAL_DATE
DT_RATE_CHG_VAL_DATE
RECORD_STATUS ▼

S_POST_ADDRESS_CIF ▲
PF * CONTACT_POINT_HASH_KEY
P * LOAD_DATE
RECORD_SOURCE
HASH_DIFFERENCE
POST_ADDRESS_TEXT_FREE
DOOR
FLOOR
ENTRANCE
BLOCK
BUILDING
STREET
QUARTER
CITY
POSTAL_CODE
MUNICIPALITY
REGION
COUNTRY_CODE
CITY_TYPE
ADMINISTRATIVE_DIVISION_ID
POPULATED_PLACE_ID
APARTMENT_TYPE_ID
STREET_TYPE_ID
QUARTER_TYPE_ID

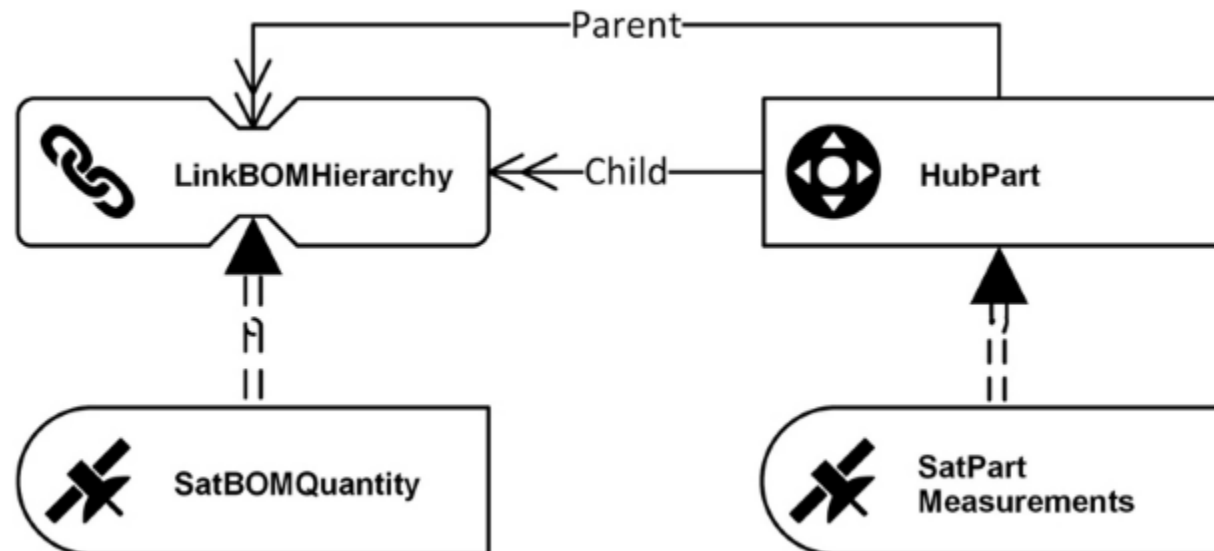
Visual Data Vault – Logical Modeling Language

- <http://www.visualdatavault.com/>
- Microsoft Visio Stencil



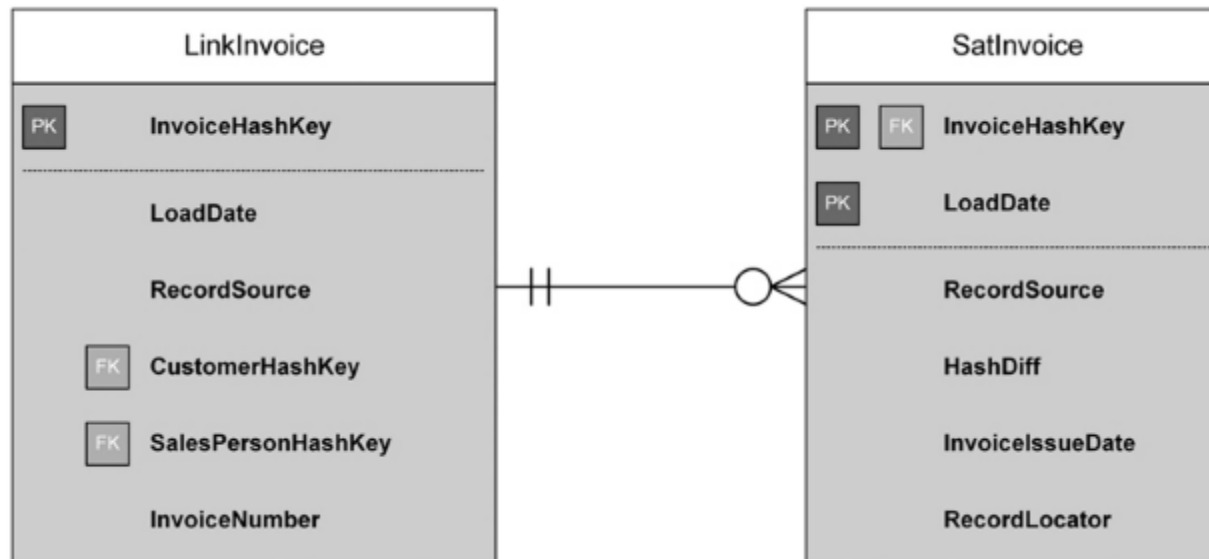
Hierarchical Links

- Parent-child hierarchies
- Level-based hierarchies



Non-historized (Transactional) Links

- The Link Satellite doesn't have LoadEndDate
(the example is from DV 2.0, anyway it is not used after DV 2.0.1!)



Non-descriptive Links

- Only the relationship between the entities is important



Reference Tables

- Code and description lookup structures.
- Resolution occurs through queries at run time. They do not house nor require foreign keys. In general, the codes (natural keys) are found housed in the Satellites as they typically describe other keys or other relationships.

Terminology

- Data Vault
- Hub, Link, Satellite

Project Work

- Review
 - Did you download the provided info?
 - Did you install the data modeling tool?
 - Did you analyze the source model and the data?
- Today
 - Clarify all open questions
- Next steps
 - Analyze the model and data
 - Understand their specifics
 - Start implementing the CBS (source) logical model (on paper or electronically)