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

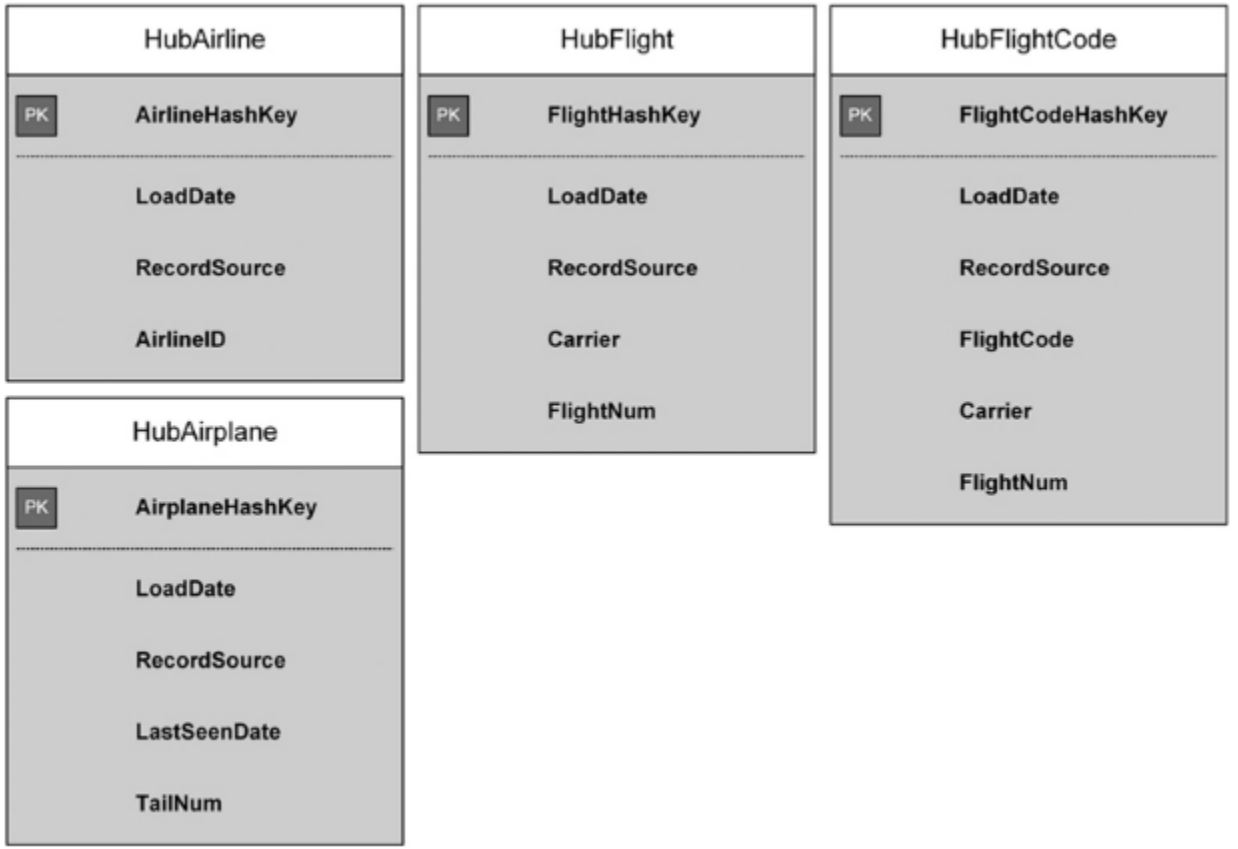
- Data Vault Entity Types
- Hub, Link, Satellite in details
- Usage Specifics
- Reference Tables
- Terminology
- Project Work

Data Vault Basic Entity Types

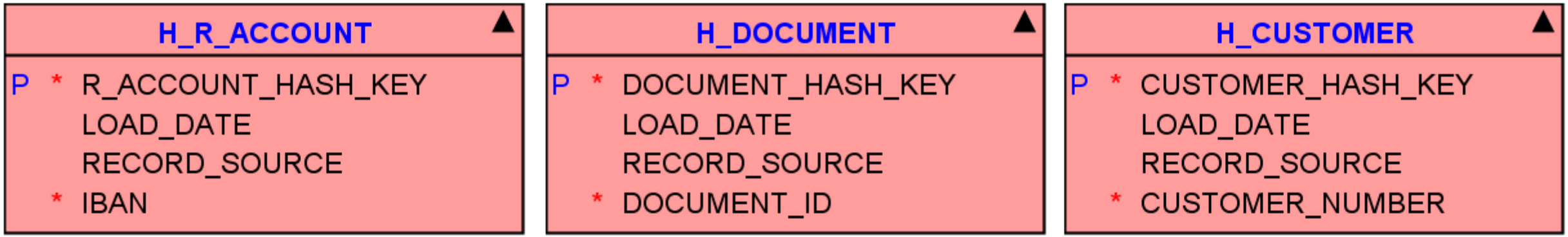
- **Hub** – a unique list of business keys.
- **Link** – a unique list of relationships (intersections) between two or more business keys (hubs). Responsible for modeling transactions, associations, hierarchies; doesn’t capture timelines
- **Satellite** – keeps the content at a given time or over a time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

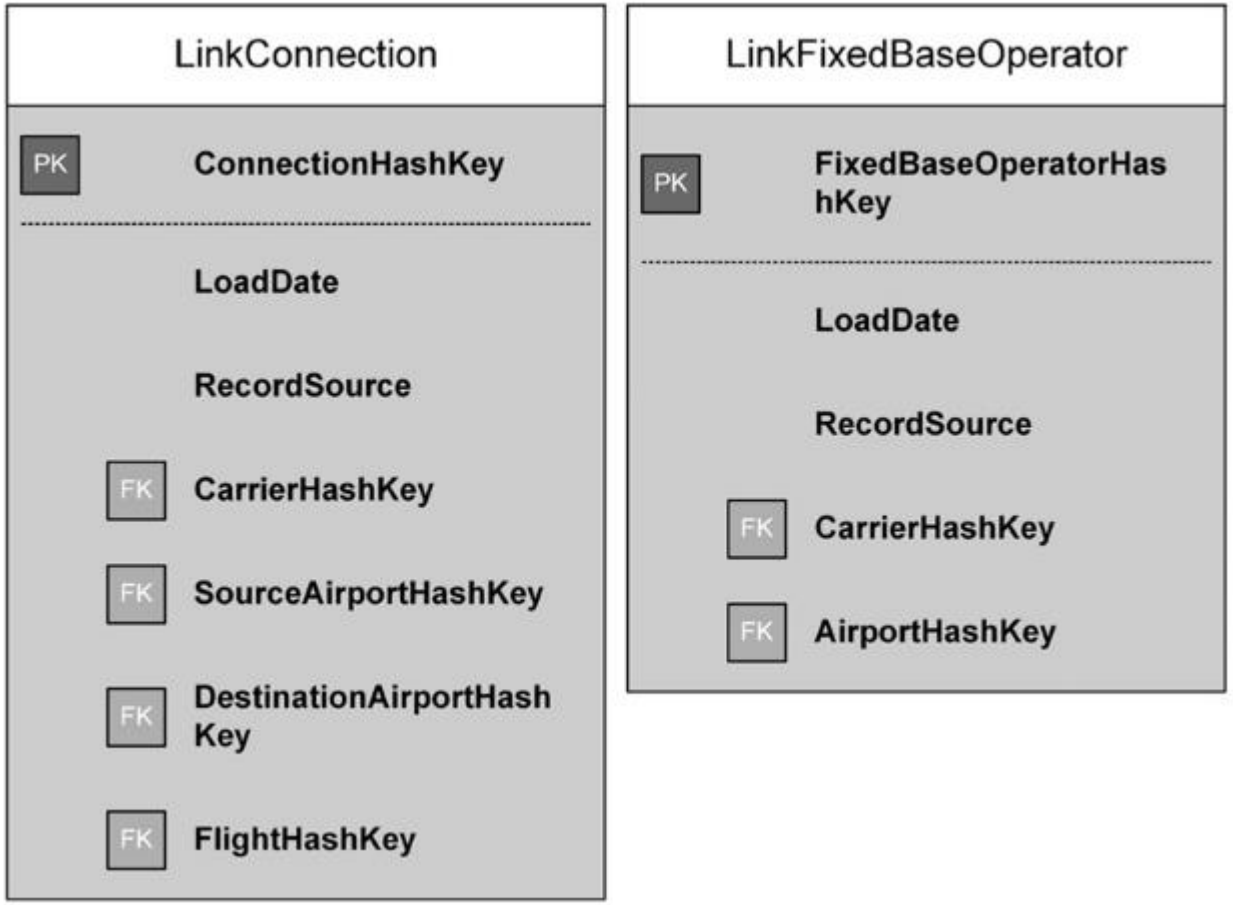


Link

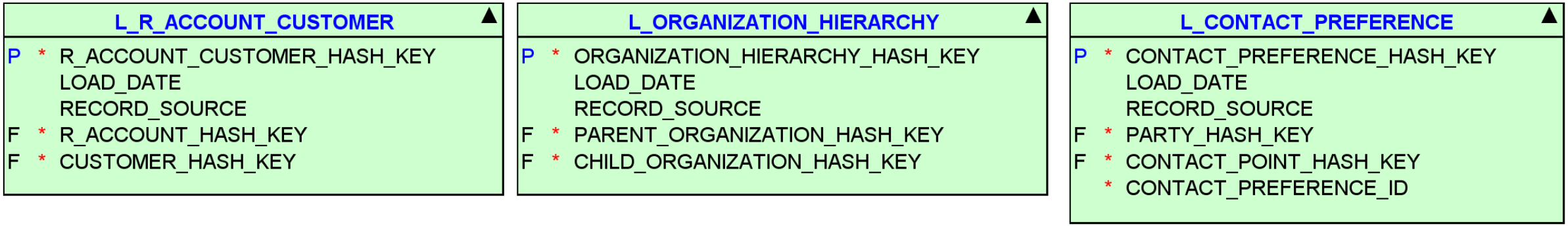
- **HashKey** - PK
- **LoadDate**
- **RecordSource**
- **Entity1HashKey** – FK

…

- **EntityNHashKey** - FK
- (**DependentChildKey**) - Degenerate Field

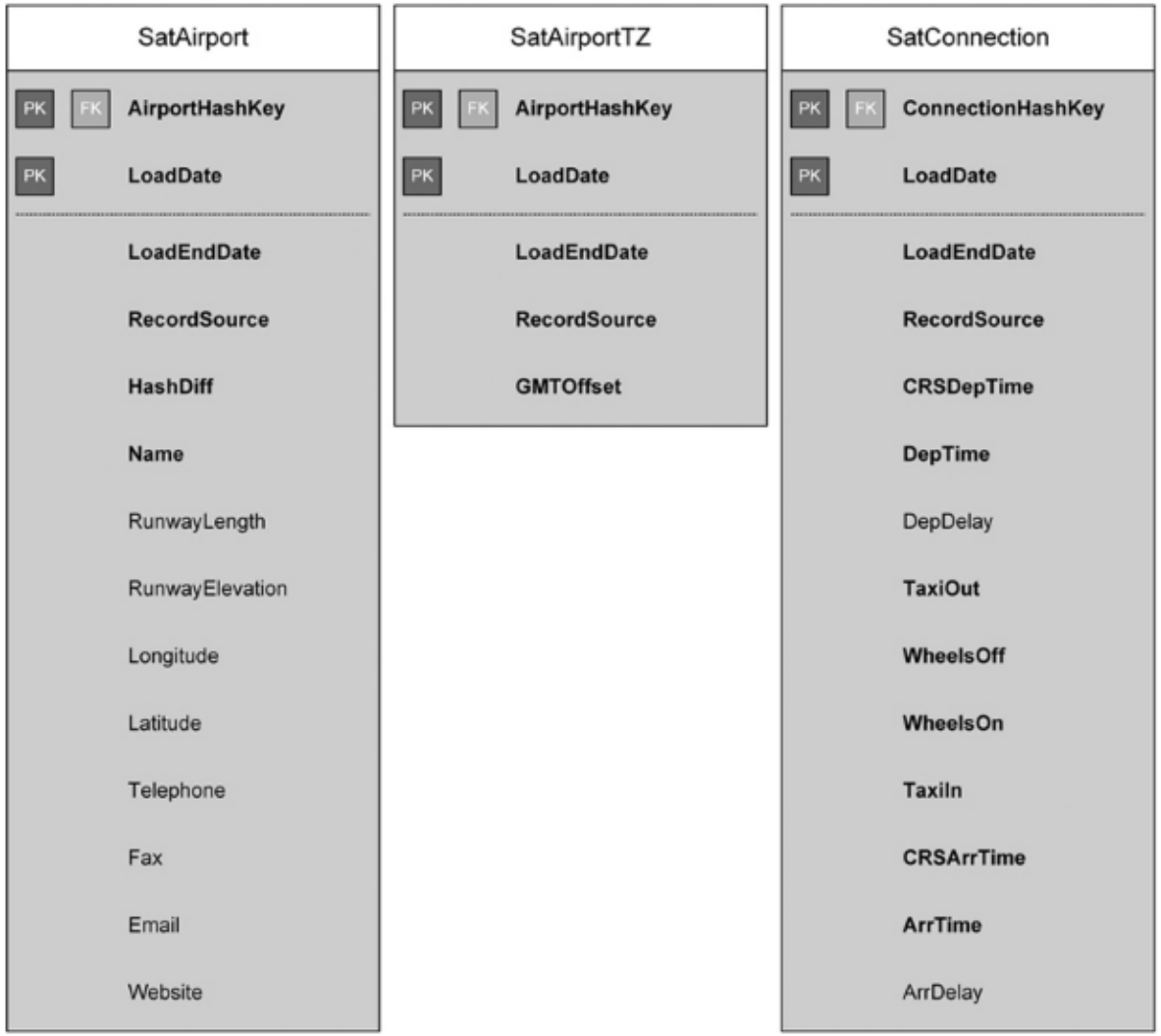


Link (examples)

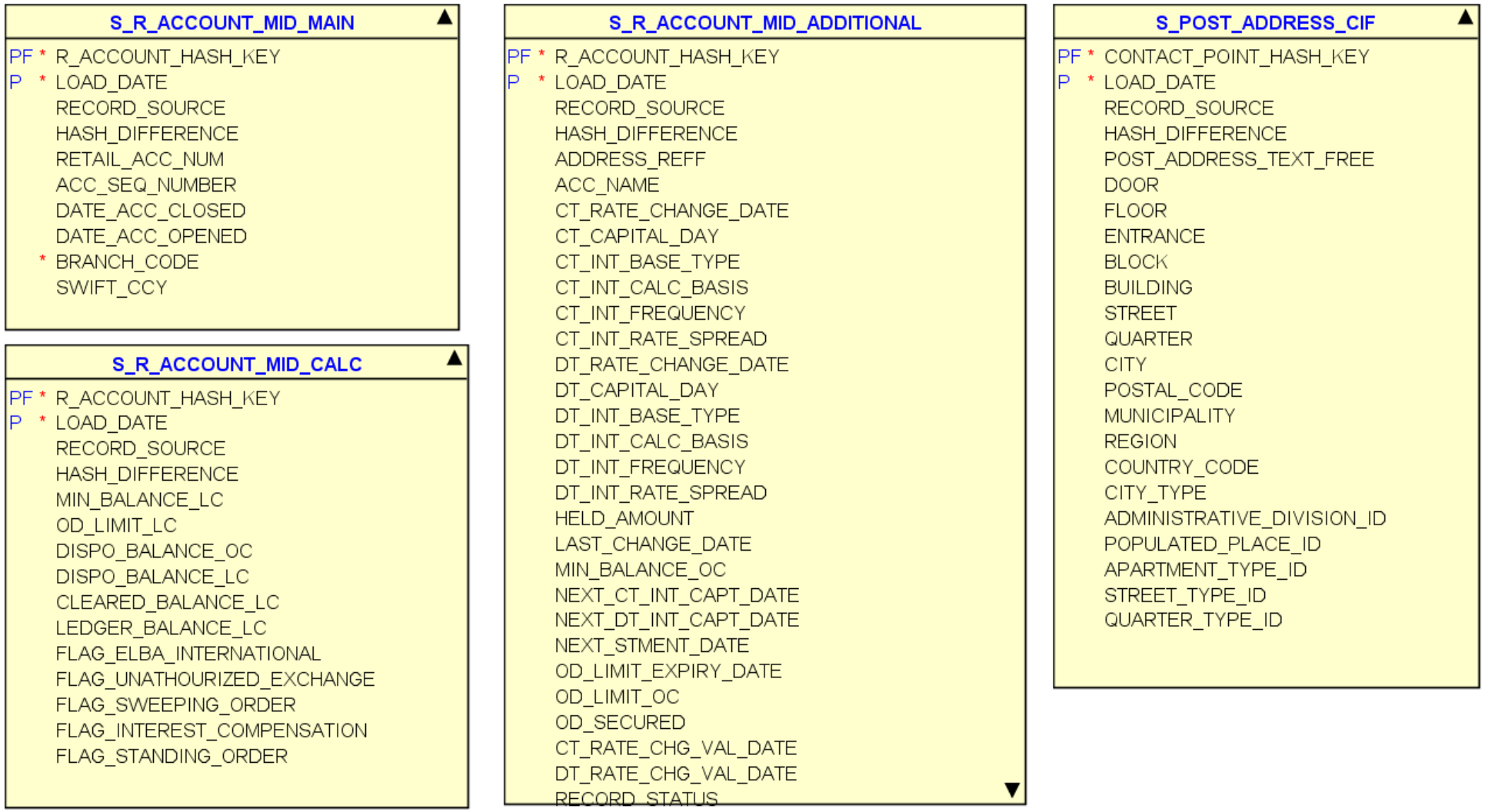


Satellite

- **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**

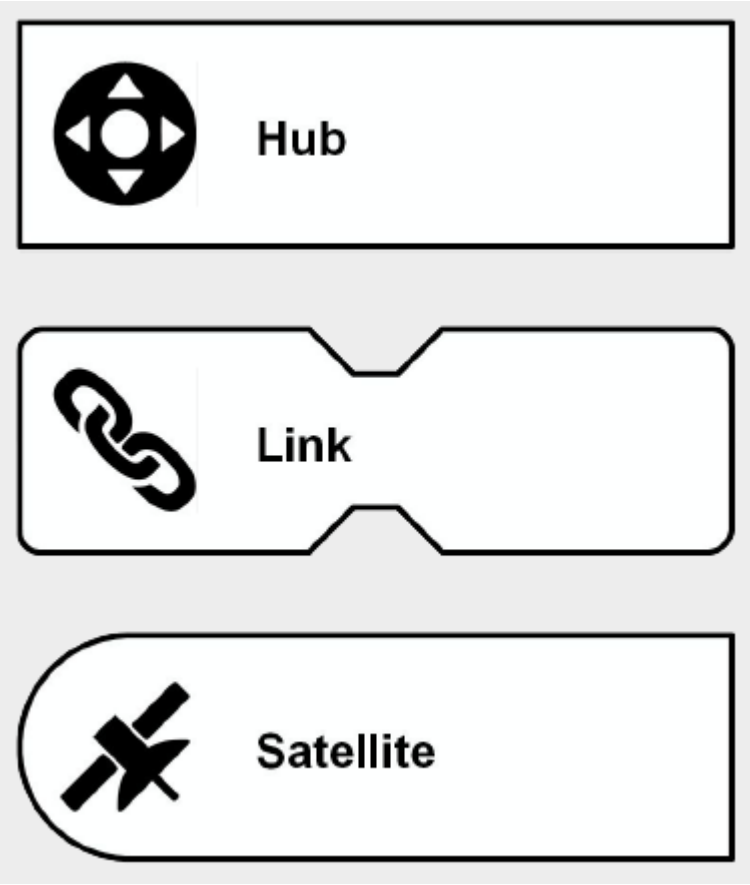


Satellite (examples)



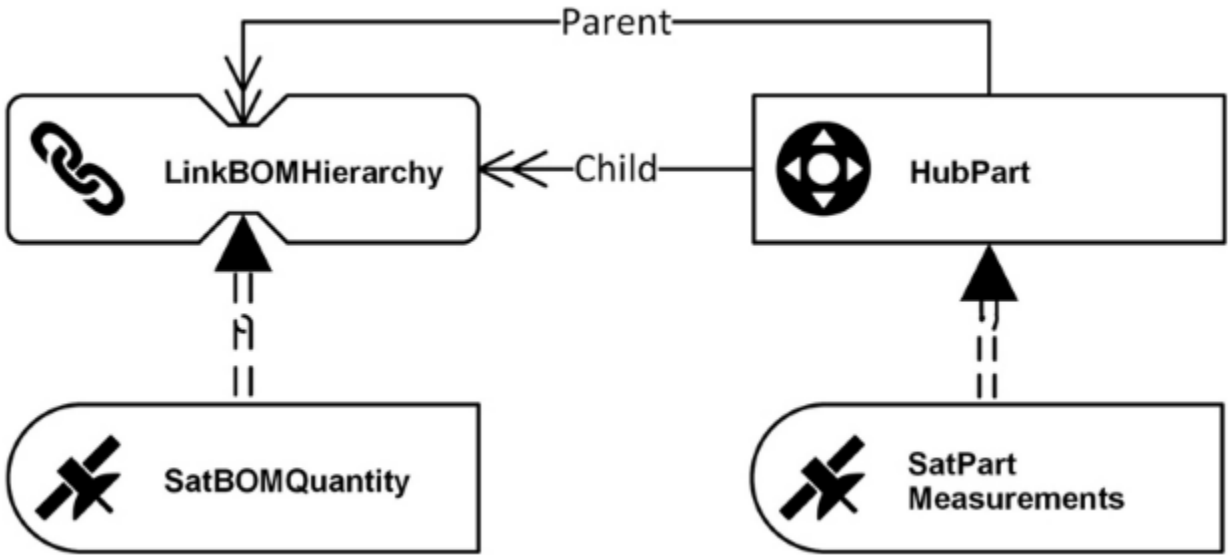
Visual Data Vault – Logical Modeling Language

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



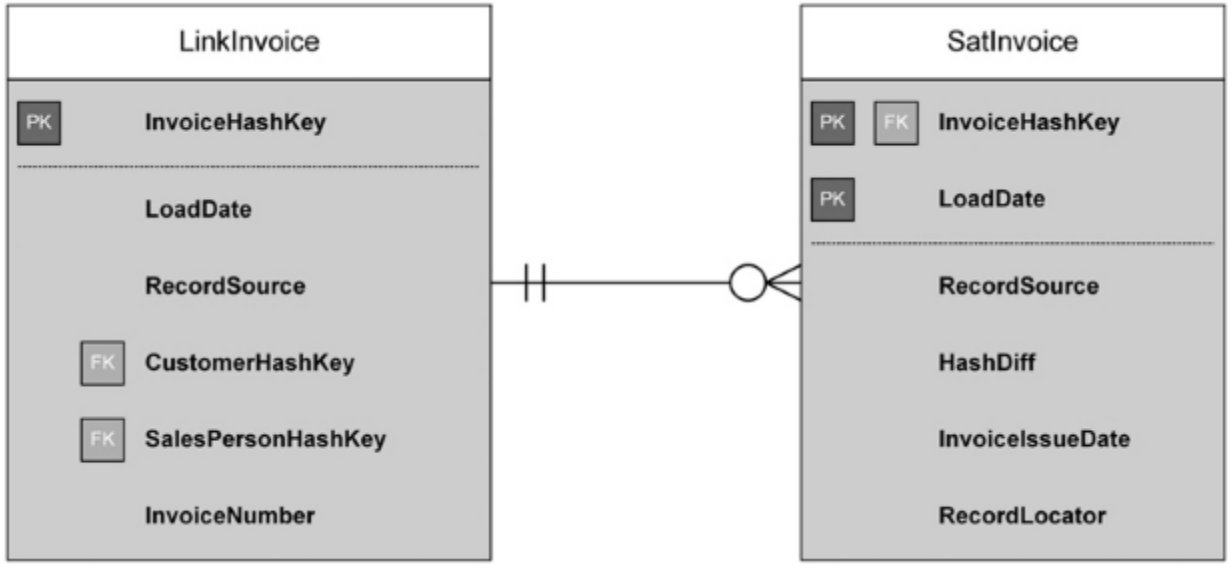
Hierarchical Links

- Parent-child hierarchies
- Level-based hierarchies



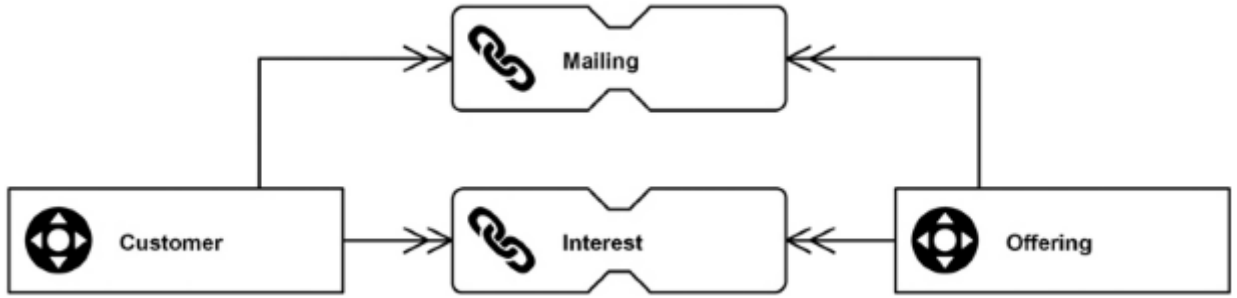
Non-historized (Transactional) Links

- 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)