

SAS and Data Vault

Why juggle stars when you can model galaxies

Bronwen Fairbairn

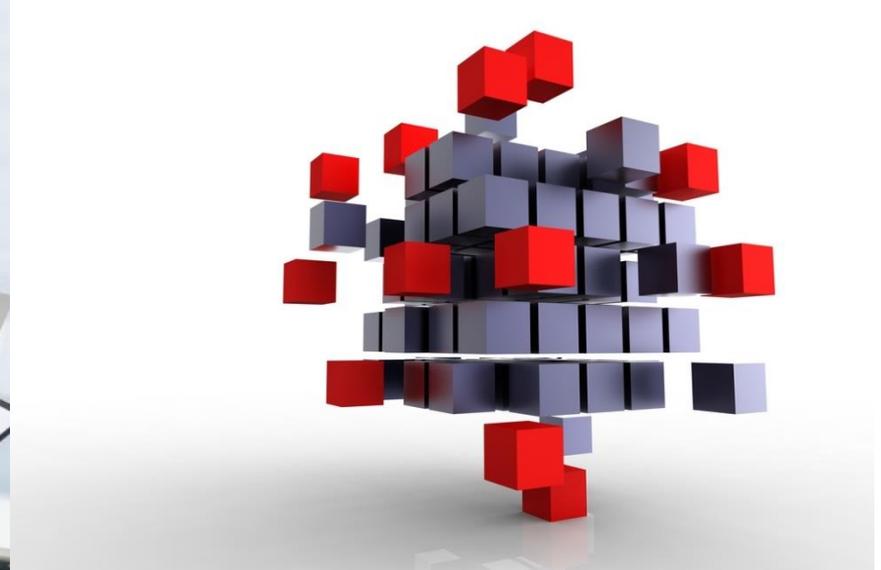
Business Intelligence Consultant

bronwen@bronwen.org

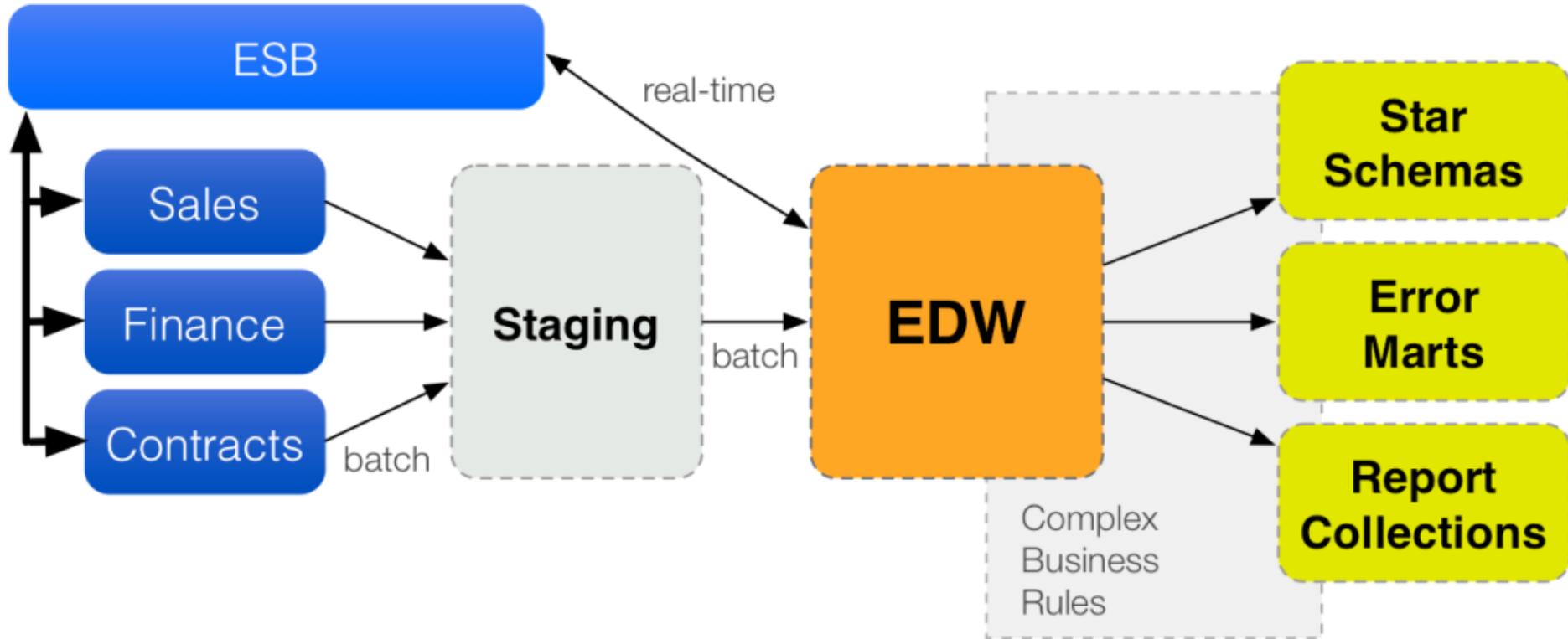
Data Warehouse that can...

A magnifying glass is positioned over a data table, highlighting a specific row. The table contains numerical data with four columns.

1	0000	1	Administrative and Support	24378	0.00	1.00
2	0000	2	Customer Service Support	26852	0.00	0.00
3	0000	3	Human Resources Support	23386	46.00	92.00
4	0000	4	Information Systems Support	510515	0.00	0.00
5	0000	5	Legal Support	506781	0.00	0.00
6	0000	6	Marketing Support	92001	0.00	0.00
7	0000	7	Operations Support	95001	0.00	0.00
8	0000	8	Public Affairs Support	94011	0.00	0.00
9	0000	9	Research and Development Support	514278	0.00	0.00
10	0000	10	Security Support	518003	99.00	99.00
11	0000	11	Training Support	534941	0.00	0.00
12	0000	12	Unassigned Support	555555	0.00	0.00



Data Warehouse that can...



Data Vault Concepts

HUB HUB - Should represent business keys, relatively static and widely known. Focus on the keys your business cares about.

SAT — **HUB** SAT - The attributes for a business key. Capture as much of the data from source as possible. Generally each source writes to different SAT tables.

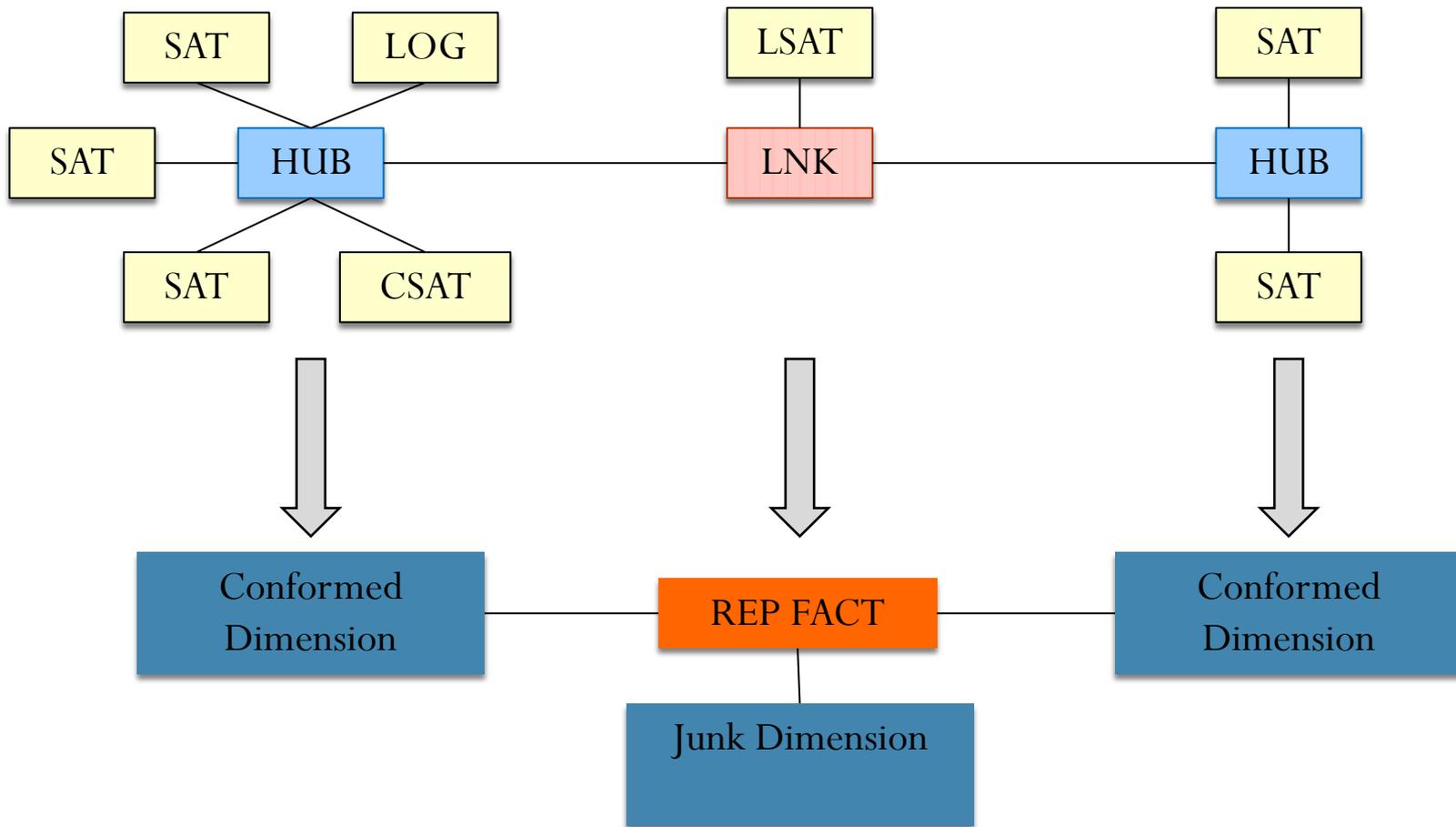
HUB — **LNK** — **HUB** LNK - A relationships between the HUBs.

LSAT — **LNK** LSAT - Exactly like a SAT, but describes the relationship, including it ending.

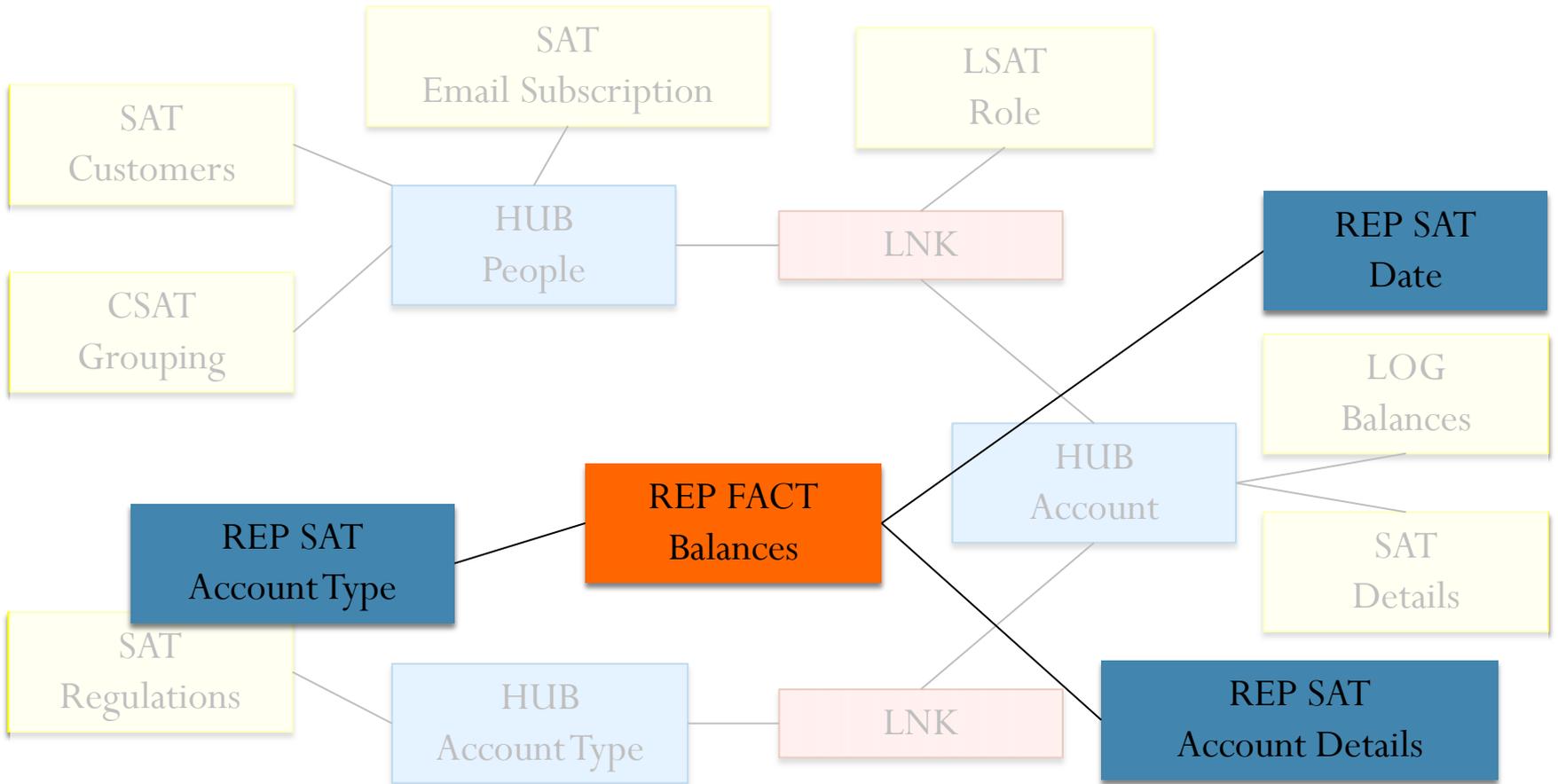
LOG — **HUB**
LLOG — **LNK** LOG - Exactly like a SAT, bit has date as part of the key. For example Web site activity logs.

CSAT — **HUB** Clean SATs - The summarised, cleansed result of information elsewhere in the Data Vault. These should be entirely re-creatable.

Data Vault Concepts



Data Vault Example



Benefits

- Agile and Extensible
 - Data Vault is capable of holding an entire organisation's information
 - Multiple Data Sources can be easily integrated
 - HUBs can be added to existing LNKs and visa versa
 - The Data Vault can sit across multiple databases
- Business Focused
 - Leverage the entire business's data
 - Analytical Models should integrate perfectly into the Data Vault
- Flexible and Reliable
 - Only core business rules have been built in to the Data Vault
- Parallelism
 - HUBs > LNKs > SAT/LSAT/LOG/LLOG > CSAT > Reports

Implementation

- Consider developing your own transformations
- Loading the Data Vault with SAS can be lightning fast
 - SAS indexing and lookup tables
 - Adding new HUB keys from LNK processing
 - Table structures automatically adapt to changes in source
 - SCD tables that don't write records for all null values
 - SCD with source and effective date stamps
 - Repeat tables for when the data has been checked
 - Process that switches between loading styles
- Table and Job Logging

Implementation

- Choosing your hub keys is the most of the design
 - Link tables should then be all relationships
 - Different SAT tables for each source system, completely lossless.
- Migrating an existing Data Warehouse
 - Design the DV to work with your current systems and requirements.
 - Then import all historical data from the old data warehouse with negative Sequence numbers.

Summary

- Data Vault should make your EDW auditable and more flexible
- There is Data Vault training available in Brisbane
- Data Vault is evolving
 - It can adapt to your situation
- Building a toolkit for loading the Data Vault may save you time.
- Data Vault is the source data but business focused.
 - So it is a great communication tool.

Any Questions?

Bronwen Fairbairn

bronwen@bronwen.org

www.linkedin.com/in/BronwenFairbairn