Breaking Silos

A Data Engineering Initiative By GovTech

Data Engineering

The *Art* of **Designing** and **Building** systems to

Collect, **Store**, **Analyze** and **Provide access** to data

Whenever, **Wherever** and **However** it is needed.

"Data Engineers are like Plumbers.

Plumbers bring water from source to destination, they move water reliably from place to place.

Likewise, Data Engineering is about __moving data__ and making sure pipelines are reliable such that data can get to where it needs to be. "

An analogy by Edwin Law
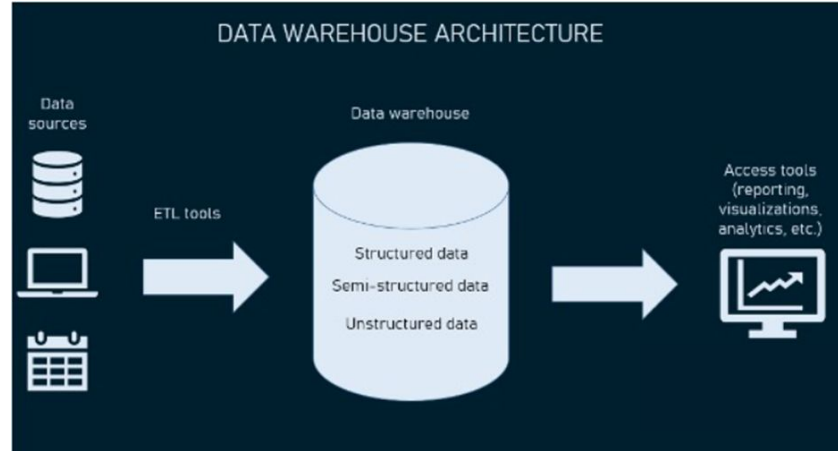**Grab – Data Engineering Lead**

Silos?
Data Engineering Can help!

Data Warehousing solutions are fantastic for knocking down Silos within an organisation!

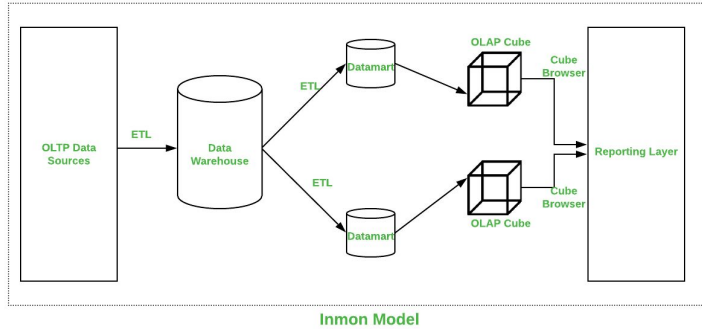Recommendation: **Inter-RI Data Warehouse** to improve information sharing between RIs
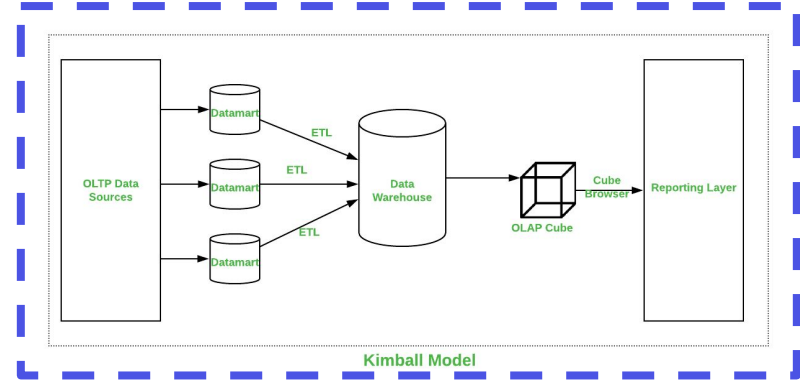
Data Warehouses (DW)

DATA WAREHOUSE ARCHITECTURE

Data sources · ETL tools · Data warehouse (Structured data, Semi-structured data, Unstructured data) · Access tools (reporting, visualizations, analytics, etc.)

Definitions: <u>Central repositories</u> of data combined from multiple sources that provide **<u>meaningful</u>** business insights.

Value-add: Enable cross-organisational **information sharing** while ensuring data privacy and security by defining access through **data marts**.

Inmon & Kimball DW Models

Inmon Model

Kimball Model

1) Ingests from source data
2) Normalised Model
3) Pros: Less Data Redundancy, Completeness
4) Cons: Initial cost, More ETL required

1) Ingest from existing Data Marts
2) Denormalized Model
3) Pros: Faster to set up, Understandability
4) Cons: Data Redundancy, can be Incomplete

Recommendation: Kimball Approach - Easier for RIs to share data from existing data marts than to come up with new structures for 20+ RIs

Value Adds

RIs

Collaboration

Data is standardised and accessible - fostering collaboration between RIs

Less Redundancy

Easier to find and utilise similar past research data from other RIs

HQ

Visibility

Consolidated Data allows HQ to form clearer view of the research scene

Accountability

RIs encouraged to be more accountable and transparent with research resources due to increased HQ visibility

Thanks!