

Splintered Business Intelligence Architectures



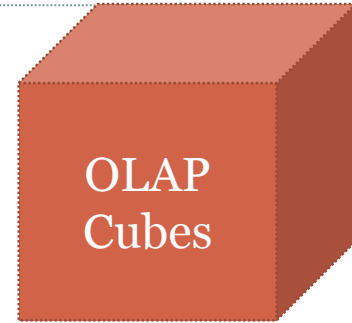
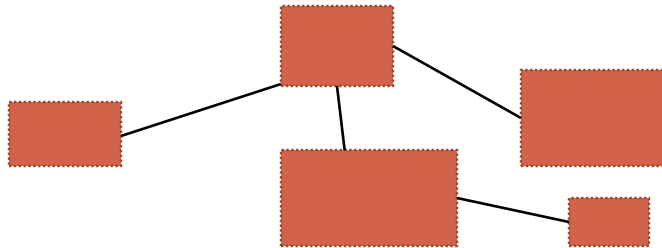
By
WILLIAM STERLING

Analytic Architectures and Data Warehousing Today

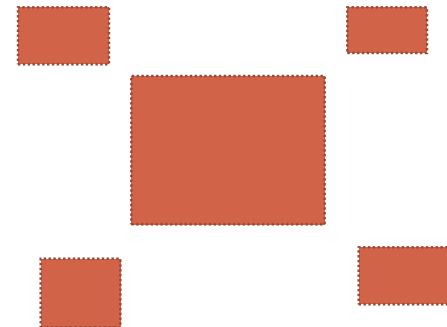


**AN EVOLUTIONARY BUSINESS
AND
SPLINTERED
ARCHITECTURES**

A Very Short History of Warehouses and Data Marts



1991 Arbor Software creates the modern version of OLAP



Ralph Kimball creates Redbrick and the **Star Schema**, a relational way to model dimensionally..

EF Cod and C. Date create the relational Calculus: guarantees data integrity, separates access from where data is and how it is structured.

*3rd normal form:

Bill Inmon "Data Factory"

- *minimize or forbid data redundancy;
- Programs no longer had to have copies of shared data structures
- Data safety can be proven and trusted
- DBA's core value might be 'no surprises'
- What about reporting?

A full SAP installation can have **80,000** tables.

A quick definition of relational 3rd normal form



“3NF is the form best used by DBA’s to build a database optimized for update. The normalization gives you one copy of each data element, so that updates don’t take so long and the integrity is guaranteed.”

“It’s all about an update.”

- Rex Wiederanders, IBM

In other words, it’s all about OLTP getting data in, not about OLAP, getting data out.

Technological and Cultural Divides in the Analytic World



Finance – IT cultural and technical divide:

(But not just Finance models Manufacturing Models, inventory models, transportation models, etc. All department level analytics).

Within Warehousing: Bill Inmon – Ralph Kimball Divides:

There is nothing wrong with this; we all promote what we know.

Normal Form designs vs the Star Schema: DBA Angst

These differing perspectives have a real impact on optimizing business performance and on selling analytic solutions into the enterprise environment.

An IT Architecture Dilemma: Star Schemas versus 3NF Data Factories



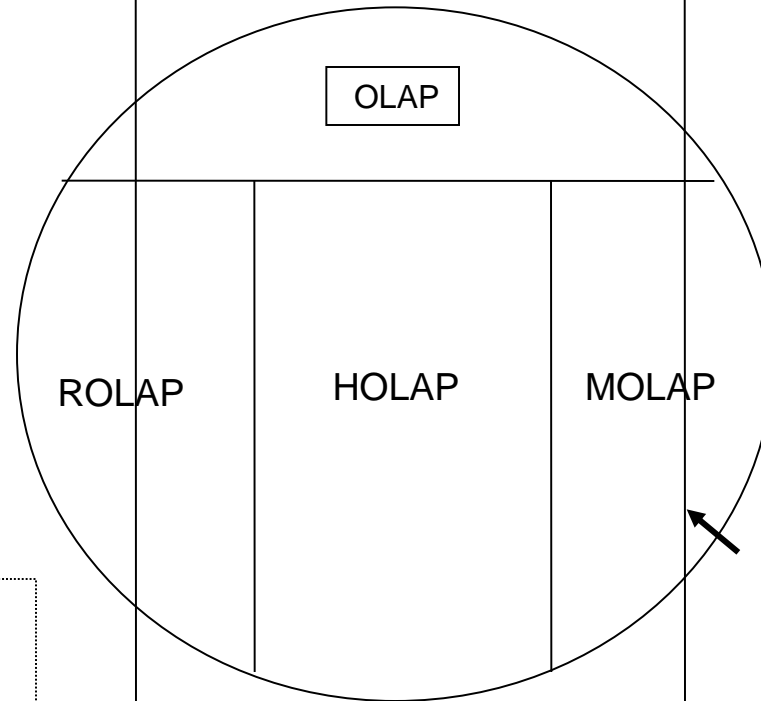
- This is a false dichotomy
- However, it can be cultural, even religious in nature
- Following is an explanation of how this divide comes into being and maintains itself
- We all champion what we know...

The OLAP – IT Technical *and Cultural* Gap

Systems Analysis
Perspective

Business Analysis
Perspective

IS
MIS
Systems Analysis
Data Warehousing
Enterprise
Do ERD diagrams
run the world?



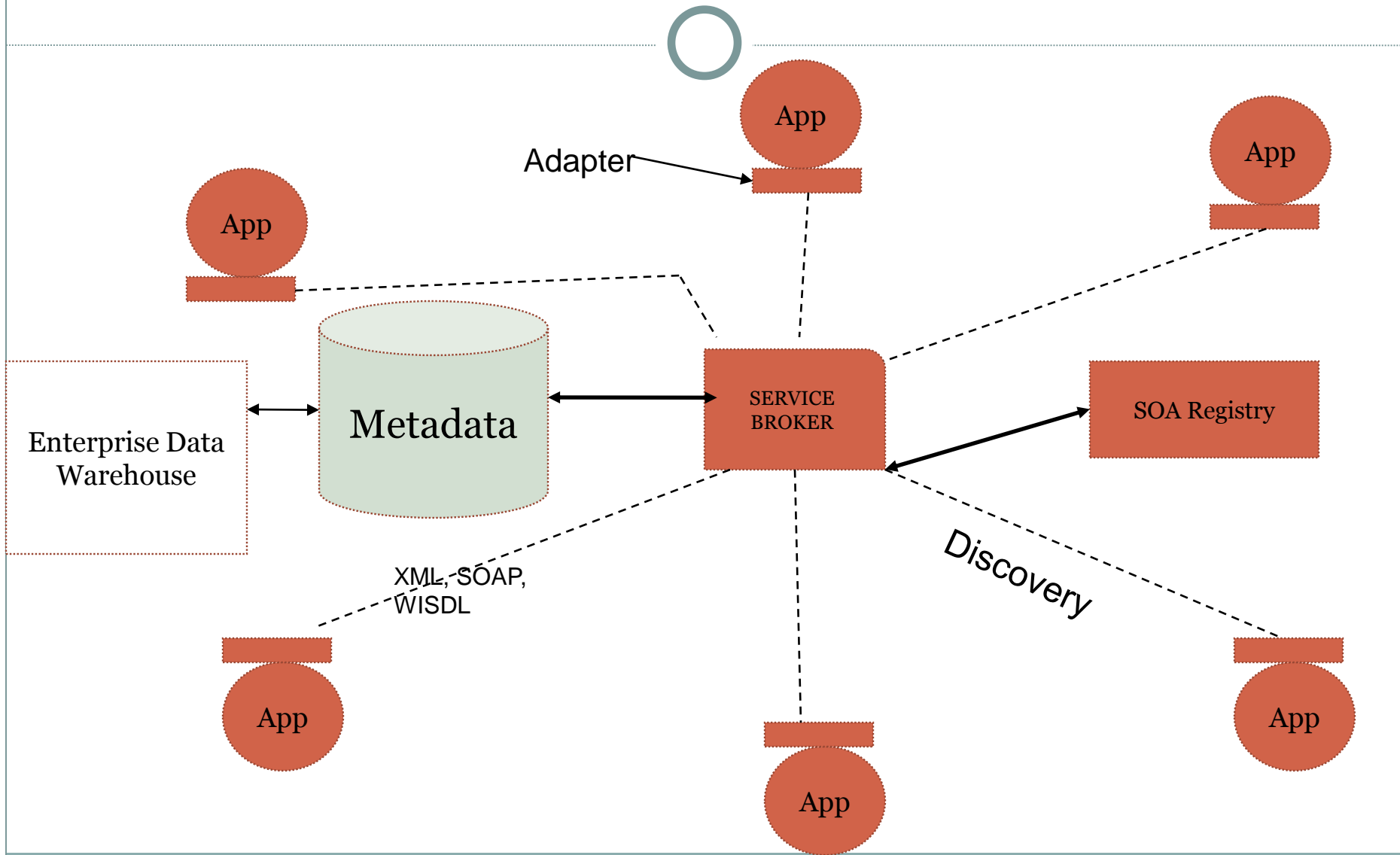
MBA
Business Analysis
KPI
Departmental
Do 50 Algorithms run the
world?

Source: IBM Data Warehousing 2001 Chapter 12: By Bill Sterling and Richard Sawa

These 'wars' are over, but the perspective difference persists ..

The way one views "OLAP" depends upon what side of the analytic space you come from, systems analysis or business analysis. The whole original "rolap, holap molap" debate comes from this difference.

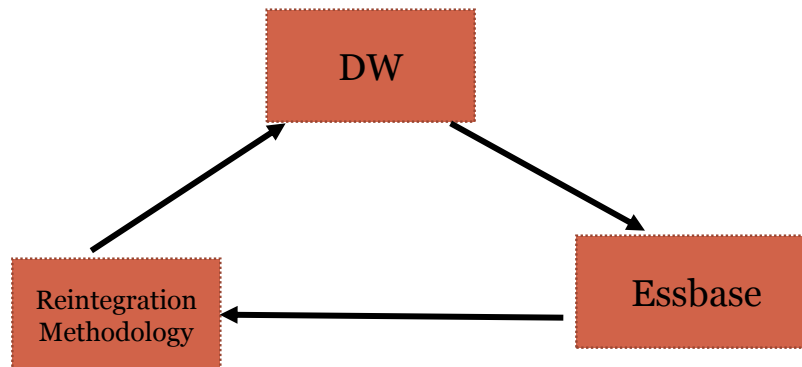
And Now Service Oriented Architectures require conformed metadata



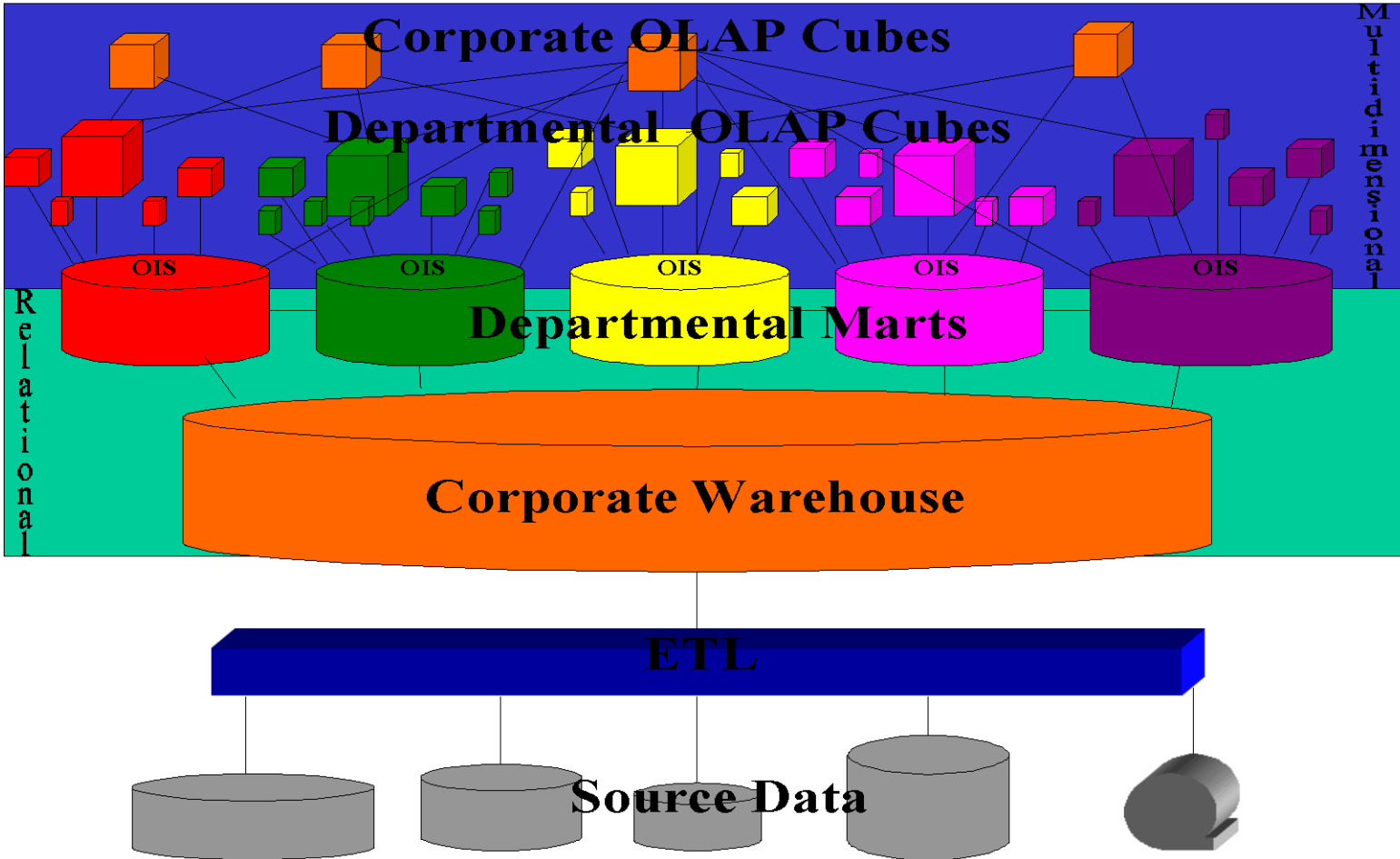
The criticality of conforming Dimensions



- There are ~123 ways to spell AT&T
- **Ralph Kimball** : “Dimensions are conformed when they are either exactly the same (including the keys) or one is a perfect subset of the other. Most important, the row headers produced in answer sets from two different conformed dimensions must be able to be matched perfectly.” – p. 394 Data Warehouse Toolkit, 2nd Edition
- Essbase won't work any other way – how many kickouts have you had in an Essbase data load?
- A method for returning cube data to the warehouse takes the conformed dimensions in an Essbase cube and ports to the DW



An OLAP Bad Dream ... It is unwise to develop OLAP Data Marts at the Enterprise level. That is the proper job of the Enterprise data warehouse.



Source: Chapter 12 IBM Data Warehousing; Sterling and Sawa

Next a Brief Review of Star Schemas



1. Value to enterprises
2. Compared to cubes
3. Role in mediating between cubes and warehouses
4. Some structural details

The Fact Table and Dimension Tables



Surrounded by
Dimension tables

Fact Table

“The facts”

Product (PK)	Name
1	X Server
2	P Server
3	Z Server

Store (PK)	Name
1	Tokyo
2	Sydney
3	Beijing

Time (PK)	Month
1	Jan
2	Feb
3	Mar
4	Apr

Product (FK)	Time (FK)	Shipper (FK)	Store (FK)	Sales
1	1	1	1	100
1	2	2	2	200
1	3	1	2	150
1	4	2	3	250
2	1	2	1	50
2	2	1	2	150
2	3	2	2	300
2	4	3	3	200
3	1	1	1	100
3	2	2	2	200
3	3	2	2	300

Shipper (PK)	Name
1	UPS
2	Fed Ex
3	Pickfords

Primary and Foreign Keys



Primary Keys

Foreign Keys

Primary Keys

Product (PK)	Name
1	X Server
2	P Server
3	Z Server

Product (FK)	Time (FK)	Shipper (FK)	Store (FK)	Sales
1	1	1	1	100
1	2	2	2	200
1	3	1	2	150
1	4	2	3	250
2	1	2	1	50
2	2	1	2	150
2	3	2	2	300
2	4	3	3	200
3	1	1	1	100
3	2	2	2	200
3	3	2	2	300

Store (PK)	Name
1	Tokyo
2	Sydney
3	Beijing

Time (PK)	Month
1	Jan
2	Feb
3	Mar
4	Apr

Shipper (PK)	Name
1	UPS
2	Fed Ex
3	Pickfords

Querying A Dimensional Schema



- This is a simple dimensional schema
- Relational joins allow queries across fact and dimension tables
- What were product ID 1 sales \$\$ in March in the Sydney store?

Product (PK)	Name
1	X Server
2	P Server
3	Z Server

Product (FK)	Time (FK)	Shipper (FK)	Store (FK)	Sales
1	1	1	1	100
1	3	2	2	200
1	3	1	2	150
1	4	2	3	250
2	1	2	1	50
2	2	1	2	150
2	3	2	2	300
2	4	3	3	200
3	1	1	1	100
3	2	2	2	200
3	3	2	2	300

Store (PK)	Name
1	Tokyo
2	Sydney
3	Beijing

Time (PK)	Month
1	Jan
2	Feb
3	Mar
4	Apr

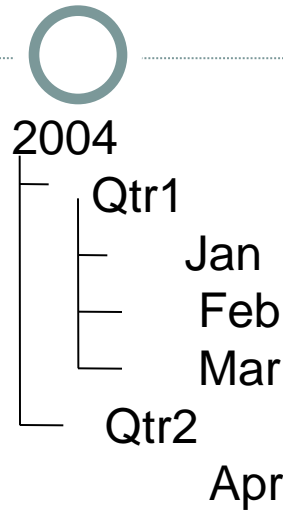
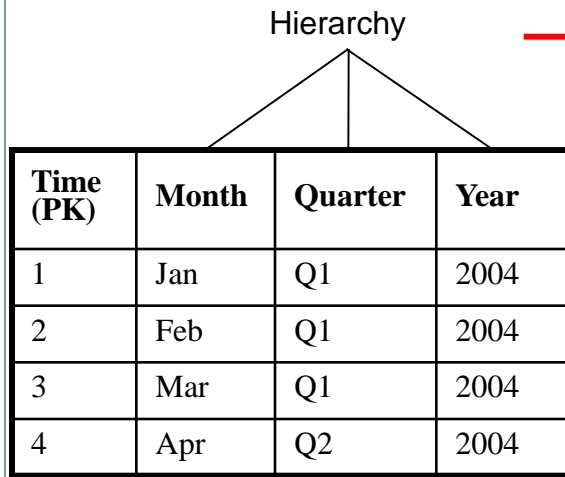
Answer:
\$350

Shipper (PK)	Name
1	UPS
2	Fed Ex
3	Pickfords

Note: a star can be exported with primary and foreign keys, (often called “synthetic keys”) or with Essbase member names in the fact table columns.

Hierarchies

A critical difference in representation and in thinking between cube OLAP Outline dimensional structure and relational dimensional structures. Key part of the cultural divide.



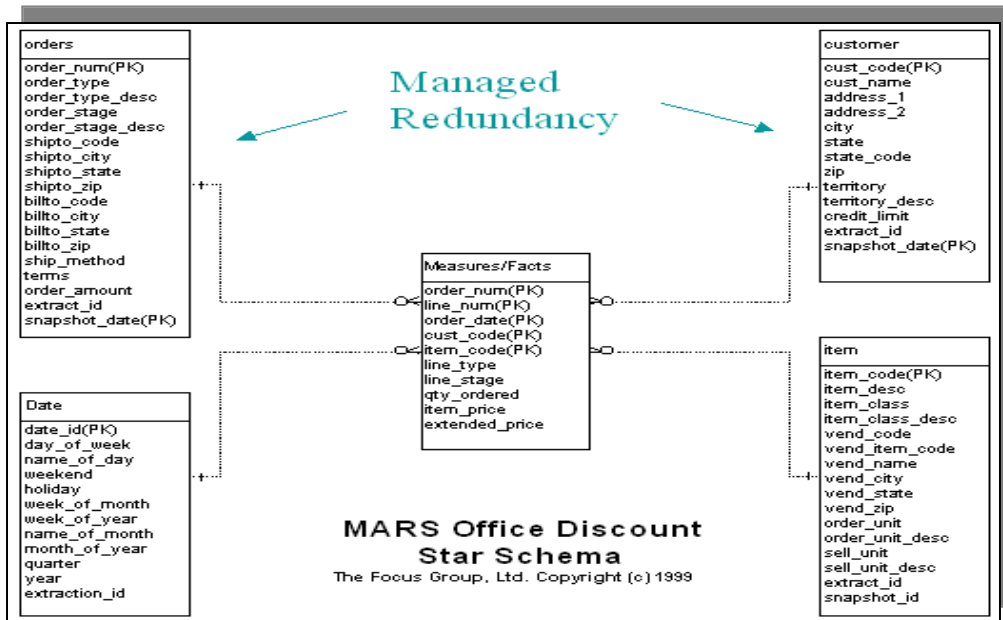
The Essbase outline is a picture where the business model and the data model are the same, expressed in one document. The star schema isn't.

Product (PK)	Name
1	X Server
2	P Server
3	Z Server

Product (FK)	Time (FK)	Shipper (FK)	Store (FK)	Sales
1	1	1	1	100
1	2	2	2	200
1	3	1	2	150
1	4	2	3	250
2	1	2	1	50
2	2	1	2	150
2	3	2	2	300
2	4	3	3	200
3	1	1	1	100
3	2	2	2	200
3	3	2	2	300

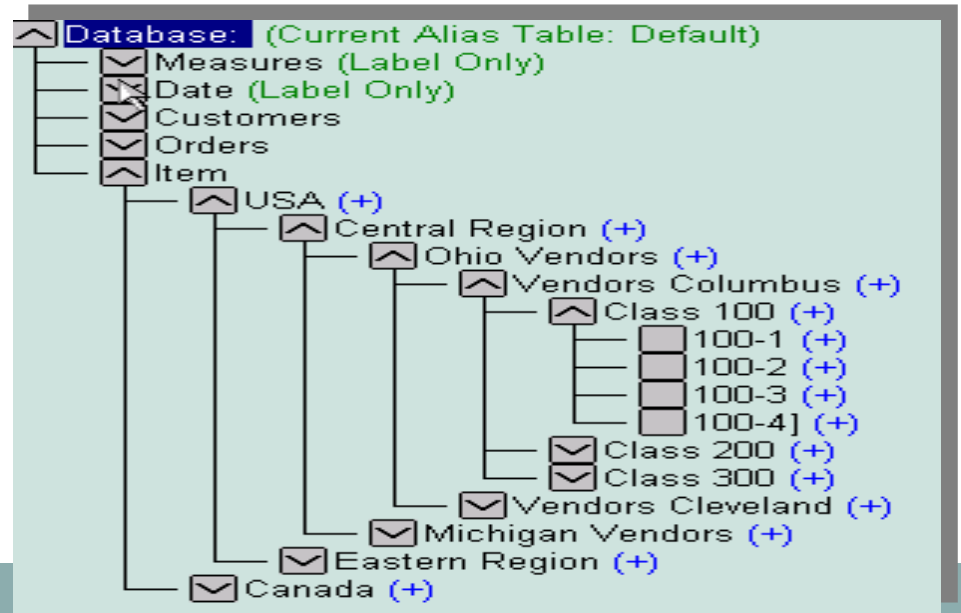
Store (PK)	Name
1	Tokyo
2	Sydney
3	Beijing

Shipper (PK)	Name
1	UPS
2	Fed Ex
3	Pickfords



The same model

The Essbase outline is a way of picturing an enterprise model that is a closer approximation to how business people experience the business than relational schemas are. Business users experience specific business names, like “Diet Coke”. They don’t experience the “categories” of entities like a table column called item_code, which are the meat of relational dimensional modeling.



Source: Op Cit.

Due to IT dogma about not duplicating data, the following explanation is often needed for DBA's

Why putting a cube into the warehouse as a star is not duplicating data ... it is adding vital new management information

```
IF(@ISUDA(Scenario,"Last1s53Weeks"))  
"Current Year"%@SHIFT("Last  
Year",1,Time,@LEVMBRS(Time,0));  
ELSEIF(@ISUDA(Scenario,"53Weeks") AND  
@ISMBR("Wk 53"))  
"Current Year"% "Current Year"->"Wk 01";  
ELSE  
"Current Year"% "Last Year";  
ENDIF
```

Never a
pushback once
explained

An Essbase
Calculation. Try
Doing it relationally!

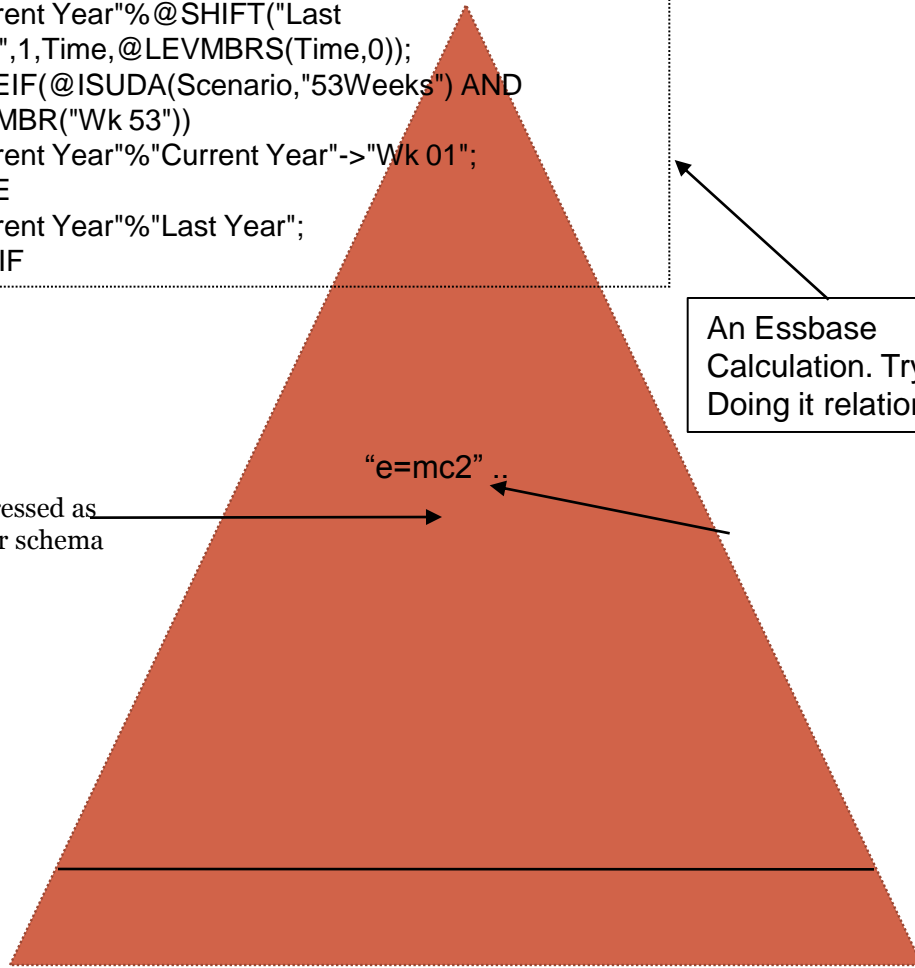
"e=mc2"

Expressed as
a star schema

DBA's likely see as
Redundant data – violating
3NF religious precept of no duplication

But it's not. The hierarchical aggregations are "new data"
that the DBA doesn't want to create anyway..

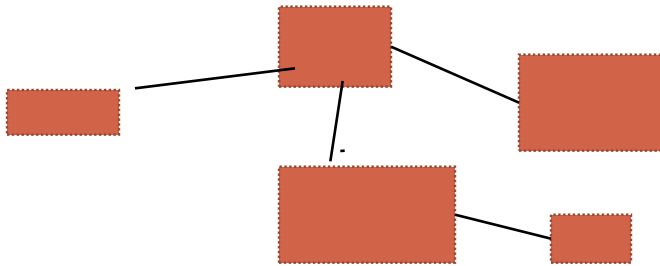
A thin slice of duplicated leaf level data



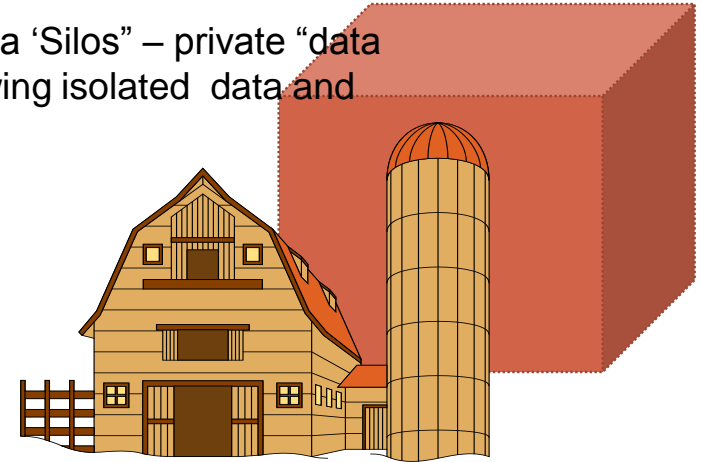
Splintered Architectures



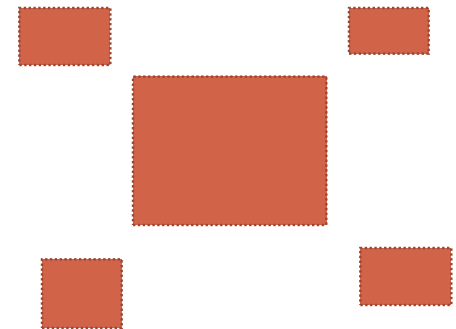
Data Warehouse



Classic data ‘Silos’ – private “data farms, growing isolated data and metadata!



Star Schemas



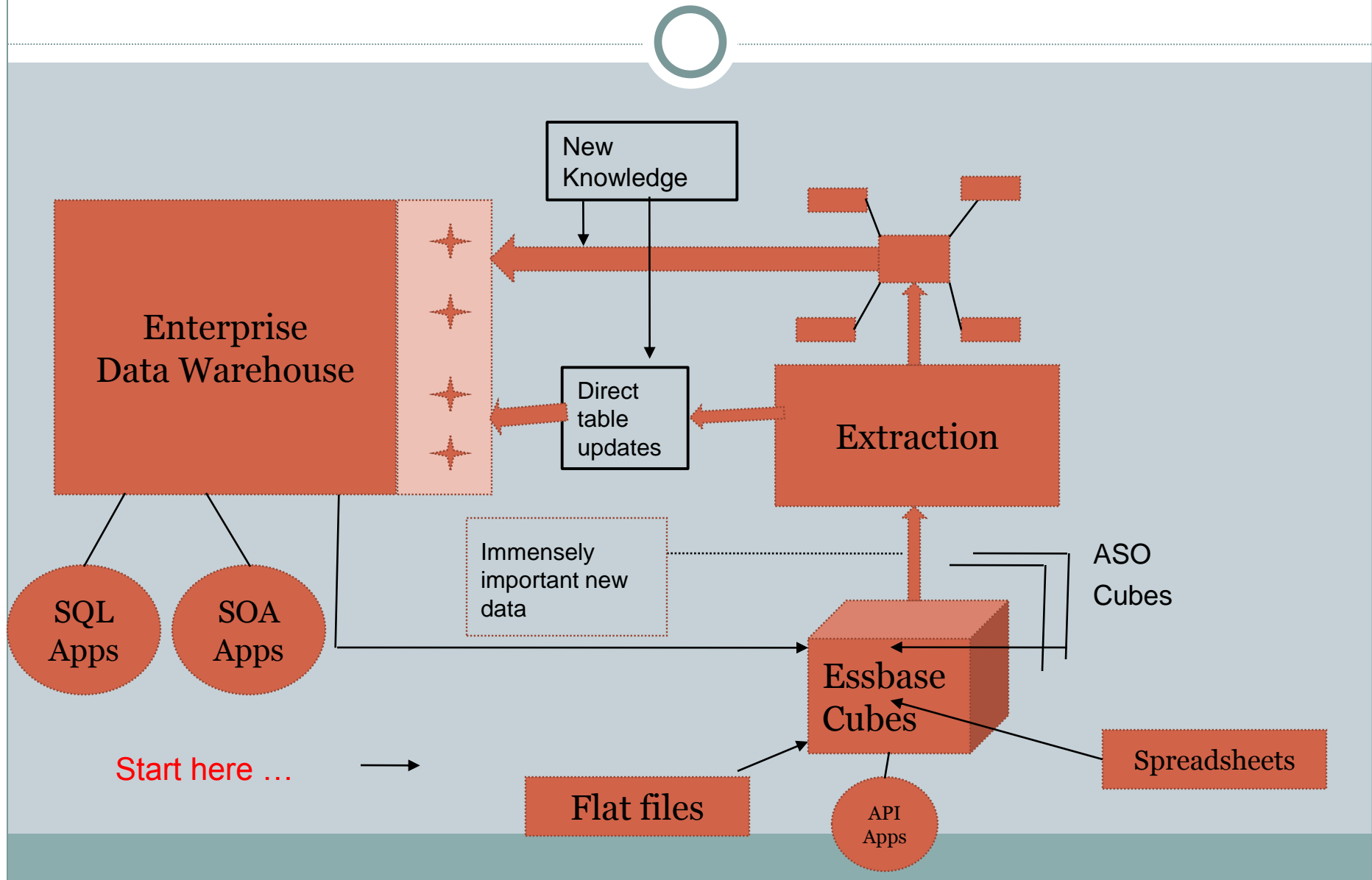
Cubes, data factories, silos, stars, little integration, isolated solutions, little or no **dimensional conformation..** (recall SOA requirement for conformed metadata)

Standard Star Schemas To the Rescue



- The Star Schema is the Rosetta Stone of the enterprise – Consumable by standard enterprise software
- Speaks the lingua franca of IS: **SQL**
- Gives IS confidence to produce any “one – off” report for executives, managers
- API's not needed
- Single source of the truth, single source of reporting the truth

How stars provides a redeeming architecture, integrating analytics into the existing trusted enterprise infrastructure



“Bringing Home the Data Marts...”



- All the major database vendors want to do this
- Whether called “Dynamic Data Warehousing”
- Or “OBIBEE”
- Teradata claims integration
- SQL Server also

Benefits of Integrated Design for Unified Business Intelligence



- New Knowledge is generated across the data mart / warehouse divide. Knowledge that cannot be known by looking within either domain separately
- Ensures a single set of calculation rules
- Relationship with IT is scheduled, automated and dependable
- Cultural divide is bridged and cooperation enhanced

Summary Benefits of Universal BI Architectures to IT



- IT becomes custodian of *all* corporate data
- Acquire metadata from a source where it must be clean (Essbase) ; thus enable critical conforming of dimensions corporate wide
- Enables enterprise wide Service Oriented Architectures
- Provide a single source of the truth
- Provide a single source of reporting
- Acquire ability to validate all planning and forecasting data with corporate wide data – provide enterprise wide SOX compliance
- Get the benefits of complex KPI calculations without writing and maintaining them
- Cultural divide is bridged and cooperation enhanced

Conclusions



- As OLAP has grown up it has created the seeds of corporate requirements for its integration into standard corporate structures;
- OLAP is mature. The connection of OLAP to warehouses is in its infancy
- This has generated a cultural divide between IT and OLAP data marts housed in Lines of Business that seriously retards corporate unification:
- That era of management data fragmentation is over;*
- It is possible to share by integration or share by migration, brilliantly defined KPI's with the rest of the corporation.
- This is now a new imperative of Enterprise Performance Management.
- Unified Business Intelligence

Updating Warehouses from Cubes



- Direct table updates from cubes are a simple way to provide unified Business Intelligence by getting the metrics across the boundary
- Another way is to port the cube metadata to the warehouse and provide a reporting mechanism to “federalize” cubes by reading them directly as if they were warehouse tables, as does Oracle OBIEE
- A third way is to push the entire cube or a subset of it into the warehouse as a star schema.