



TERADATA. THE BEST DECISION POSSIBLE

WHY DOES SIMPLICITY CAUSE SO MUCH COMPLEXITY?

ROB ARMSTRONG

Rob.armstrong@teradata.com
www.smartdatacollective.com/robarmstrong



Opening Thought

Technology offers the potential to make life easier yet each new technology adds complexity and increases difficulties

– Paradox of Technology from “Design of Everyday Things” Don Norman

2

TERADATA. THE BEST DECISION POSSIBLE

We are the Creator of Much of our Complexity

- Things are built for maintenance and management, not necessarily usage
- Can't evaluate every possibility
- Tend to have an "either / or" mindset
- Have natural bias towards what we already know – or discard what we think is not "do-able"
- Politics and Organization tend to color our selections

A data warehouse is a technical solution to an analytical business problem, yet we use the same processes as for transactional problems

OLTP		DATA WAREHOUSING
Transaction Posting	Business Need	Iterative Analytics
Process Speed	Focus	Query Flexibility
Point-in-Time	Timeframe	Periods of time
Known	Database Operation	Unknown
Very Repetitive	Consistency	Becomes Dynamic

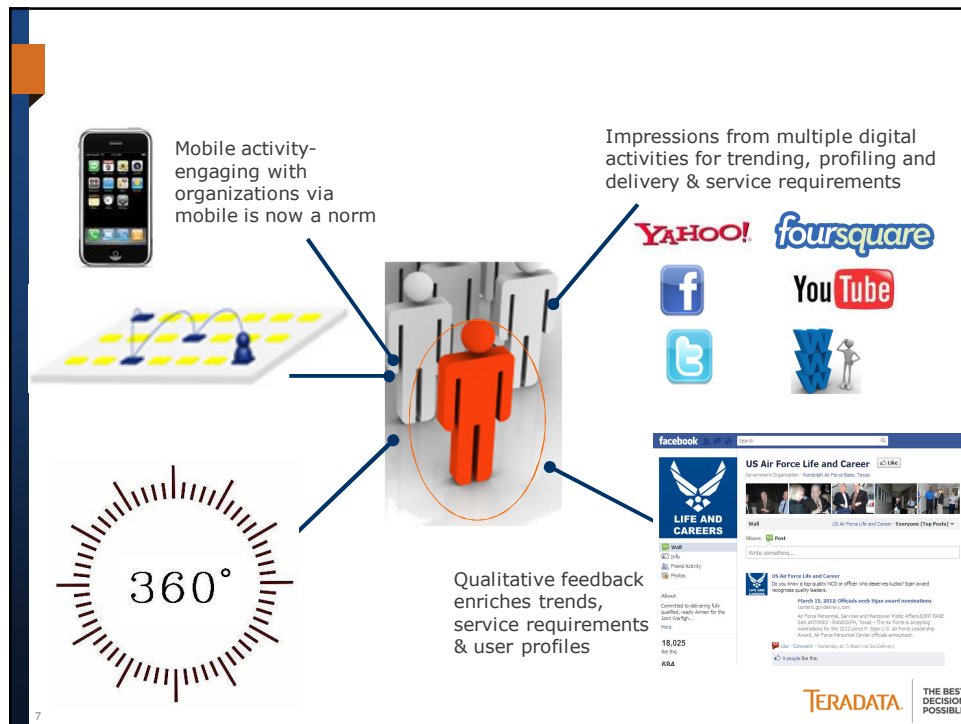
5

TERADATA THE BEST DECISION POSSIBLE

The warehouse needs to constantly overcome obstacles that are as yet undefined

6

TERADATA THE BEST DECISION POSSIBLE



But we do know what is going to happen...

- More varied analytics against the data
- More granular data
- More frequency to the load processes
- More users across the corporation
- Tighter integration leading to corporate escalation for governance

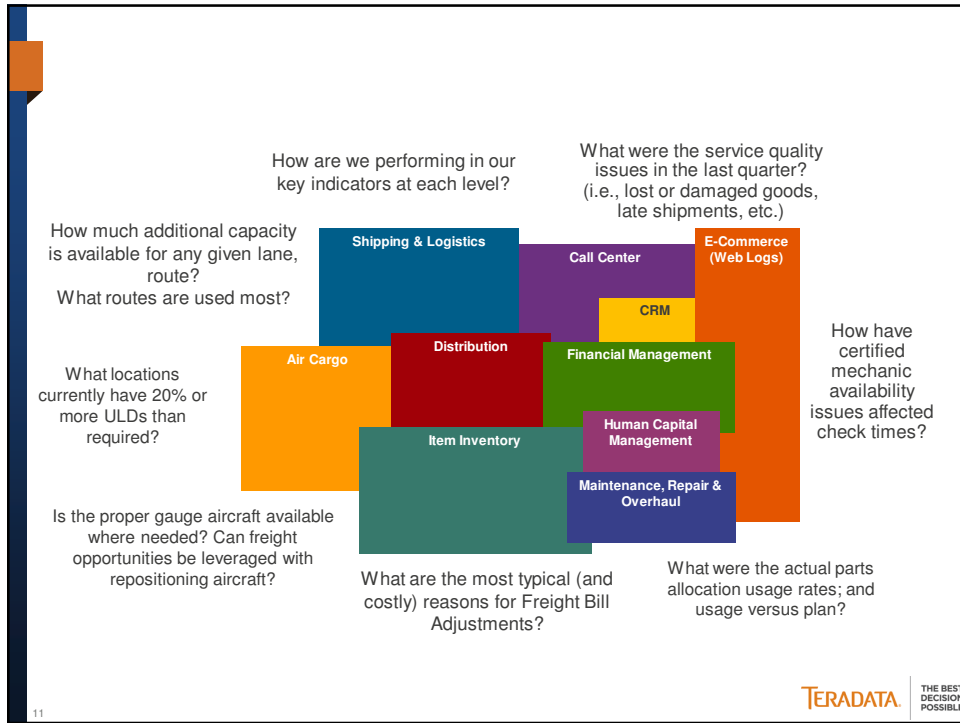
Decisions are easy when Priorities are Known

A foundation built on elements that
change, cannot change;

A foundation built on elements that don't
change, drives change

Attribute	Change over time
Data Loading	Weekly - Daily - Hourly
Data Granularity	Corporate - Region - Store
Data Joining	Rarely - Quarterly - Monthly - On Demand
Query Scope	Report - Functional - Cross Functional
User Community	Department - Business Unit - Corporate
Data Access	Daily - Hourly - On Demand

So what doesn't change?



When you distribute data you force centralized decisions, when you centralize data you enable distributed decisions

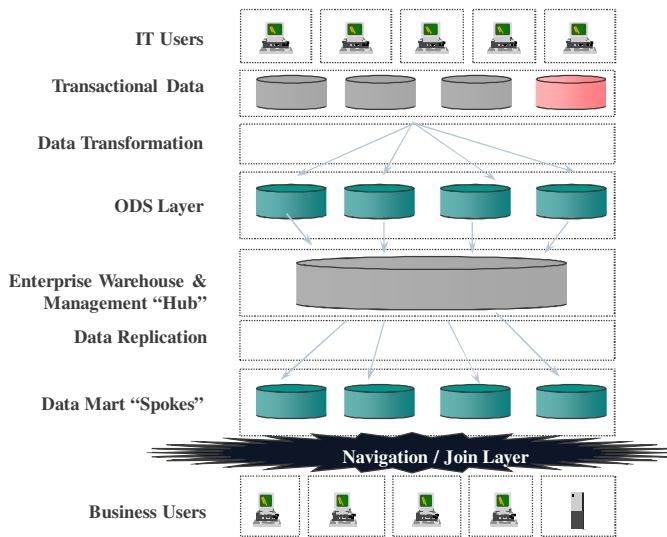
Corollary:

The more complex an architecture the more simplistic your application must be. A simple architecture allows for a complex application

12

TERADATA THE BEST DECISION POSSIBLE

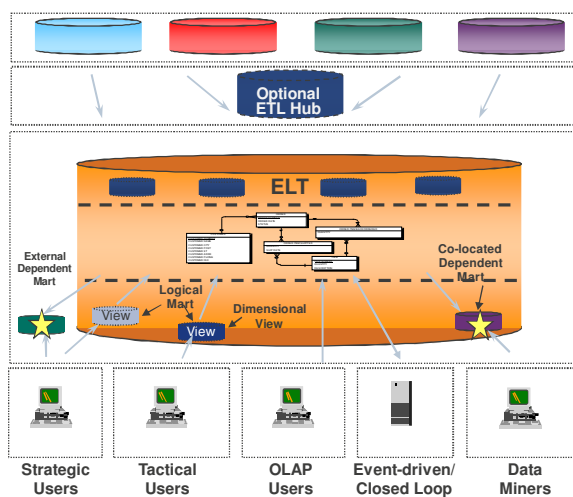
Bring Data to Process



13

TERADATA THE BEST DECISION POSSIBLE

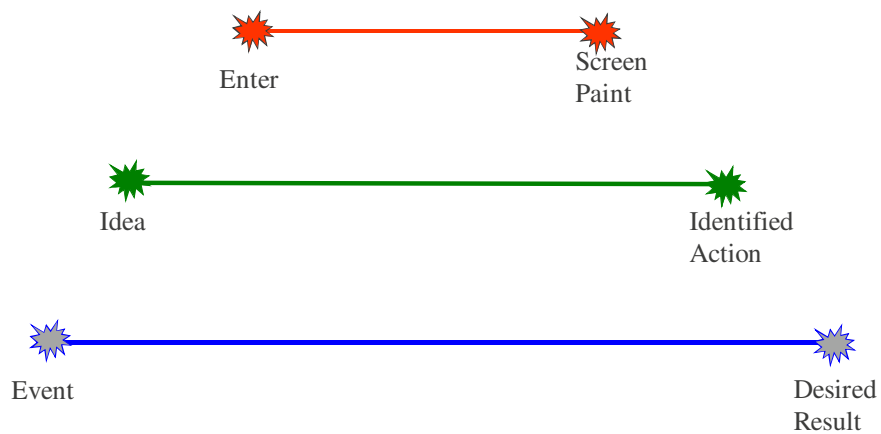
Bring Process to the Data

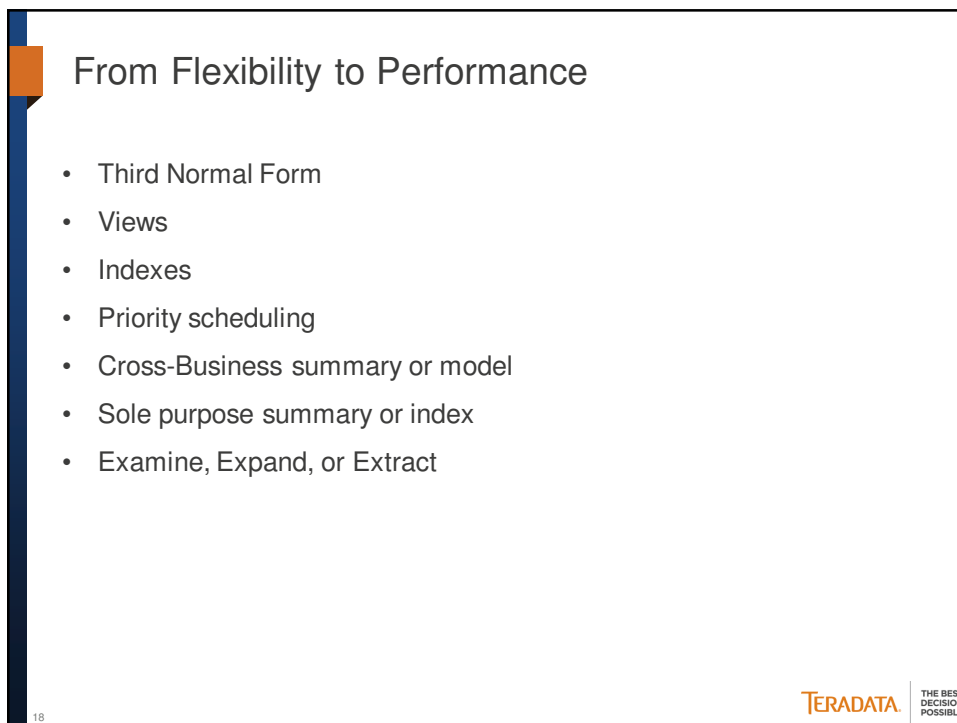
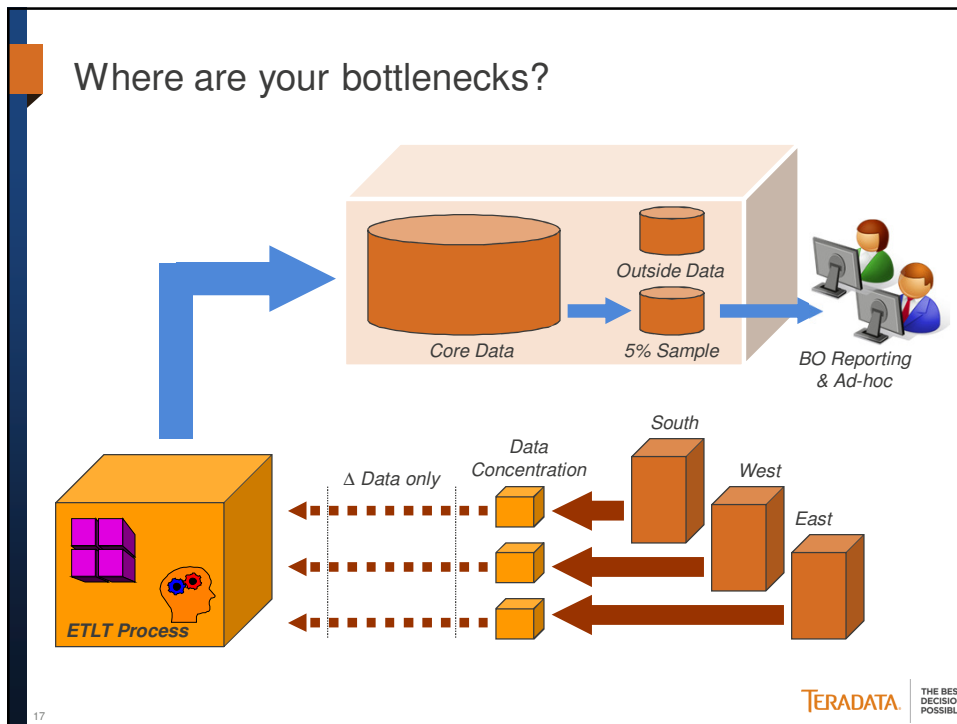


14



TERADATA THE BEST DECISION POSSIBLE

Summaries and Extracts were supposed to drive faster response but ultimately lead to larger lag times between data creation and relevant action







User and query freedom is not the absence from responsibility but the adherence to shared responsibility

19



Usage Type	IT Involvement	Service Level Expectation	User Responsibility
Push button	Highly. Optimized query and prioritization	< 2-5 seconds	Agree with SLA policy
"Canned" Ad-hoc	High, query path known, included in WM	< 10 seconds *	Define paths
"Planned" Ad-hoc	High, query path known	< 30 seconds *	Define path and payback
Company Analytics	Medium	Minutes	Agree / Define inputs and metrics
User written	Little to None	None	Know SQL and optimization

* May be dependent on quantified rows

20

The data warehouse is supposed to provide consistent data yet what it really does is highlights the inconsistency of the data

21

Do you have redundant data? (you wish you did!)

App 1	Cust. Id	First Name	Middle	Last Name	DOB	SSN	Address
	30391-244	William	James	Sosulski	04/12/39	563-49-1234	123 Oak St., Eves, IL 30319



App 2	Cust. Id	First Name	Middle	Last Name	DOB	SSN	Address
	30391244	William	J.	Sosulski	4-12-39	563491234	123 Oak St., Eves, IL

App 3	Cust. Id	First Name	Middle	Last Name	DOB	SSN	Address
	14239	Bubba	J.		April 12		BubbaJ@bubbagroup.com

App 4	Cust. Id	First Name	Middle	Last Name	DOB	SSN	Address
	3721B	William	James	Corp.	April 12	56349123	3224 Pkwy G, Los Osos

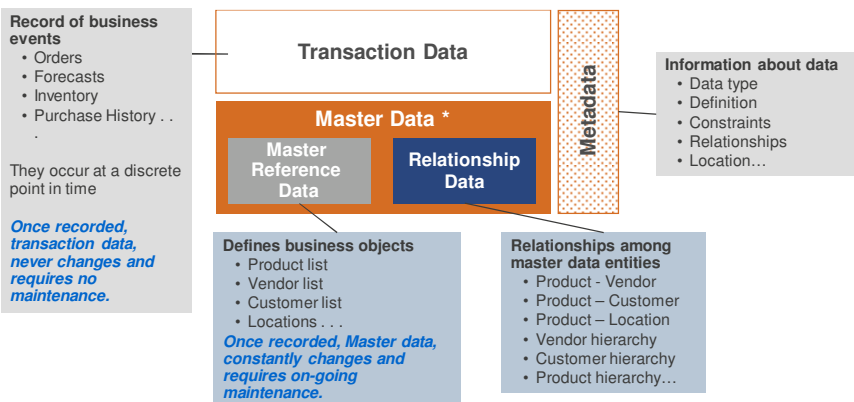
App 5	Cust. Id	First Name	Middle	Last Name	DOB	SSN	Address
	30391-244	William	James	Sosulski	04/12/69	563-49-1234	123 Oak St., Eves, IL 30319

- > Multiple mailings to the same address = ineffective and costly communications
- > Inconsistent and ineffective service
- > Inaccurate individual based action

22

Master Data Management is critical to success



*Can either reside within or outside the core data warehouse

A successful data warehouse program is delivered through projects but project funded data warehouses will not succeed

<div style="text-align: center; margin-bottom: 10px;"> Program view </div> <div style="text-align: center; margin-bottom: 10px;"> <h3 style="color: #2980b9;">Future value</h3> </div> <p>Centralize data to provide as much information to as many users to enable maximum information leverage.</p> <ul style="list-style-type: none"> ▪ Single source of truth ▪ Easy to reconcile ▪ Data protection is simplified ▪ Easier to manage data quality ▪ Lower Group IT cost and complexity <p style="text-align: center;">Project funds Program benefit not easily attributable to the current project</p>	<p>VS.</p>	<div style="text-align: center; margin-bottom: 10px;"> Project view </div> <div style="text-align: center; margin-bottom: 10px;"> <h3 style="color: #2980b9;">Current cost</h3> </div> <p>Give me what I need as quickly and cheaply as possible</p> <ul style="list-style-type: none"> ▪ Pay for data required ▪ Pay for quality required ▪ Quick turn around time ▪ Reduce delivery risk ▪ Use familiar technology <p style="text-align: center;">Program creates future cost, which is not attributable to the current project</p>
---	-------------------	--

THE BEST
DECISION
POSSIBLE

Information Strategy

Govern	Collaborate	Invest	Organise
<ul style="list-style-type: none"> • Align the project portfolio to achieve Group outcomes • Endorse policies and standards • Govern and prioritize data quality issues and escalations 	<ul style="list-style-type: none"> • Drive cross-BU collaboration and synergies 	<ul style="list-style-type: none"> • Fund initiatives proposed by BUs • Initiate (and seek funding for) projects where there is a lack of progress against the Strategy 	<ul style="list-style-type: none"> • Optimize the operating model around our Information Assets • Refine funding /recharge models • Align KPIs with outcomes

THE BEST
DECISION
POSSIBLE

1 If the new data is "core data" then a fully integrated design is called for
* see next slide

2 Who is interested in the new data?

Is the new data best used with IM data?

Yes	A preferred option Fully integrated design e.g. Falcon's FX data load from WallStreet System \$\$\$	B exemption required Semi integrated design e.g. NetBank Calendar \$\$
	C exemption required Isolated design e.g. external data sources like Bloomberg, Roy Morgan \$	D exemption required Isolated design or off GDW e.g. Group Security's case management data feeds \$
No		

3 **E** Data labs are used to test and learn as needed

- A** New data that lots of users want access to and that is to be used with other data in the warehouse should be conformed, and loaded to the integrated model (IM). This is the preferred option.
- B** New data that is of value to only select users and is to be used with other data in the warehouse, with exemption, may be loaded to the semi integrated model. An exemption is to be granted.
- C** New data that doesn't require full integration with the rest of the data in the warehouse may be loaded to isolated tables. However, an exemption is to be granted.
- D** New data that doesn't require full integration with the rest of the data in the warehouse may be loaded to isolated tables. However, an exemption is to be granted.
- E** Customers are permitted to create data labs as required to understand what the data contains and how it can be put to use. However strict controls are to be in place to prohibit misuse of data labs.

27

TERADATA THE BEST DECISION POSSIBLE

If all else fails, the ultimate paradox prevails....

There is never enough time to do it right but always enough time to do it over

28

TERADATA THE BEST DECISION POSSIBLE