Building a Business Intelligence Architecture Fit for the 21st Century

Jon Page
Building a Business Intelligence Architecture Fit for the 21st Century

Jon Page

This seminar is based on the Contents of this book

The premise is that:

BI is not delivering what is needed

Current BI architectures are not appropriate

Business and IT need to work Together on agreed strategies

Technology must be used to enable the reduction of copy management and the support of real time accuracy of critical data

People and organisation are Key to success in BI
Our Objectives

To record some history – what has happened in the past that makes the future quite challenging.
To provide real examples of BI at work – good and bad.
To illustrate the nature of data and why it has become so important in driving forward the business in the 21st century.
To outline a way to align technology with the business so that efforts and budget are spent in a way that will enable the future rather than support the past.
To propose a set of principles and ideas that can guide a company in a way to make data available to all who have the penchant to turn it into useful and valuable information.
To describe the new organisation unit that will be needed to realise the dream.
The Changing Business Climate

Globalisation

Deregulation

Technology

CHOICE
Competition: Very aggressive
To Be Successful – Two Critical Components

- Blue Chip
- The Winners!
- Out Of Business (slowly)
- Out Of Business (fast)
Maximising Customer Value

- Number of Relationships
- Relationship Duration
- Current Customer Value

Full Potential
Imagine You Could Do This...

Profitable

Not Profitable

Now

Future
BI is an afterthought
BI is reporting
BI is not taken seriously
Everyone does their own ‘bi’
Lots of tools, lots of platforms
No-one trusts the data
No-one understands the data
No-one shares information
No-one owns the problem
Business and IT don’t work together
There is no BI Strategy

And on, and on, and on
Customers

Time

Get some technology

Get more Technology – and fast

HELP!!
Data Warehousing Terminology

- MIS
- TNF
- Star Schema
- Data Warehouse
- Dimensional
- Business Intelligence
- E-Intelligence
- Data Sandpits
- DSS
- OLAP
- ROLAP
- MOLAP
- Snowflake
- Data Mart
- Strategic Information Systems
BI technology S-Curve

Product Performance vs. Time or Engineering Effort

- Executive Information System
- Independent Data Mart
- Enterprise Data Warehouse
- Data Provisioning Platform
Executive Information System (EIS)

- Very few users, typically management
- Standard reports, usually monthly
- Often run on production system
- High IT requirement
- Full centralized control, minimal user flexibility
- No persistent analytic data store
- Minimal data integration

Dimensional
Proprietary
PC based (if you’re lucky)
Data Warehousing: Independent DMs

- Rise of the power user
- Query capability and analysis added
- Dimensional systems; typically departmental
- Introduction of BI tools and normalization for more user flexibility
- Rise of data redundancy and synchronization problems
- Heavy management overhead
- Data integrated from multiple sources via ETL

Dimensional (star schema)
OLAP
SMP based
Data Warehousing: Enterprise DW

- Eliminates redundant data; single version of the truth
- Ad hoc query and cross-functional analysis possible
- Reduces ETL costs; continuous data loading possible
- Re-centralizes control
- Integrates data from multiple sources, but requires agreement on a single data model
- Complex to manage, not adaptable to change

Normalised ROLAP, MOLAP MPP
Data Provisioning? Operational BI??
Cloud???
Choosing the Right Architecture

"Data Warehouse" Topology Choices

1. Virtual Data Warehouse (i.e., Universal Data Access)

2. Lots of Data Marts

3. No User Access to DW

4. Selected Data Marts and Direct User Access of the DW
When to build data marts?

- Geographical considerations.
- Specific performance requirements.
- Specific availability requirements.
- Departmental control over reports and queries.
- Specialized applications.

- The Three P’s
  - Performance, Politics and Packages
DM Topologies – Independent DMs

**Advantages**
- “Fast” implementation.
- Quick ROI.
- Departmental control.
- Not reliant on IT for data.

**Disadvantages**
- Multiple data models.
- No consistent corporate data model.
- Multiple interfaces to manage/maintain.
- No single version of the truth.
- Duplication of data.
DM Topologies – Dependent DMs

**Advantages**
- Single version of the truth.
- Clean/scrubbed data.
- Consistent data model.
- Robust data transformation.

**Disadvantages**
- Must have an existing data warehouse.
- Must fit with corporate strategy.
- Duplication of data.
If you have an EDW and you want to propagate data to Data Marts, you have a choice:

1) Propagate data to distributed Data Marts...
   - Original data is stored in EDW.
   - Propagated data is stored in different physical sites.

2) Propagate data to integrated Data Marts...
   - Original data is stored in EDW.
   - Propagated data is stored in separate “databases” within the EDW environment
If you have an EDW and you want to achieve the Data Mart result without creating more databases to maintain, you can create Materialised Views within the EDW.

- The data remains stored in the EDW.
- MV’s can provide superior performance.
- Applications can use logical model of the data.
DM Topologies - Federated DMs

IT Users
Operational Data
Data Transformation
Enterprise Warehouse & Management

Visual/Intelligent/Magic Joiner

Business Users
Question???

What do all of these architectures have in common?
Failure is Always Blamed on Technology

• The Computer is too slow
• The Database can’t scale
• SQL is not powerful enough
• The Database doesn’t have the functionality I need
• The data is not to be trusted (not upto date)

• Nearly always technology is blamed for faults in people/organisation
• There’s really nothing wrong with technology today...
Moore’s Law Illustrated

‘Every 20 years the cost of processing a single bit of information declines by a factor of 1,000’
Questions?

“Even if you’re on the right track, you’ll get run over if you just sit there.”

Will Rogers
Building a Business Intelligence Architecture Fit for the 21st Century
Part 2

Jon Page
So it seems that we still have a problem? The trouble is that BI doesn’t stand alone – it feeds some very complex processes
Linking BI Strategy and Corp Strategy

WHY???

• Raise visibility of BI
• Build required organisation
• Ensure BI solves REAL business issues
• Qualify and Prioritise BI activities
• Get funding
• Get recognition for a job well done
Corp Strategy Looks Something Like This:

- Vision
- Goals
- Objectives
- Strategies
- Project
Projects can be broken down into Facets:

- Strategies
- Projects
  - Process
  - Technology
  - Applications
  - Organisation
  - Information

(Now we’re moving into BI Strategy)
The Information Facet can be broken down into Business Questions:
Business Questions can be broken down into SQL and Data
Data and SQL can be regrouped into meaningful and discreet BI ‘tasks’
What is a Pebble?

They are opportunities that:

can be enabled by a technology based solution
need to exploit information
can be realised in no more than six months
are well understood by business and technology
are of high priority
are measurable in terms of success whether this be ROI, TCO or some other internalised mechanism
have a descriptive and meaningful name.
solve at least some part of the information facet of a project.

*Perhaps the most important characteristic of pebbles is that they support agreed and strategic projects*
Real ‘Pebbles’

- Management of churn in a specific segment of customers – let’s reduce churn by 5% in our customers who make more than 5 calls per day
- Revenue recognition – let’s make sure that we improve by 5 points the ratio between service usage, billing and collection across the board
- Increase acquisition rate of customer in a specific segment – let’s increase rate by 2% for corporations with 50 to 300 employees.
- Decrease fraud – let’s reduce subscriber fraud by 3 points across all customers
- Understand product affinity – let’s determine which products are bought in combination to allow us to more effectively deploy ‘loss leaders’.
Let’s reduce churn by 5% in our customers who make more than 5 calls per day

- Questions (SQL):
  - Who makes more than 5 calls per day
  - What is the churn rate in this segment?
  - What do the churners have in common?
  - Who in the current segment looks like these churners?
  - How can I keep them?
Global Information Technology Planning

Process
Applications
Organisation
Technology
Information

Pebbles
Business Questions
Data
  Critical
  Supporting
  LDM
  PDM
Prioritising Pebbles

- **High/Low**
  - High Strategic Business Value
  - Low Need for IT Support

- **High/High**
  - High Strategic Business Value
  - High Need for IT Support

- **Low/Low**
  - Low Strategic Business Value
  - Low Need for IT Support

- **Low/High**
  - Low Strategic Business Value
  - High Need for IT Support
Global Information Technology Planning

<table>
<thead>
<tr>
<th>Customer Segmentation</th>
<th>Revenue Assurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage Analysis</td>
<td>Customer Profitability</td>
</tr>
<tr>
<td></td>
<td>Risk Analysis</td>
</tr>
<tr>
<td>Value</td>
<td>Statutory Reporting</td>
</tr>
<tr>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Sales Reporting</td>
</tr>
<tr>
<td></td>
<td>Fraud Detection</td>
</tr>
<tr>
<td></td>
<td>Competitor Assessment</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Customer Segmentation</td>
</tr>
</tbody>
</table>

Difficulty
Let’s look at a worked example

Workbook 1
BI Strategy Documents and Owners

Key Information Imperatives – owner Information Master
Business Intelligence Roadmap – owner Information Mast
BI Pebble Definitions – owner Information Manager
Data Dictionary – owner Information Master
Skills Matrices – owner Information Master or Information Manager as relevant.
Security Policy – DBA
Access Policy – DB
Information Architecture – Information Architect
Data Reuse Map – Database Designer
Logical Data Model – Database Designer
Physical Data Model – Database Designer
User Requirements – Information Creator
Various technical documents - various
Various Service level Agreements – Information Manager/Master as appropriate
Building a Business Intelligence Architecture Fit for the 21st Century

Part 3

Jon Page
What Have We Learned So Far?

BI has not matured as fast as we maybe believe
The gap between what we need and what we have is getting bigger
The goal is to be able to predict and maybe change the future
Business and IT MUST work together
A BI Strategy linked to Corp Strategy is vital
Contemporary BI platforms – Data Marts and Data Warehouses are a good start
but that’s all. They’re:

Full of useless inaccurate data
No one knows where the data came from or how old it is
It’s copy miss-management gone crazy

We need something that:
Is trusted
Manages a single copy of master data
Can provide different ‘types’ of info – strategic, regulatory, ad-hoc, DSS, history, trend
Recognises that data is of many natures
Design    Make     Distribute    Retail    Sell    Distribute    Bill    Pay

Pay                                                                                               Receive
Question 3 – What are the main differences between these two sets of data?
We need to be able to store and manage many different types of data automatically

- Transactional, detailed, historical - vast volume
- Back-office, operational Data - on-line?
- Master Data - on-line
- Regulatory Data - accurate
- Summarised Data - automatically created
- KPI’s - automatically maintained – complex algorithms
- Metadata - in one place, meaningful
What We Need to do Now:

- Build a single BI Platform that:
  - Stores and Manages Critical Data (Master Data) in a real time way and as a service to applications
  - Stores and makes available detailed, historical, transactional data
  - Manages automatically all summary data
  - Provided fresh KPI’s
  - Is the basis for trusted and repeatable regulatory reporting
  - Services ad-hoc complex queries and transactional queries
  - Is self managed
Collaborative Information Blueprint

Some fundamentals:

It depends on collaboration between Business and IT
It needs a BI strategy as described in Part 2
It builds on what we’ve learned in Data Warehousing
It is an incremental strategy
It is a centralised, data driven initiative
It’s a set of guiding principles rather than a methodology

 Depends on a centralised platform that is TIGHTLY coupled with the operational world
Steps 1 to 4 of CIB

Plan

Define Key Information Imperatives (KII)
Define/Prioritise BI Pebbles (BIPs)

Identify Critical Data Elements (CDEs) and build Data Model

Determine Data Rules

Implement
Step One – Information Imperatives

We must deploy all reporting across the web.
We will rigorously abide by all pertinent standards at all times.
We will keep a single copy of important data accessible to all.
Wherever possible we will pursue a single vendor strategy.
We will not use bleeding edge technologies to support core processes.
We will forbid the copying of data unless suitably authorised.
Non-compliant applications will only be used if a pathway to full compliance is possible within an 18 month window.
We will aggressively centralise all important data.
Step Two – Determine and Prioritise Pebbles
What is a Pebble?

They are opportunities that:

can be enabled by a technology based solution
need to exploit information
can be realised in no more than six months
are well understood by business and technology
are of high priority
are measurable in terms of success whether this be ROI, TCO or some other internalised mechanism
have a descriptive and meaningful name.
solve at least some part of the information facet of a project.

*Perhaps the most important characteristic of pebbles is that they support agreed and strategic projects*
Real ‘Pebbles’

- Management of churn in a specific segment of customers – let’s reduce churn by 5% in our customers who make more than 5 calls per day
- Revenue recognition – lets make sure that we improve by 5 points the ratio between service usage, billing and collection across the board
- Increase acquisition rate of customer in a specific segment – let’s increase rate by 2% for corporations with 50 to 300 employees.
- Decrease fraud – let’s reduce subscriber fraud by 3 points across all customers
- Understand product affinity – lets’ determine which products are bought in combination to allow us to more effectively deploy ‘loss leaders’.
Step Three) Identify CDE and Model

<table>
<thead>
<tr>
<th></th>
<th>Customer</th>
<th>Invoice</th>
<th>Bill</th>
<th>Office</th>
<th>Payment</th>
<th>Supplier</th>
<th>Product</th>
<th>Credit Score</th>
<th>Employer</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIP1</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIP2</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIP3</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIP4</td>
<td></td>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIP5</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIP6</td>
<td>y</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Now We Have A Clear Justifiable View on what is OUR:

MASTER DATA – (working definition for this seminar!!!)

The Data That Supports Most of our Critical BI Requirements (Pebbles)

This is KEY – because I propose that we store this data (at least), in just one place and one place ONLY!
Step 4 - Define Rules Around CDEs

- **Sample Business rules may include:**
  - Account Open Date must be 12 years later than Birth Date of Account Owner
  - Customer Name cannot be ‘null’
  - A company must be allocated a SIC code from the published list
  - An account is ‘open’ if any type of contact has been associated with this account in the past 20 days
  - Address must always have a post code

- **Sample Technical rules may include:**
  - Account Number must be 6 digits
  - ‘Male’ will always be represented as ‘M’
  - Value must be capitalised
  - Value must be from following list......
  - All table names should be singular

- **Process rules may include:**
  - If there are more than ten of these marked ‘O’ then....
  - This attribute is the result of function X on data item Y
  - This cannot be deleted whilst.....
After Phase One we have:

- A prioritised map for BI implementation
- A set of IT Imperatives to guide architecture, infrastructure, purchasing etc
- A clear definition of Master Data, including rules etc
- A core Logical Data Model
Steps 5 to 8 of CIB

Implement

Enable Compliance

Select and Deploy Critical Data Platform (CDP)

Implement Pebbles Incrementally

Build end-user layer

Automate and Secure
Step 6 - Technology must be selected to create the Critical Data Platform (CDP), and this will include at a minimum the:

- Computing platform (hardware)
- Operating System plus management tools
- RDBMS
- Communications capability
- Business Intelligence tools (reporting, ad-hoc query, mining, OLAP)
- Integration tools
- Metadata Management tools/capability
- Backup and Recovery capabilities
- Security software
Question 5:

WHO CAN PROVIDE THE PLATFORM?

One ‘computer’
One Data Base
CDE’s on-line to operational systems
All data feeds automatic, in and out
Step 7 – Implement Pebbles Incrementally

Q1          Q2         Q3         Q4         Q5         Q6

Customer
Segmentation

Revenue
Assurance

Usage
Analysis

Customer
Profitability

Risk
Analysis

Customer
Segmentation

Statutory
Reporting

Sales
Reporting

Fraud
Detection
Physical Database Implementation

Deployment of Integrity rules

Develop Queries And Reports

Initial Data Load

Determine Data Source

Implement Interface Technology

Test

Acceptance
Step 8) Enable Compliance

- Get the right answers
- Get these answers quickly and reliably
- Demonstrate how numbers are calculated (audit)
The Implement phase brings into a single database:
- Key shared data for enabling Corporate strategy (often called ‘Master Data’ but actually much more.
- Operational data needed by multiple operational applications
- Derived data for complex algorithms and scoring
- Summary data often used to form the basis of Compliance calculations as Key Performance Indicators
- Detailed, historical, behavioural data for deep mining and predictive analysis.
So in our preferred architecture, the tables we have modelled to represent our Critical Data Elements will be implemented in a single database and as a set of relational tables.

- As these tables are deemed critical to the enterprise it is imperative that:
  - We have only one set of these tables in the whole enterprise.
  - These tables are shared by all applications that need this data whether these applications are internal or external, operational or analytical.
  - This data is protected from miss-use by a defined layer of metadata.
  - The data in these tables is always up-to-date.
  - The data in these tables records history.
Step 9 and 10 of CIB

Automate and Secure

Ensure 100% Availability

Automate and Secure
A Cluster for Reliability?
Remote Backup?
Critical Data Elements, plus other as needed

Source Systems

BI Platform

BI Pebbles
Critical Data Elements
plus other as needed

Source Systems

BI Pebbles

BI Platform
Critical Data Elements
plus other as needed

Source Systems

BI Pebbles

BI Platform
Business Value

Credit Risk Analysis

Product Bundling and Pricing

Profitability Analysis

Behavioural Analysis

Information Infrastructure

Additional Data
Summary

- Define and prioritise requirements – already done in our BI Strategy
- Define first deliverable – already done in our BI Strategy (see BI Pebble Road Map)
- Build Logical Data Model for first Pebble - already done
- Build Physical Model to support initial requirement
- Deploy rules as triggers, procedures, macros etc.
- Define preferred provider system for each
- Define interface requirements of each item of data with provider system
- Perform initial data load to database
- Deploy the required data feed technologies
- Test system
How do we pay more for less

• Tailoring pre-canned applications
• Building our own logical data model
• Not using the technology that we have already paid for
• Buying data we already have
• Issuing RFIs and RFPs
• Evaluating identical technologies

• **Copying data**
Summary

- Build a comprehensive (but flexible) BI strategy linked to the Corp Strategy
- Develop IT Imperatives
- Put in place the right organisation
- Cherish Critical Data
- Build a closer but more demanding relationship with your vendors
- Use technology
- Promote yourselves – BI is the most important facet of most internal business
Building a Business Intelligence Architecture Fit for the 21st Century

Part 4

Jon Page
BI Projects fail for many reasons, we continually see:

Insufficient dedication of resources (technical and most importantly, human).

People with the wrong skills.

Multiple, unprioritised roles and responsibilities.

Lack of decision making power.

Rudimentary future planning.

Conflicting commitments.

Too many on-going projects
A Real DW Audit findings:

Lack of business involvement/ownership
No three year roadmap for expected deliverables
No formal Data Architecture or Logical Data Model
Poorly defined job descriptions/roles – confusion of responsibilities
Non-optimal deployment of Hardware
Using out of date software
No consistent application of Parallel Database features
Many important Oracle facilities not being used effectively
Lack of documentation
Non-scalable environment
Critical data was missing
No deep usage of data in terms of analytics/data mining etc
Define Idea → Identify Data → Estimate Cost → Go or No Go → Implement Interfaces → Create and Test Report → Operationalise

LEARN!
The Information Board

- Creation of Corporate Strategy
- Ensuring alignment of BI Strategy
- Providing support for deployment of CDP
- Providing guidance and support to the BI ‘team’
- Solving critical issues
- Arbitrating on internal conflicts
- Monitoring return on Investment and/or Total Cost of Ownership goals
- Appointment of the Information Master
The Information Master:

is ultimately responsible for the creation and maintenance of the following documentation:

Key Information Imperatives
BI Roadmap
Data Dictionary
The complete Business Intelligence Strategy
Skills Matrices for the BI team
BI Strategy Documents and Owners

Key Information Imperatives – owner Information Master
Business Intelligence Roadmap – owner Information Master
BI Pebble Definitions – owner Information Manager
Data Dictionary – owner Information Master
Skills Matrices – owner Information Master or Information Manager as relevant.
Security Policy – DBA
Access Policy – DB
Information Architecture – Information Architect
Data Reuse Map – Database Designer
Logical Data Model – Database Designer
Physical Data Model – Database Designer
User Requirements – Information Creator
Various technical documents - various
Various Service level Agreements – Information Manager/Master as appropriate
How do we pay more for less

- Tailoring pre-canned applications
- Building our own logical data model
- Not using the technology that we have already paid for
- Buying data we already have
- Issuing RFIs and RFPs
- Evaluating identical technologies

**Copying data**
Summary

- Build a comprehensive (but flexible) BI strategy linked to the Corp Strategy
- Develop IT Imperatives
- Put in place the right organisation
- Cherish Critical Data
- Build a closer but more demanding relationship with your vendors
- Use technology
- Promote yourselves – BI is the most important facet of most internal business