The Data-driven Business and other Lessons from History

Executive Summit, Strata Conference
February 1-3, 2011, Santa Clara

Dr Barry Devlin
Principal, 9sight Consulting

Dr. Barry Devlin is a founder of the data warehousing industry and among the foremost authorities worldwide on business intelligence (BI) and beyond. He is a widely respected consultant, lecturer and author of “Data Warehouse—from Architecture to Implementation”. Barry has 30 years of experience in the IT industry, previously with IBM, as an architect, consultant, manager and software evangelist.

As founder and principal of 9sight Consulting (www.9sight.com), Barry provides strategic consulting and thought-leadership to buyers and vendors of BI solutions. He is currently developing a new architectural model for fully consistent business support—from informational to operational and collaborative—Business Integrated Insight (BI²). Recently relocated to Cape Town, Barry works both locally and internationally.

Email: barry@9sight.com (preferred contact method)
Phone: (S. Africa) +27 71 557 7479
         (Ireland) +353 86 237 5128
Agenda: The Data-driven Business and other Lessons from History

1. Past – origins of Data Warehousing
   - The foundations and evolution of the architecture

2. Present – Business and technical challenges
   - With special focus on the information explosion

3. Future - a new architecture?
   - Business Integrated Insight …and a new balance

―That men do not learn very much from the lessons of history is the most important of all the lessons of history.‖  Aldous Huxley

―The Past lies upon the Present like a giant’s dead body‖  Nathaniel Hawthorne

THE PAST
The Original Data Warehouse Architecture (1988)

- Based on internal work in IBM Europe beginning 1985
- Data-driven business: “Users can focus on the use of the information rather than on how to obtain it.”
- “Business Data Warehouse (BDW)… is the single logical storehouse of all the information used to report on the business… the end user is presented with a view / number of views that contain the accessed data…”
- Raw & enhanced, detailed & summary, public & personal data all within a single component

The layered Data Warehouse since the early ’90s

“Big Data” performance led to two layers within the DW

1. Enterprise data warehouse (EDW)
   - Reconciled/cleansed data

2. Data marts
   - What the users need

- Other key characteristics
  - Vertical and horizontal segmentation of information
  - Separate metadata
  - Hard information only
  - Unidirectional data flow

- Well architected! But, rather regimented…

Operational systems

Data marts (Informational Apps)

Enterprise data warehouse

Metadata
Explosion of components since mid-'90s onwards creates serious problems.

- **Timeliness**
  - Multiple ETL stages delay data availability
- **Consistency / quality**
  - Inconsistent data – content and timing
- **Management costs**
  - Creation, maintenance and rework of multiple ETL and storage components
- **External, collaborative and soft information support as an “add-on”**
- **Data mart nirvana**
  - Anybody can have one
  - Chaos and innovative exploration?
- **And, hmmm… not so well architected!**

---

**Lessons from history**

1. **Information quality, reliability, consistency is key**
   - Enormous investment in data warehousing over 25 years
   - And other quality initiatives – ERP, MDM
2. **Data volumes, variety present big challenges**
   - Expectations and demands outstrip technology
   - Organizational and political issues hamper progress
3. **Exploration and analysis drives innovation**
   - From the first spreadsheets to cloud-based analytics
   - With ever-increasing data volumes and expectations

"History repeats itself because no one was listening the first time" – Anonymous
"And today? 
Today is a gift. 
That’s why we call it the present.”

Babatunde Olatunji

THE PRESENT

Three key trends in business are driving rapid change.

1. Closed-loop business – strategy to execution
   – Merge operational, informational & collaborative
2. Massive information volumes for use
   – Volumes, sources, types
3. Collaborate to innovate
   – Millennials move into power
Technology advances in recent years offer the possibility to address business needs in new ways.

1. Closed-loop business – strategy to execution
2. Massive information volumes for use
3. Collaborate to innovate

**Business**

1. Service Oriented Architecture
   - Deconstructing applications
2. Advances in “Data Processing”
   - Relational databases and beyond
     - Programmatic approaches
3. Web / Enterprise 2.0 and 3.0
   - Semantic web

---

What is hard and soft information?

- **Hard information (or data, “structured”, etc.)**
  - Computer oriented
  - Information meaning and values separate
  - Stored in tabular / relational / hierarchical form
  - Formally modelled: Metadata made explicit

- **Soft information (or content, “unstructured”, etc.)**
  - People oriented
    - Meaning and values intermingled
    - Stored in many, varied forms
  - Unmodelled or informally modelled: Implicit or tacit metadata

  “Unstructured information” is an oxymoron!

- Application and database design is the process that traditionally reconfigures soft into hard information
Information today… what has changed?

- **Volumes**
  - Hard Info: CAGR 22%
  - Soft Info: CAGR 60%
- **Type**
  - Hard Info: 15% in 2005
  - As low as 5% in 2010
- **Source**
  - Internal → Largely external

Many assumptions about business intelligence and analytics must change.

- No longer possible to route all data through EDW
  - Volumes too large and mix of hard and soft info
- Not possible to convert / store all soft info as hard info
  - Volumes too large
- Reduced trust in information quality
  - Uncertified quality of externally sourced or repurposed data
- A “single version of the truth” is no longer achievable
  - Information is too varied, business needs too diverse
- Unfettered analysis becomes unreliable
  - Varied, uncertified, voluminous external / repurposed data
- BI and analytics must integrate into business processes
  - Closed-loop business requires integrated innovation
A technology view suggests two approaches.

Databases – appliances
- Columnar, parallel, in-memory, and/or compressed
- Reduced footprint, setup
- Massive distributed storage
- Up to 100x faster queries
- Declarative design

Applications – massive parallelization
- MapReduce / Hadoop environment
- Open source
- Parallel infrastructure
- Procedural design

In fact, these approaches are complementary.

**Database-centric**
- Strong support for data quality and information creation
- Central point of consistency in multiple interrelated truths
- Hub for traceability, auditability
- Central information point for process-linked decisions (operational BI) and reporting
- Starting point for innovation

**Application-centric**
- Strong support for large, rapidly-changing data
- Distributed, variable data environment for mining
- Sandboxing
- Specific analytic processes that could be connected to SOA environment
- Starting point for integration
"The future, according to some scientists, will be exactly like the past, only far more expensive."

John Sladek

It’s time to re-visit the BI (and IT) architecture

- The old layers separate information
  - Operational from informational
  - Both from collaborative (never considered)
  - Hard information from soft information
- Leading to information problems:
  - Quality & maintenance of multiple copies
  - Incomplete, inconsistent and out-of-date
  - Time-delays in information availability

- The information resource must be considered as a whole
  - Complete, integral, timely
  - Easily and efficiently managed and maintained

➤ A single layer of all the business information
The axes of the BIR describe a logical information space containing *all* the information of the business.

Mind over Matter map: key to information meaning

- A further extension to the Knowledge density / Structure axis
- Based on deeper analysis of big data possibilities
- See my O’Reilly Radar article for further details

See: B-Eye-Network ongoing series of articles
First article (of 10 parts) [www.b-eye-network.com/view/12760](http://www.b-eye-network.com/view/12760)
A new layered framework…
Business Integrated Insight (BI²)


So, are you willing to see(k) the lessons of history?

Information quality, reliability, consistency is key
- The basis for all exploration is a reliable, consistent, high quality map of what’s already known

Data volumes, variety present big challenges
- There will always be more information to explore and more questions to answer than technology allows
- And technology is by far the least of the problems

Exploration and analysis drives innovation
- Business value lies in the unknown; business disaster lurks in the misinterpreted

"History is a cyclic poem written by Time upon the memories of man”
Percy Bysshe Shelley