The Evolution of Data Warehouse Automation

Barry Devlin, 9sight Consulting

April 16, 2015
Speakers

Barry Devlin
Founder & Principal, 9sight Consulting

Mark Budzinski
President, WhereScape
Dr. Barry Devlin, founder and principal of 9sight Consulting (www.9sight.com), is among the foremost authorities on business intelligence (BI), big data and beyond. He is a founder of data warehousing, having defined its first architecture in 1985. A respected visionary and thought-leader in the evolving data industry, Barry has authored two ground-breaking books: the classic "Data Warehouse--from Architecture to Implementation" and “Business unIntelligence--Insight and Innovation Beyond Analytics and Big Data” (bit.ly/BunI_Book) in 2013.

With over 30 years of experience in IT, until 2007 with IBM as a consultant, manager and distinguished engineer, Barry provides strategic consulting and thought-leadership to buyers and vendors of BI and Big Data solutions. He is an associate editor of TDWI's Journal of Business Intelligence, and a regular keynote speaker, teacher and writer on all aspects of information creation and use.

Barry operates worldwide from Cape Town, South Africa.
35 years of evolution of BI needs: ever bigger, always faster and increasingly complex

- **1985**: Data mining and basic BI querying
- **1990**: Access to closer to real-time data
- **1995**: Consolidating reporting across business lines
- **2000**: E-Commerce converges operational-informational
- **2005**: Web logs offer view of interactions – not just transactions
- **2010**: Social media data offers sentiment and relationships for marketing – predictive analytics
- **2015**: Devices on Internet of Things reveal individual behaviors and measures - instantaneous analytics

"Big Data Eclipse"
Gaining value from social media and the Internet of Things depends on “(sm)all data” management.

- The need for data warehousing continues to grow despite “data lakes”
  - Core business information
  - Consistent and integrated business management

- Business needs change faster than ever
  - Short iteration projects
  - Ongoing dev/maint process

- Business and IT must work together
  - Biz-tech ecosystem
The layered Data Warehouse has been a constant since the early ’90s.

Functional drivers

- Consistency across sources
- Cleanliness of base data
- “Single version of the truth”
- All increasingly needed in an always-connected, information-overloaded world

It has also long presented development challenges.

Working across the layers

- Project-specific business needs for data and function
- Enterprise data model and database design
- ETL transformation and cleansing (x 2)
- “Data archaeology”

➢ Automation is mandatory
Big Data support adds to the development challenge.

- Pillar architecture supports multiple data types
  - Data Warehouse + Operational Systems = Process-mediated
  - IoT = Machine-generated
  - Social media = Human-sourced

- More development challenges
  - Diverse Big Data sources
  - Access to data at source vs. Load to Data Warehouse
  - Assimilation of context and relationships across pillars
  - Data governance

➢ **Automation again mandatory**
Three development (and maintenance) challenges

- Highly iterative development cycle between business and IT
- Disparate and unconnected development tooling
- Multiple, unrelated stores of metadata
Data Warehouse Automation is an evolution from more manual, traditional approaches.

- Reduce number of tools and environments

- Create an integrated, agile requirements / design / development / maintenance environment

- Provide a common, shared store for all context-setting information (metadata) – business, technical and more

- Support full collaboration between business and IT
Automation of the full development and maintenance process increases productivity of business and IT.

- Traditional ETL approach separates design scopes
  - Disparate tools for modeling, design and ETL
  - Additional servers and licenses

- ELT (extract, load & transform) integrates the entire design scope
  - Single design and development environment
  - Benefits from performance of relational database
The demise of metadata… and the rise of context

- Metadata is two four-letter words!
  - Information (not data)
  - Describes all “stuff” (not just data)
  - Indistinguishable from “business information” by non-IT people (and some IT people)
  - Many (or most) metadata projects fail

- NSA popularizes/repurposes metadata
  - “It’s metadata, not personal info… so, we can collect it”… How ironic!

- Context-setting information (CSI):
  - From business meaning to information collected
  - From data stored to context understood
A common CSI (metadata) foundation for business and IT enables extensive collaboration.

- Requirements
- Data needs and availability
- Models and databases
- Queries and reports
- Etc.

Context-setting information
 Scope: Business & Technical

Interaction

Collaborative Team

Business Person

IT Person

Shared Information

Innovate

Interact
Balancing consistency with time to value engages business with IT in a joint pursuit.

- Integrated development process between business and IT
  - Requirements to data design and query in one session
  - Iterations over further joint sessions
  - Collaborative working approach
  - All context-setting information stored in one place

- All context is carried over to maintenance process
  - Maintenance as an extension of dev.
  - Business and IT involvement
  - Data quality as a focus
Conclusions

1. Big Data value and use depends on having core business information (Data Warehouse)

2. Data Warehouse Automation is key to timely and standardized development and maintenance

3. Shared context-setting information enables ongoing business/IT collaboration

4. Business/IT collaboration supports uncertain and changing requirements
Thank you

Questions?

Dr Barry Devlin
Founder & Principal
9sight Consulting
Using Automation Techniques to Satisfy the Thirst for Data

Mark Budzinski, President
markb@wherescape.com
Traditional Approach to Development

- Overcomplicated
- Too many tools
More with less - Automated Data Integration

Automated Data Integration software:

- that designs, builds and operates data warehouses
- automates repeatable best practice development standards
- with built in Data Governance
- using metadata to drive agile delivery
- with consistent quality and full documentation
- and delivers value, faster, to business & IT stakeholders
- that saves time and money for our customers
WhereScape RED Overview:

DATA SOURCES
- Data Warehouse
- ETL load file
- Databases
- Files
- ODBC
- XML
- Spreadsheets
- Hadoop

LOADING
- Load Tables

TRANSFORM
- Stage Tables

STORING
- Data Store Tables

MODELLING
- 3NF Models
- Star Schema
- Data Vaults

ACCESSING
- Views
- Join Indexes
- Aggregates
- OLAP Cubes
- Export

External Models
Existing Objects

Retrofit
WhereScape Approach

• Simplification
• Automation
• Data Driven!
Some of Our Customers
“The six-month project has now been completed, smashing the previous three-year forecast by using agile development to overhaul each data warehouse business area in ‘sprints’.”

Andy Ruckley, Head of Technology – Data Platforms, Tesco PLC
“WhereScape is enabling us to get value from sensor data and shorten times to market; we are able to deliver our BI solutions faster than ever before.

Using WhereScape RED, what used to take 1 hour coding by hand now takes 6 minutes using Data Automation.”

Stijn Roelens, Enterprise BI Architect, Volvo Trucks
“WhereScape has accelerated and amplified our output by orders of magnitude. A week of development in our prior environment can now be done in 30 minutes using WhereScape RED.”

Dana Keith, manager of applications and data warehousing – WVNET
“Our results using WhereScape have been extremely impressive. WhereScape enabled us to design, develop, document and deploy a production-ready solution in 8 weeks. Using traditional data warehouse development methods would have taken us 6-8 months.”

Ryan Fenner, VP, Data Solutions Architect – Union Bank
Automated Data Integration from WhereScape
us.wherescape.com
Questions and Answers
Contacting Speakers

• If you have further questions or comments:

  Barry Devlin
  barry@9sight.com

  Mark Budzinski
  mark.budzinski@wherescape.com