



Building a Data Warehouse? Jump Off a Cliff!

Want to Seriously Throw Good Money After Bad? Start a BI Project.

Ian Nicholson
Director

BIReady Pty Ltd
Sydney, Australia

info@biready.com.au

www.biready.com.au

Reissued July 2016.

Building the Data-Driven Business
Architecture and Tools for Smart Operations,
Analytics and Information

Dr. Barry Devlin
Sydney
August 24-25



ajilius
Data Warehouse Automation

Build Data Warehouses FAST!
Site license - US\$10.00opa.
Book a demo today.

Today's business cannot afford to wait weeks, months or even years to get the lifeblood information it needs to make better decisions. When it comes to Business Intelligence, your biggest challenges are:-

- Creating more business value
- Lowering your costs
- Empowering your users with self-serve reporting
- Delivering information anywhere, anytime
- Reducing complexity

Everyone accepts the norm when it comes to building enterprise Business Intelligence solutions. You need to build a data warehouse first. Traditionally, this has been done by teams of experts using ETL tools to Extract, Transform and Load the data from their sources into another database (the data warehouse or data mart) in a way that makes it easier for the business to perform reporting and analysis.

On a group thread on LinkedIn recently I witnessed several consultants arguing over which ETL tool is best for building data warehouses.

They all discussed the same, tired old approach they've been doing for years, completely oblivious to a few key stats:-

1. ***According to Gartner, 76% of the budget of any BI project is sunk into the data warehouse.***
2. ***Also, according to Gartner, 70-80% of BI projects fail.***
3. ***Of those that succeed, many are over time or over budget.***

And yet these same consultants are continuing to use those same 'best' ETL tools and approaches to 'deliver' BI projects, just like lemmings jumping off a cliff.

So what's going wrong? Let's look at what's currently happening...

You decide to start your BI project. It is deemed that a Business Analyst (1st consultant) talks to the business to 'get an understanding' of their requirements.

The BA give her results to a Solution Architect (2nd consultant) who creates the design and gets a Data Modeller (3rd Consultant, but possibly the same guy) to design a data model. Now the likelihood is, he will design a set of star-schemas - because that's what you'll need for the reports, right? Also, they can be delivered straight away for a quick win with the business.

So the Data Modeller gives his design to a DBA (4th Consultant) to create the schema in the data warehouse server, and to the ETL developer (5th Consultant) to write the ETL code to populate the schema in the data warehouse. Once the schema is present, a report is written by the BI Developer (6th Consultant).

Of course, all of this is overseen by the Project Manager (7th Consultant).

Building the Data-Driven Business
Architecture and Tools for Smart Operations,
Analytics and Information

Dr. Barry Devlin
Sydney
August 24-25



ajilius
Data Warehouse Automation

Build Data Warehouses FAST!
Site license - US\$10.000pa.
Book a demo today.

Now, fast forward a few months to the first deliverable being revealed to the business. More often than not, the solution isn't what the user expected, wanted or signed-for. Changes are required. Why? Usually, a business user doesn't understand what is possible or what is required in a data warehouse. Only when they see a result do the pennies drop, and they realise they really needed something else. Either that, or during the time it took to deliver the report, the world moved on and focus changed.

Whatever the reason, you must go back through that convoluted process above to make the changes. Now, did I mention that when the ETL Developer is writing his code, he has a problem? If he needs to get data from lots of data sources, he must do some further 'manipulation' in order to shoehorn it into a star-schema (a star-schema, by the way, is a table design made up of facts (things we want to measure: Invoice amount, quantity etc.) surrounded by dimensions (how we want to sort and slice that data: Customer, Region, Product, Date, Sales Rep).

Put simply, **you cannot take data from multiple sources and just put it into a star-schema**. A lot things must happen first. Bill Inmon, the Father of Data Warehousing, has written numerous books on this science, but in a nutshell he says you should make the data generic and put it into its Third Normal Form. Few people know how to do this properly, and this is where, I believe, that 76% of money and 80% failure rate begins. Your ETL Developer writes a lot of convoluted code which only he understands which becomes a maintenance nightmare when changes are required.

Now, even if the project is using best practices and designing a Third Normal Form Enterprise Data Warehouse, they will continue to use outdated ETL tools to try to do the job. Even though they have a better architecture, they use the same SDLC approach that ETL tools force you to follow. As a result, they take much longer to deliver and at a greater expense. They also require more specialised skills (read: expensive) to understand and build them, which is why, I suspect, many companies avoid building them.

So how do we fix this? Well, I should be clear that Software Development Life Cycle (SDLC) as a development approach does NOT work for BI projects, since such projects tend to be fluid in nature and more of a journey than a destination.

Agile development methodologies yield far more success in BI projects. Current ETL development tools are too limited and cumbersome in their approach. There is little if no automation available from these products. They are development tools, originally designed just to move data from one place to another, not build data warehouses.

ETL tools do not build data warehouses. People using ETL tools build data warehouses, and ETL tools simply cannot deliver in an Agile fashion.

The good news is that the design of Third Normal Form Enterprise Data Warehouses (or the newer variant, Data Vault) involves well-defined processes, and therefore, can be automated.

Nowadays there is a new breed of Data Warehouse Automation tools available that do the job much faster - giving deliverables in hours or days, not weeks or months. In doing so, they slash development times and costs to a fraction of what they are currently.

Building the Data-Driven Business
Architecture and Tools for Smart Operations,
Analytics and Information

Dr. Barry Devlin
Sydney
August 24-25



ajilius
Data Warehouse Automation

Build Data Warehouses FAST!
Site license - US\$10.000pa.
Book a demo today.

Data Warehouse Automation tools work by first analysing the source data **model**, and then automatically designing the required schema for it in Third Normal Form (or Data Vault), applying the teachings of Bill Inmon to the design process. They then generate the code to populate it.

Once a Third Normal Form or Data Vault construct is in place, then it is a simple process (usually performed using a wizard) to derive the required star-schemas for reporting.

The design process can therefore be crushed down from weeks (in the old way) to minutes. Even with multiple data sources, there is some mapping-type work to be done in order to join the different source models together, but the process is achieved much faster using Data Warehouse Automation tools than combined data modelling and ETL tools.

Using Data Warehouse Automation tools saves an enormous amount of time, money, and delivers value sooner while virtually eliminating risk of project failure.

As we hurtle into the age of Big Data, we need to get our house in order when it comes to data warehouse development. ETL tools – no matter how expensive – simply do not cut it anymore in the 21st century for the simple reason they cannot deliver at a speed expected or needed by a 21st century business.

About the Author

Ian Nicholson started his IT career programming MUMPS on PDP-11s in 1982. His first BI role (although it wasn't called that back then!) was writing reporting programs for various healthcare applications. In 1984 he designed and coded the FEMAC reporting solution used by Dow Chemical used to report imported product data to the Dutch Customs. The Dutch Customs liked the design so much, they set it as their required standard for other companies to use. He worked as a database and programming consultant for many years in the UK, Holland and Switzerland before moving to Australia.

In 2001, he made the jump to Business Intelligence sales, selling Crystal Reports. Since then he has worked for various BI/DW firms selling Hyperion, Cognos, Tableau, WhereScape, BusinessObjects, Yellowfin, BIReady, Attunity and Ajilius. Today he is the Channel Sales Director for BIReady Australia.

About BIReady

BIReady is a reseller of Business Intelligence and Data Warehouse Automation software products that automate the delivery of life-blood information to your business quickly and efficiently – delivering Business Intelligence projects in hours or days, not weeks or months. The company is headquartered in Sydney, Australia, with customers all over the world.

For more information, visit <http://www.biready.com.au>

This publication is copyright Ian Nicholson © 2013 and may not be used or copied (in whole or in part) without the author's express permission.

Building the Data-Driven Business
Architecture and Tools for Smart Operations,
Analytics and Information

Dr. Barry Devlin
Sydney
August 24-25



ajilius
Data Warehouse Automation

Build Data Warehouses FAST!
Site license - US\$10.000pa.
Book a demo today.