Data Warehousing is Mature

- Warehousing in some form has been around for 15 years +
- Many early failures have led to hard lessons learned
- Rigid, step by step siloed processes have been defined to ensure success
- Proliferation of toolsets to address the various stages
- Highly skilled practitioners are now available
So why is it still such a challenge ??
Data Warehousing Challenges

• In building a data warehouse most organizations:
  – Often view the process of creating a data warehouse as an extension of “data extract and load”
  – Tend to ignore the repeatable and iterative nature of the process of creating a data warehouse
  – Do not attempt and make data warehousing best practices repeatable
Data Warehousing Challenges

• In building a data warehouse most organizations:
  – Tend to view the data warehouse design as static
  – Forget to factor in the often exponentially rising project costs related to ongoing DW modification and changes in requirements
  – Underestimate the costs of developing and maintaining accurate project documentation
DWLC Software Introduction

• DWLC software fundamentally transforms the quality and economics of creating and maintaining a data warehouse

• DWLC solutions create automated repeatable processes that help companies
  – Create production data warehouses in a fraction of time and cost that it takes with manual approaches and traditional solutions like ETL tools
  – Encapsulate and makes addressable the process and best practices of creating and maintaining a data warehouse
The Changing Face of Data Warehousing

Home-grown solutions are brittle and expensive – one time monuments
The Changing Face of Data Warehousing

Traditional ETL-centric solutions are faster (than home-grown), but more expensive, and extend lifecycle capabilities somewhat.
The Changing Face of Data Warehousing

DWLC Software is fast, complete and inexpensive by design
What is “the Data Warehouse Lifecycle”?

Data Warehouses have traditionally been implemented using a Software Development Lifecycle (SDLC) approach.
The REAL Data Warehouse Lifecycle…

- Is a Cycle of Cycles
- Current “Siloed” approach does not allow this to happen quickly enough
Building a Data Warehouse is Not Easy…

• Data Warehousing, by necessity, uses a siloed development approach:

  Business requirements gathering - Design model - Analyse source data - Create ETL Specs - Build ETL Code - Load warehouse model - Design BI reporting layer - Build reports - Show business users...change it, start again..
Building a Data Warehouse is Not Easy…

• The traditional approach mandates that structured processes MUST be followed to ensure success

BUT this results in:-

• Issues :-
  – Slow speed to market
  – Resource-intensive
  – Difficult to change – not agile
  – Unable/slow to respond to changes in business requirements

Slow Data Warehousing Life Cycle !!!!
The new DWLC Approach to Data Warehousing

Data Warehousing is undergoing a Paradigm Shift

We need a faster, more efficient approach to address the current issues

• Data Warehousing lifecycle technologies are a new approach that treats the entire DW lifecycle as ONE process, breaking the traditional siloed development methods.

  They include:-
  – Source data analysis, data transformation/integration, DW schema design, stage tables, facts, dimensions, aggregates, views, BI cubes, full self-documentation, data lineage, impact analysis
Data Warehouse at the Speed of Business

- Business users express requirements in terms of what they currently get.
- Connect with miners, explorers and farmers so they know what is possible before they frame requirements.
- Gets users involved early and often. Iterative development means no surprises.
- This approach requires a suitable DW Lifecycle tool set
Current DW development time and cost

Traditional project timelines

- Start
- Collect and document user requirements
- Design
- Build
- Test

6 months

Prioritize & Identify Sources
The DWLC development time and cost using RED

Data warehouse lifecycle project

Prototype / iterate / design

Harden design

Start

6 months

Prioritize & Identify Sources

User Workshops

Move to Production
Traditional methods are Software Development Lifecycle-ish
Traditional practices require multiple tool change events
Result: project timeline elongation
But time is the enemy...
The integrated development framework includes the approach.

- DWLC treats data warehousing as a process – not a project.
- Which is why Bill Inmon said “SDLC approaches don’t work.”
- Live Prototyping enables rapid prototyping with real data & real users.
WhereScape RED for data warehousing

Functionally WhereScape RED integrates the following traditional development areas / tools that would normally have their own interface and metadata repository.

- **ETL Functionality**
  - Extract, Transform & Load

- **Database Modelling**
  - Design Schema & business definitions

- **Scheduler / Workflow**
  - Processing jobs & alerts

- **Metadata repository**
  - Impact analysis, promote & change mgmt

- **Documentation**
  - Support, troubleshoot, extend

**RED breaks the siloed approach**
The ETL Fixation – An old problem now!!

The other major problem is IT’s necessary fixation on the ETL part of a project.

• Traditional ETL requires that you have a detailed, specific target model design.

• This model design is a consequence of business user requirements gathering, which we already know is likely to be flawed.

• Therefore the ETL build will process will be flawed, and this is a long and complex process to build, so can be a big part of the bottleneck.

• Further, ETL only addresses data movement, it does not help in designing or managing the DW structures themselves.

• We need to focus instead on warehouse design and end user information delivery as the actual sources of business benefit.
Why do Data Warehouse Projects struggle?

1. Inaccurate business requirements - *#1 problem IDC*
2. Poor development productivity
3. Slow development cycles
4. High cost of resources
5. High TCO
6. Poor documentation – usually the last thing that is considered & never up to date.
7. Poor data quality
8. HIGH RISK

Gartner notes that **over 50% of data warehouse projects fail** or go wildly over budget – the real problem is *project approach*.
How Does DWLC Address Our Problems?

- **Inaccurate business requirements** – use a prototype and iterate approach which involves users in design phase.

- **Poor development productivity** – use a integrated development environment & lifecycle management tool.

- **High TCO** – use a framework which provides consistently implemented design, processing, and standards, reducing TCO.

- **Poor documentation** – usually the last thing that is considered & never up to date – use a toolset that does this for you

- **Poor data quality** – allow the business to decide what needs to be fixed.
So in a nutshell :-

- DWLC approach using RED is a cheaper, faster, less risky data warehousing development process.
- It is breaking the silos of data warehousing.
- It is leading the paradigm shift that is so needed in Business Intelligence.
- Companies are seeing the value of moving from the traditional SDLC to the DWLC approach for Data Warehousing.
- DWLC makes Data Warehousing truly agile.
- AND it is breaking the ETL fixation!!
- Customers are avidly adopting this approach.

This is the new wave of Data Warehousing.
The MIP Value Proposition

- Integrated DW development environment with a built in methodology that supports entire data warehouse lifecycle
- Promotes rapid prototype-iterate and deploy cycles
- Speed up development productivity - generates data warehouse objects and code
- Management & maintenance framework will reduce enhancement & support requirements
The MIP Value Proposition

- Schema and architecture agnostic
  - Warehouse, mart, master data repository
  - Star, snowflake, 3NF schema, data list etc
- We enable low risk, proven, practical approach to DW
- MIP with RED treats data warehousing as a process
- Allow for effective use of time & money
- Leverages existing, readily available SQL skills in marketplace
Questions?

Further information:
Steve Hitchman, Managing Director MIP
hitchman@mip.com.au