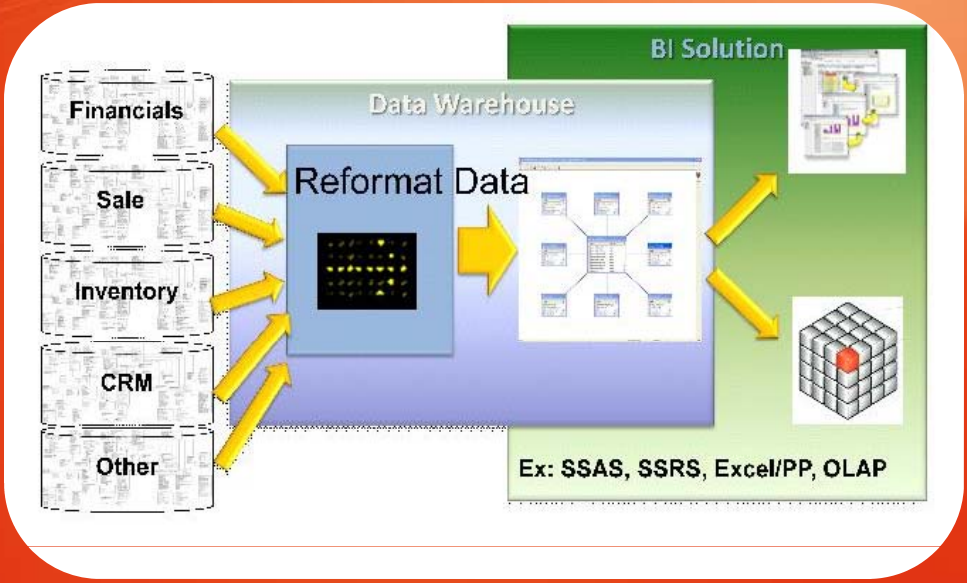
tdwi

December 8, 2009 | Colorado Chapter

Presented by Raphael Klebanov, WhereScape USA

Best Practices for Building a Data Warehouse



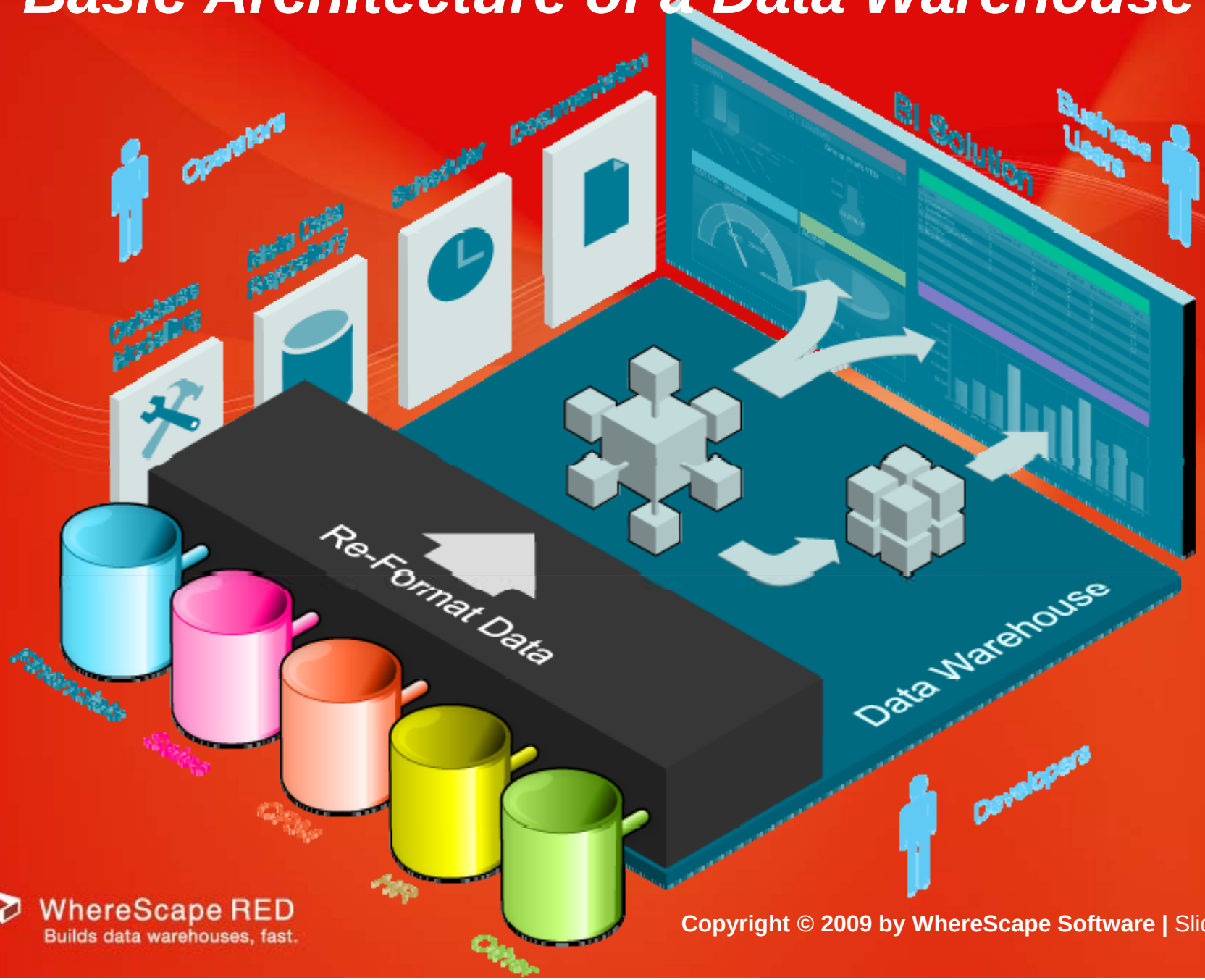
WhereScape RED
Builds data warehouses, fast.

Abstract

Key factors that influence a successful data warehouse task

- Implementing the True Development Approach
- Choosing a Rapid Development Product
- Ensuring Data Availability
- Involving Key Users throughout the entire project
- Relying on a Pragmatic Governance Framework
- Utilizing experienced Team Members
- Selecting the right Hardware, Infrastructure Technology

Basic Architecture of a Data Warehouse



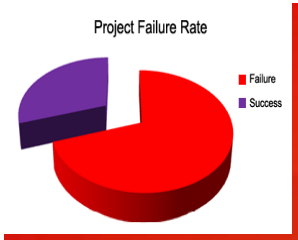
WhereScape RED
Builds data warehouses, fast.

Are you ready …

…for a intelligent decision-making process?

…for data warehouse?

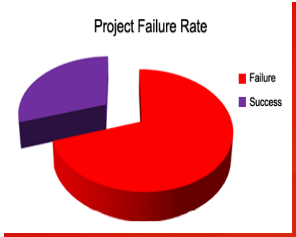
Why do Data Warehouse projects fail?



Why do Data Warehouse projects fail?

- Unreliable or unattainable user requirements

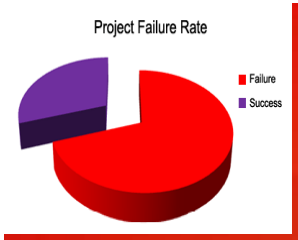
WhereScape RED
Builds data warehouses, fast.



Why do Data Warehouse projects fail?

- Unreliable or unattainable user requirements
- Quality of the data that feeds the source system

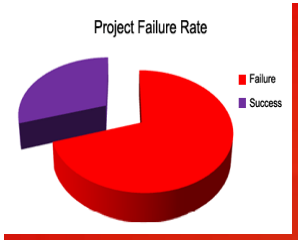
WhereScape RED
Builds data warehouses, fast.



Why do Data Warehouse projects fail?

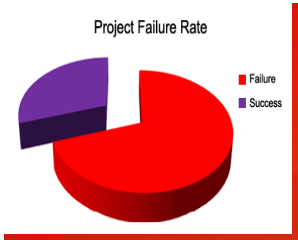
- Unreliable or unattainable user requirements
- Quality of the data that feeds the source system
- Changing source or target requirements

WhereScape RED
Builds data warehouses, fast.



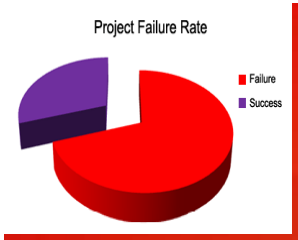
Why do Data Warehouse projects fail?

- Unreliable or unattainable user requirements
- Quality of the data that feeds the source system
- Changing source or target requirements
- Poor development productivity



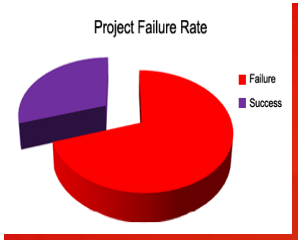
Why do Data Warehouse projects fail?

- Unreliable or unattainable user requirements
- Quality of the data that feeds the source system
- Changing source or target requirements
- Poor development productivity
- High TCO (Total Cost of Ownership)



Why do Data Warehouse projects fail?

- Unreliable or unattainable user requirements
- Quality of the data that feeds the source system
- Changing source or target requirements
- Poor development productivity
- High TCO (Total Cost of Ownership)
- Poor documentation



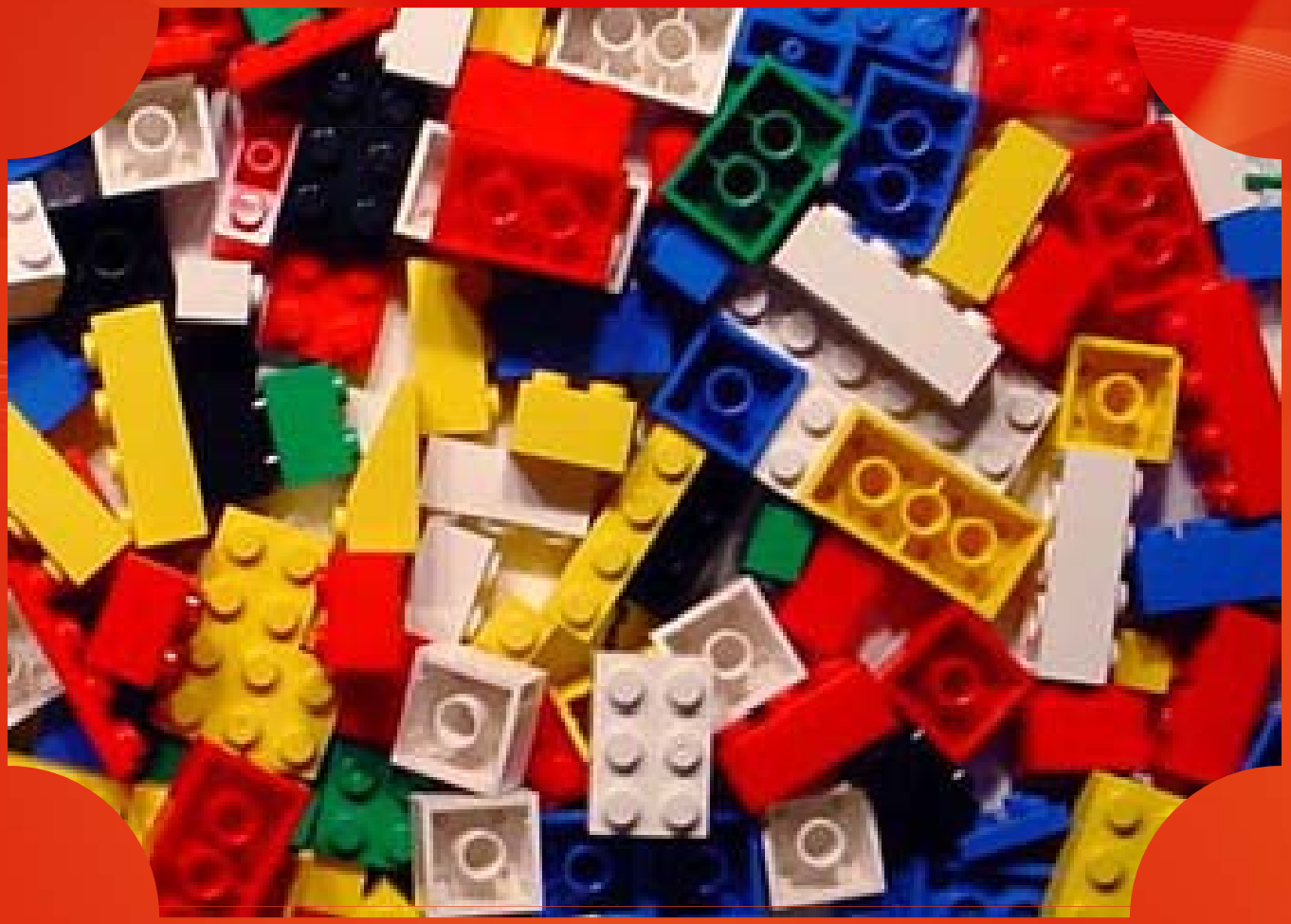
Why do Data Warehouse projects fail?

- Unreliable or unattainable user requirements
- Quality of the data that feeds the source system
- Changing source or target requirements
- Poor development productivity
- High TCO (Total Cost of Ownership)
- Poor documentation

"…over 50% of data warehouse projects fail or go wildly over budget – they blame data quality…" The real problem is project approach.

Source: Gartner. Magic Quadrant for Data Integration Tools, 2007

DW Project Components



DW Project Components

- Strong sponsorship of the DW from the business

DW Project Components

- Strong sponsorship of the DW from the business
- Divide and Conquer approach

DW Project Components

- Strong sponsorship of the DW from the business
- Divide and Conquer approach
- Iterative Development approach

DW Project Components

- Strong sponsorship of the DW from the business
- Divide and Conquer approach
- Iterative Development approach
- Productive development tools

DW Project Components

- Strong sponsorship of the DW from the business
- Divide and Conquer approach
- Iterative Development approach
- Productive development tools
- Real data to populate the prototype

DW Project Components

- Strong sponsorship of the DW from the business
- Divide and Conquer approach
- Iterative Development approach
- Productive development tools
- Real data to populate the prototype
- Access to SME during development

DW Project Components

- Strong sponsorship of the DW from the business
- Divide and Conquer approach
- Iterative Development approach
- Productive development tools
- Real data to populate the prototype
- Access to SME during development
- Compact teams

DW Project Components

- Strong sponsorship of the DW from the business
- Divide and Conquer approach
- Iterative Development approach
- Productive development tools
- Real data to populate the prototype
- Access to SME during development
- Compact teams
- Sturdy development hardware

Business Ownership

Business Ownership

- The data warehouse should be owned by the business – not IT

Business Ownership

- The data warehouse should be owned by the business – not IT
- A successful project depends upon creating a partnership with the business

Business Ownership

- The data warehouse should be owned by the business – not IT
- A successful project depends upon creating a partnership with the business
- Prioritization of project phases or agreement on a data dictionary should be agreed by the business

WhereScape RED
Builds data warehouses, fast.

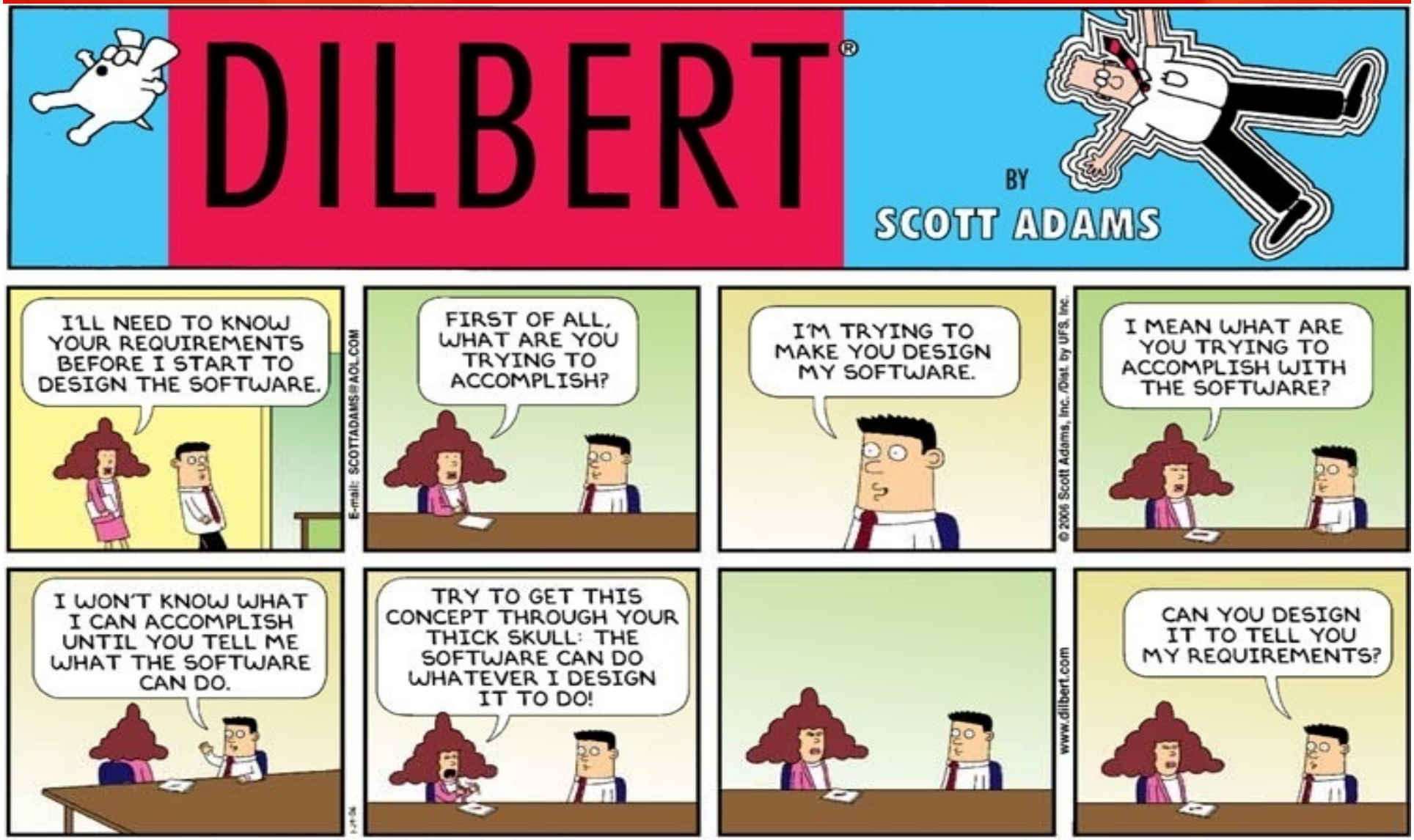
Business Ownership

- The data warehouse should be owned by the business – not IT
- A successful project depends upon creating a partnership with the business
- Prioritization of project phases or agreement on a data dictionary should be agreed by the business
- Without a strong, high level business sponsor(s) the project is likely to hit problems

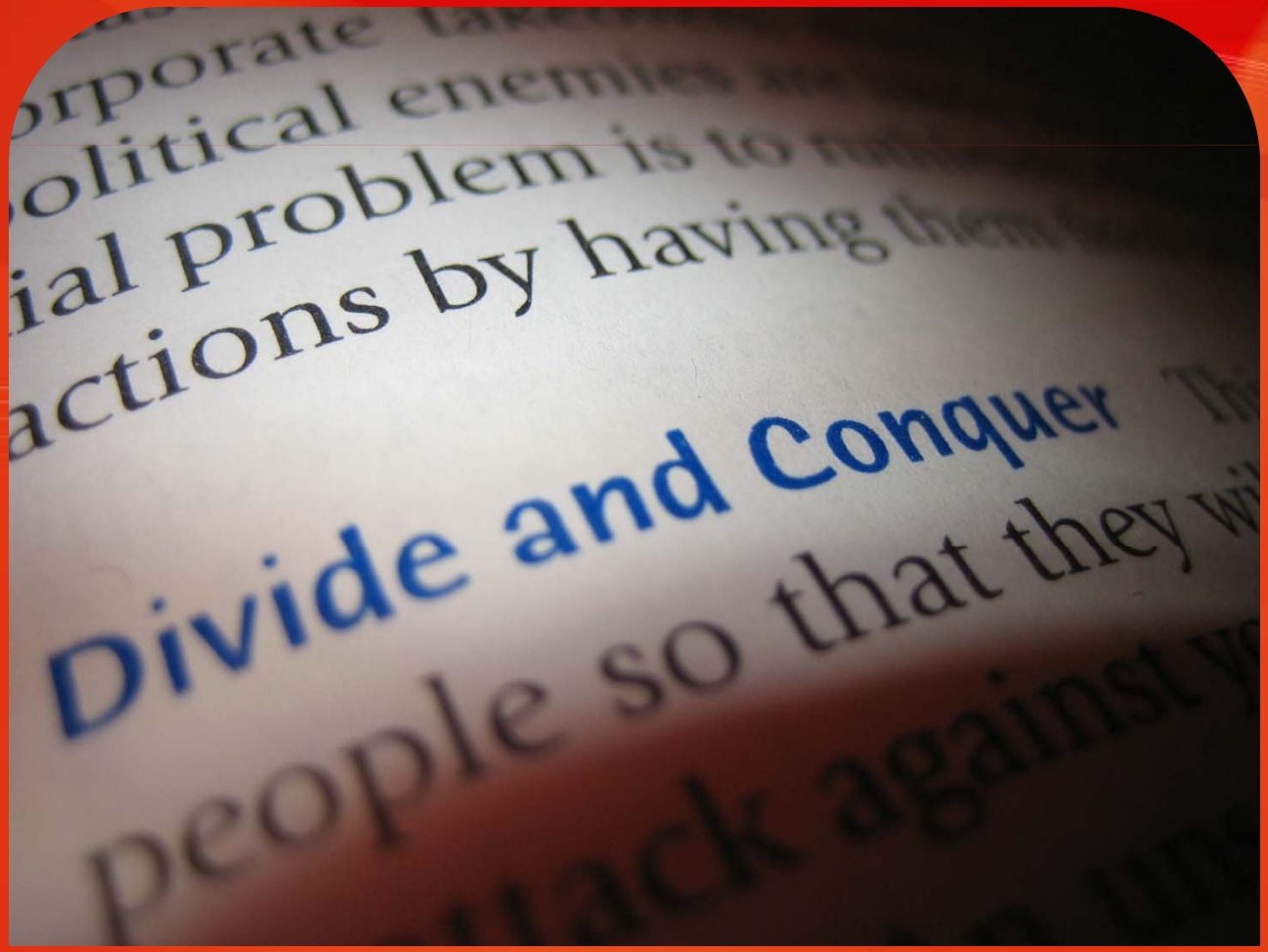
Business Ownership

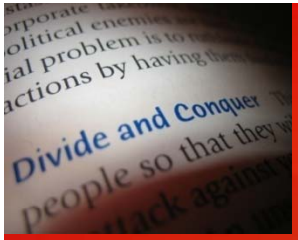
- The data warehouse should be owned by the business – not IT
- A successful project depends upon creating a partnership with the business
- prioritization of project phases or agreement on a data dictionary to should be agreed by the business
- Without a strong, high level business sponsor(s) the project is likely to hit problems
- If sponsorship is present then the data warehouse project can be broken down into a set of smaller projects

The Data Warehouse lifecycle …as we know it



Divide and Conquer





Divide and Conquer

- A ‘big bang’ approach to data warehousing has almost always ended in disaster

Divide and Conquer

- A ‘big bang’ approach to data warehousing has almost always ended in disaster
- The project phases and the order in which they are developed should be decided by the data warehouse sponsors

Divide and Conquer

- A ‘big bang’ approach to data warehousing has almost always ended in disaster
- The project phases and the order in which they are developed should be decided by the data warehouse sponsors
- Momentum is paramount for keeping the required focus

Divide and Conquer

- A ‘big bang’ approach to data warehousing has almost always ended in disaster
- The project phases and the order in which they are developed should be decided by the data warehouse sponsors
- Momentum is paramount for keeping the required focus
- Rapid prototyping and tight development cycles are vital for successful warehouse

Divide and Conquer

- A 'big bang' approach to data warehousing has almost always ended in disaster
- The project phases and the order in which they are developed should be decided by the data warehouse sponsors
- Momentum is paramount for keeping the required focus
- Rapid prototyping and tight development cycles are vital for successful warehouse
- Keep in view the bigger picture

Divide and Conquer

- A 'big bang' approach to data warehousing has almost always ended in disaster
- The project phases and the order in which they are developed should be decided by the data warehouse sponsors
- Momentum is paramount for keeping the required focus
- Rapid prototyping and tight development cycles are vital for successful warehouse
- Keep in view the bigger picture
- Use smaller phases to fund the project adequately

The True Project Approach

The True Project Approach

- Getting the business reps to use working prototypes to share an understanding of the scope

The True Project Approach

- Getting the business reps to use working prototypes to share an understanding of the scope
- Collect detailed user requirements is exactly the wrong start to a data warehouse project

The True Project Approach

- Getting the business reps to use working prototypes to share an understanding of the scope
- Collect detailed user requirements is exactly the wrong start to a data warehouse project
- Showing business users the data and relationships that are available to them in a working, populated prototype

The True Project Approach

- Getting the business reps to use working prototypes to share an understanding of the scope
- Collect detailed user requirements is exactly the wrong start to a data warehouse project
- Showing business users the data and relationships that are available to them in a working, populated prototype
- A better place to start is to collect KPIs and source system technical documentation

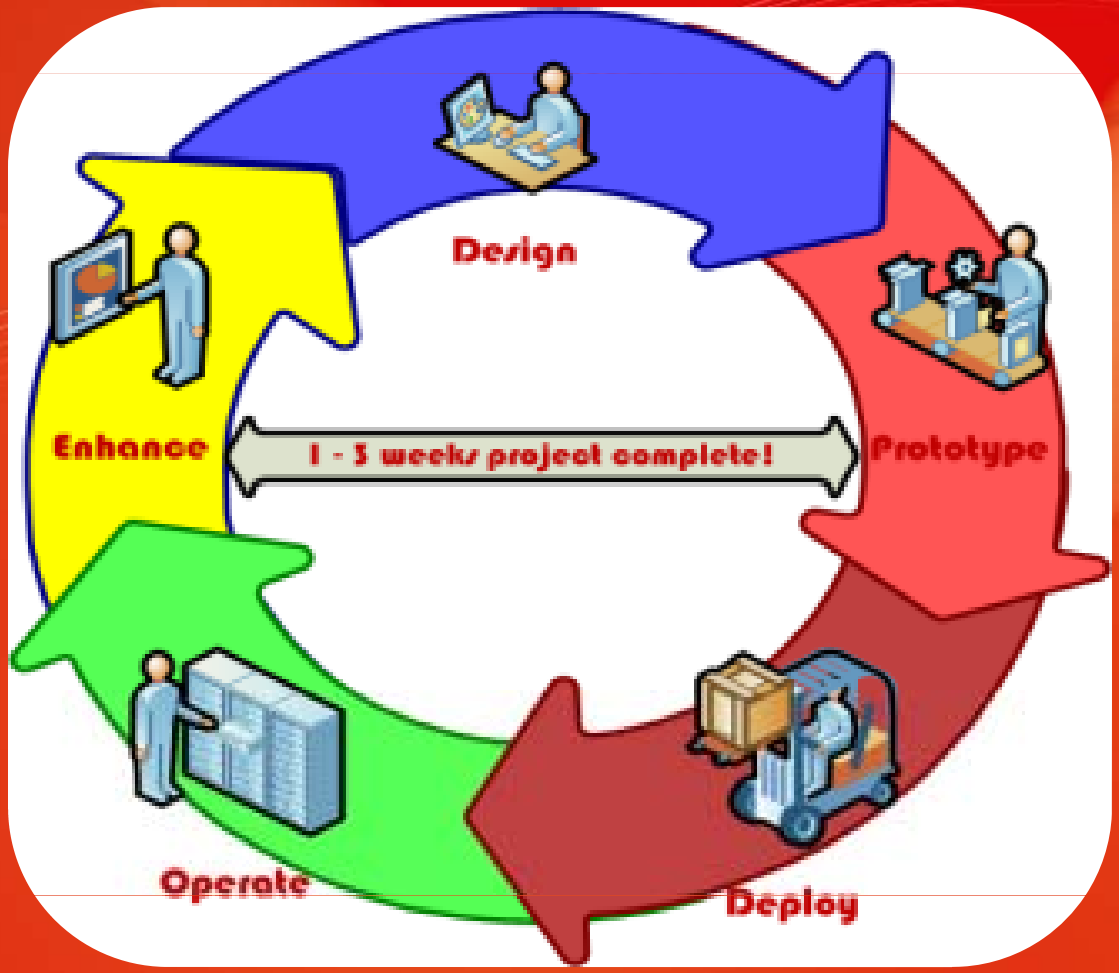
The True Project Approach

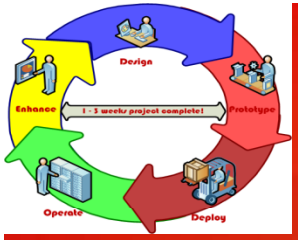
- Getting the business reps to use working prototypes to share an understanding of the scope
- Collect detailed user requirements is exactly the wrong start to a data warehouse project
- Showing business users the data and relationships that are available to them in a working, populated prototype
- A better place to start is to collect KPIs and source system technical documentation
- OLAP technology and user workshops are key tools in allowing the business to get their hands on the data

The True Project Approach

- Getting the business reps to use working prototypes to share an understanding of the scope
- Collect detailed user requirements is exactly the wrong start to a data warehouse project
- Showing business users the data and relationships that are available to them in a working, populated prototype
- A better place to start is to collect KPIs and source system technical documentation
- OLAP technology and user workshops are key tools in allowing the business to get their hands on the data
- Data quality should not be addressed in the DW; problem should be fixed on the source system

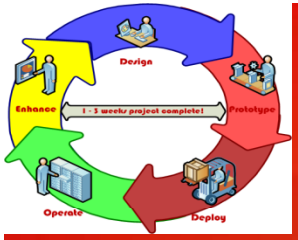
The New Breed of Tool – The DWLC Tool





The Features of a DWLC Tool

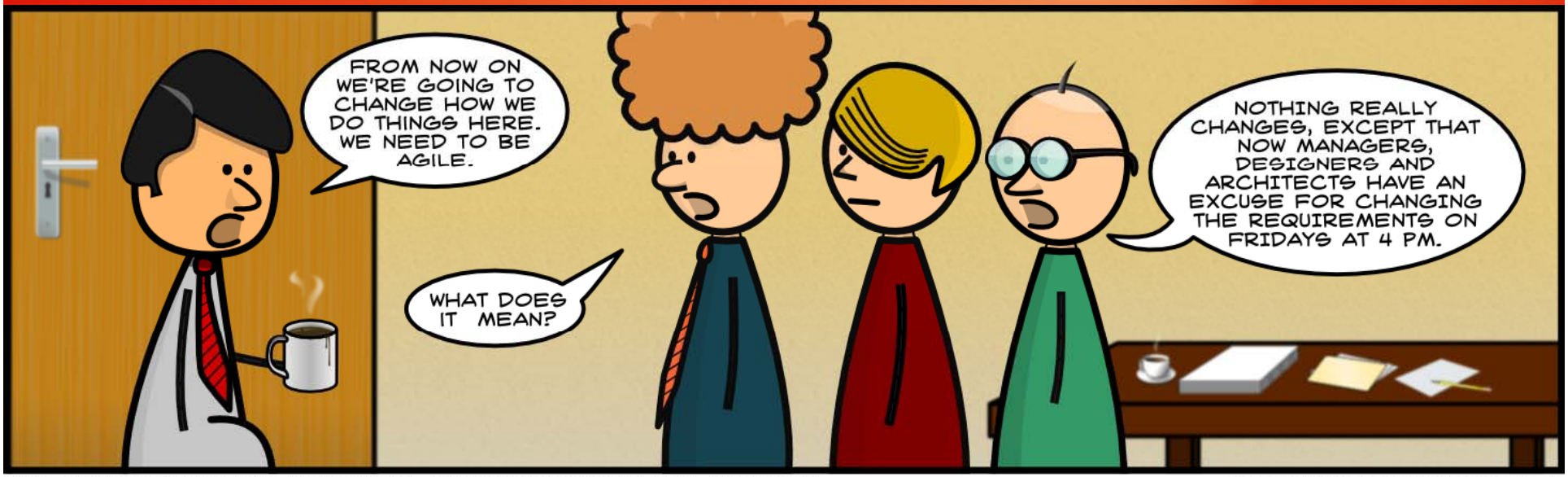
- Single Development Interface
- Documentation
- Automated Table Generation
- Automated Code Generation
- Metadata Migration
- Version Control
- Object Checkout
- Leverage Existing Core Skills
- Consistent Framework
- Extensibility

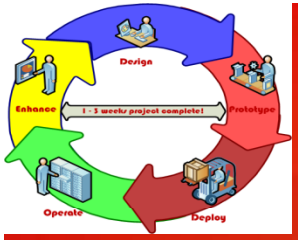


The Features of a DWLC Tool

A DWLC Tool would save a huge amount of development time and effort, and would enable the approach required to deliver a successful outcome

The DWLC methodology is a child concept for Agile Development Methodology also known as Systems Development Life Cycle (SDLC)





Traditional Approach vs. DWLC Approach

The Traditional Approach

Try to get it right first time: The SDLC approach

120 day cycle

But:

- Tools and operators in silos – inflexible
- Hard to engage business users, no shared understanding
- Locked in requirements that can't be met

Redevelopment

Risky, expensive, never OTTB & never finished

No documentation, so hard to support

The Rapid Development Approach

Prototype and iterate to prove a design with users

5 day cycle

Supported by:

- Integrated toolset and metadata repository – maximum flexibility
- Continuous business engagement, shared understanding
- No ambiguity or disagreement about scope

Successful phase completion

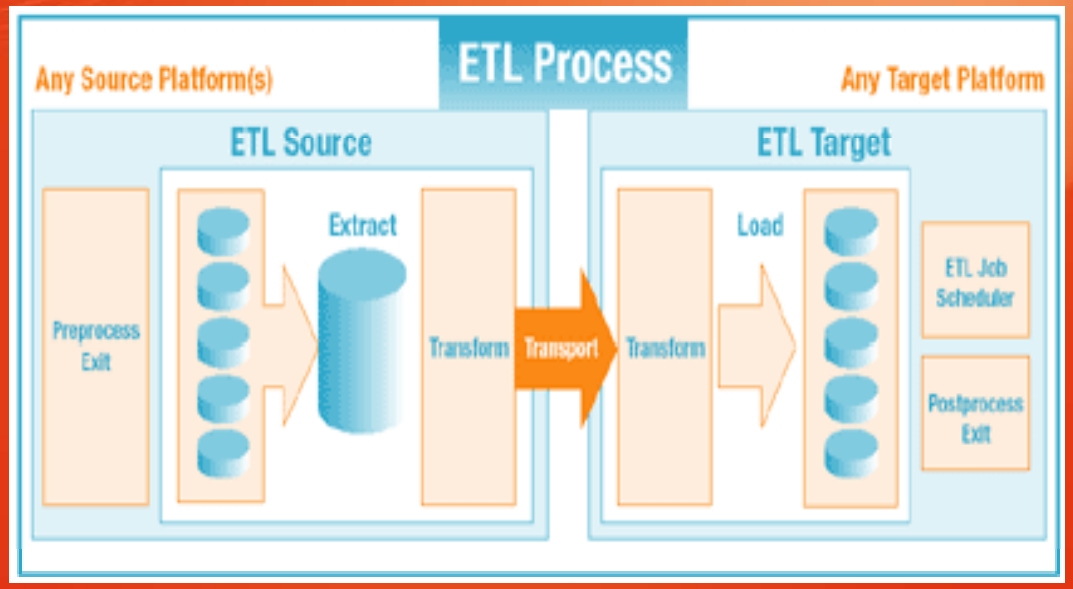
Complete, OTTB, user expectations exceeded

Documented solution that is easy to support

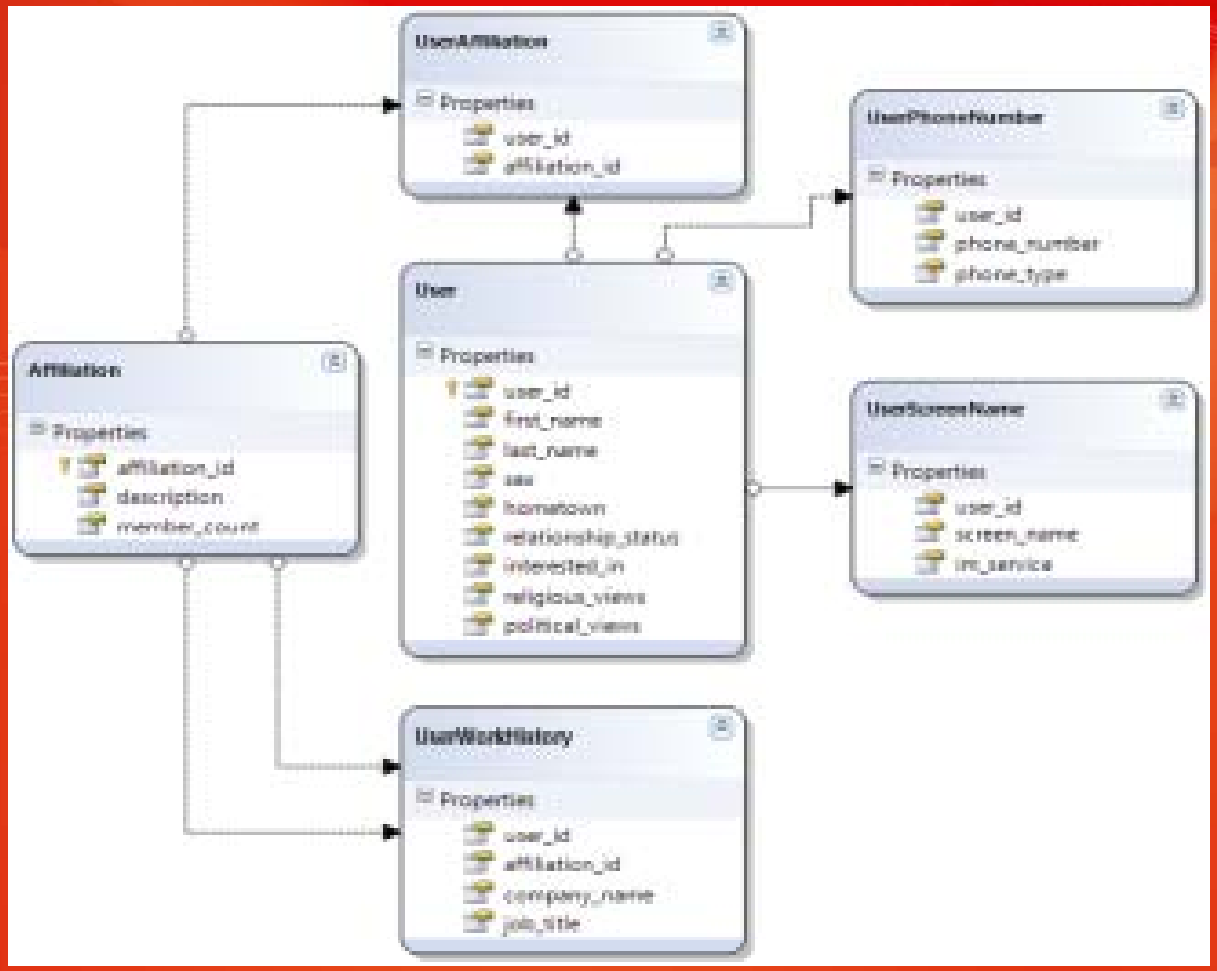
WhereScape RED
Builds data warehouses, fast.

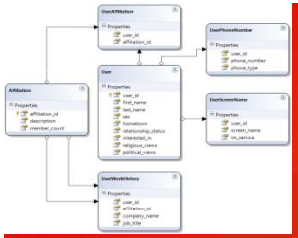
Rapid Development Product and ETL

- Scrutinize Extract/Transform and Load (ETL) tools when considering building a DW. ETL tools do not provide the ability to build a working prototype and work in short development cycles



Ensuring Data Availability





Ensuring Data Availability

- The lack of good quality live data will have a major impact on the success of Iterative project approach

- The DW's capacity to answer BI requirements is unworkable, without sufficient data to populate the DW

- If a new source system is integrated into the data warehouse, the "real" data is quite essential

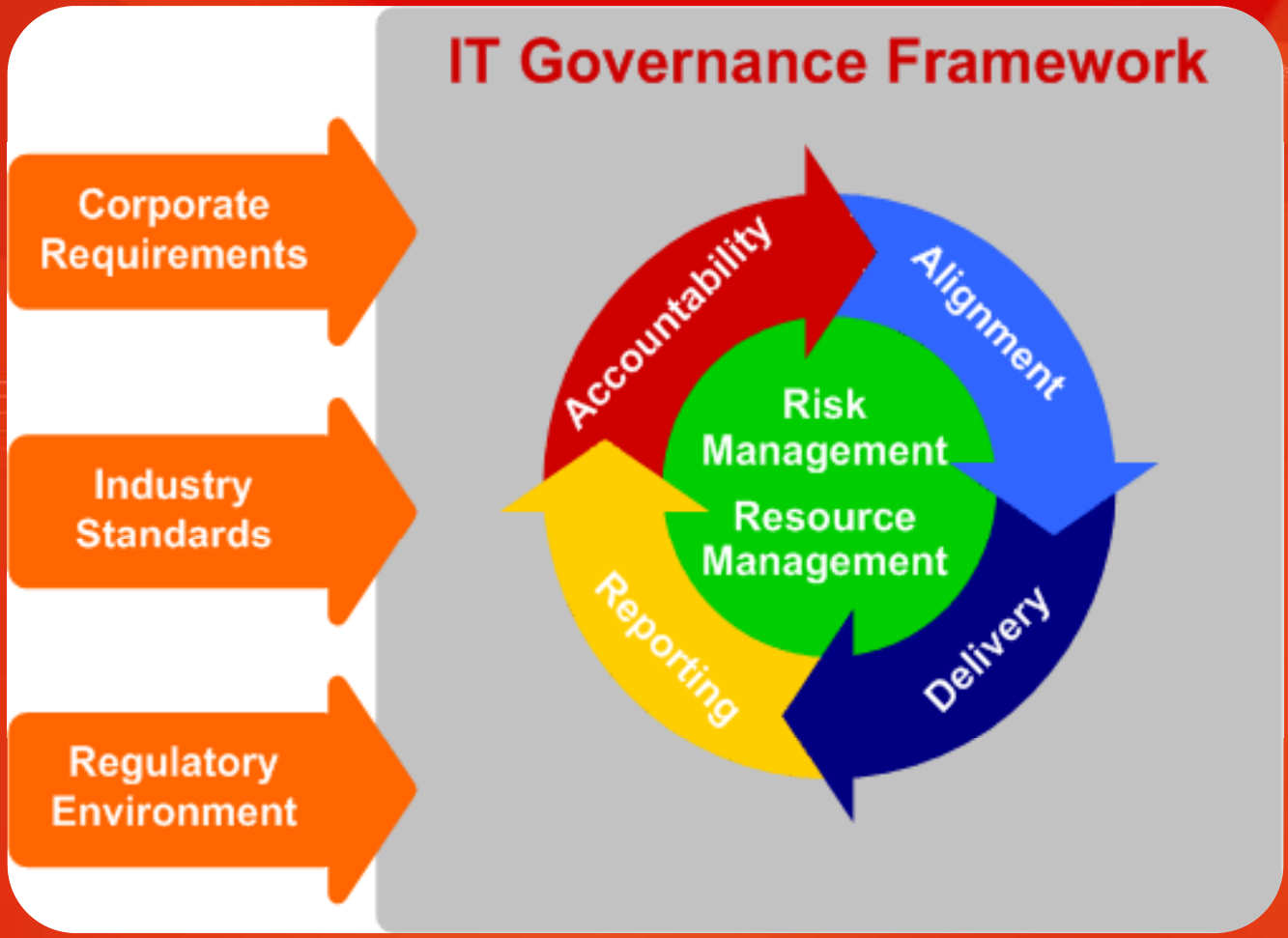
- If no "real" data for new source is available, then the significant rework will be required once the source is up and running

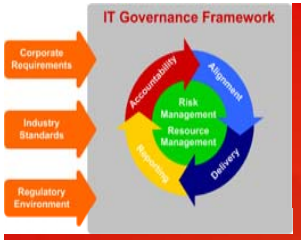
Involving the Business

Involving the Business

- Representatives from the Business provide the partnership with the DW development team
- These reps need to be able to articulate the needs of the business to the dev. team
- These reps have to trust the business department behind them when it comes to making any decisions
- The partnership during the iterative project approach provides a reliable, successful outcome
- The main forum for developers to show a working prototype and get user feedback is user workshop
- The business Involvement for the duration of the DW development will reduce the QA overheads

Project governance

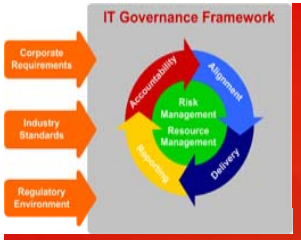




Pragmatic Governance Framework

Governance of the data warehouse project should operate at two levels:

- an enterprise level and
- a project level



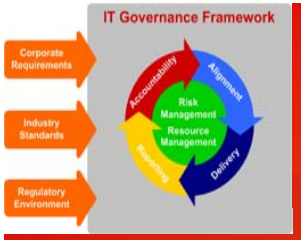
Pragmatic Governance Framework

Governance of the data warehouse project should operate at two levels:

- an enterprise level and
- a project level

Business Requirements

Technical Constraints



Pragmatic Governance Framework

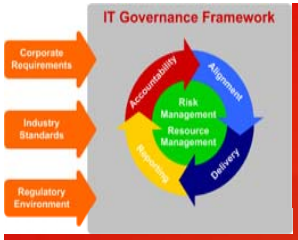
Governance of the data warehouse project should operate at two levels:

- an enterprise level and
- a project level

Business Requirements

Shared understanding, Prototype and Iterate, Best possible outcome

Technical Constraints



The DW needs to be owned by the business

Sponsorship is sourced from a highly-placed executive

The steering committee provides:

+ Vision

+ Visibility

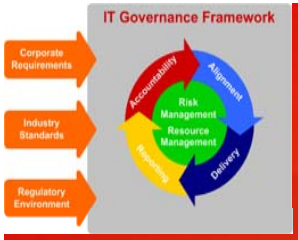
+ Priorities

+ Scope

+ Focus

+ Terminology

WhereScape RED
Builds data warehouses, fast.



Project governance

At a minimum project governance should include:

- A project plan, detailing (high level) scope and timelines
- Regular status meetings to share information
- Change request process documentation
- Standards and procedures for building a consistent DW
- Version control and backup procedures
- Ownership of specific environments and project roles

Utilizing Experienced Team Members

Utilizing Experienced Team Members

- Productivity within a data warehouse implementation is dependent on having experienced team members – both on business side and also on the technical side
- Experienced Subject Matter Experts (SME) provide a thorough understanding of the business and its needs
- Experienced data warehouse developers can take those requirements and turn them into a functioning data warehouse in a rapid timeframe

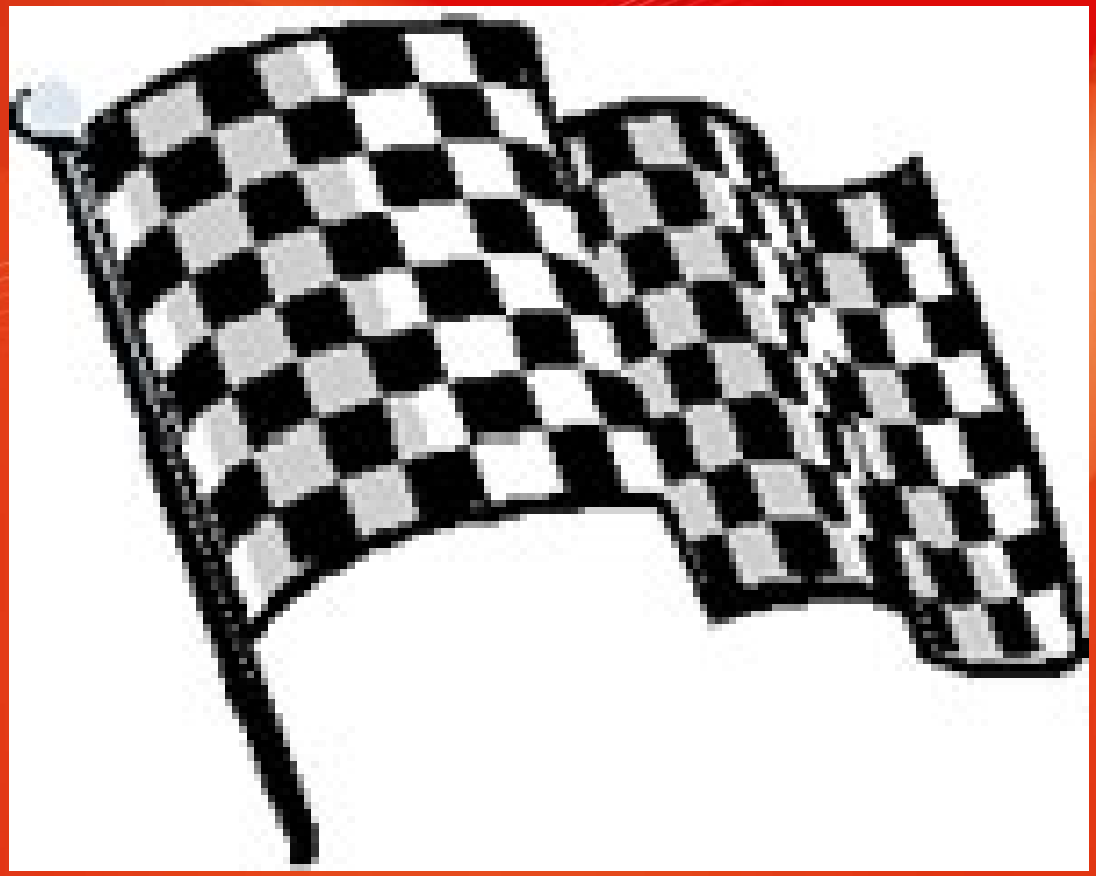
Selecting the Right Infrastructure

Selecting the Right Infrastructure

- Sufficient hardware and technology infrastructure during development
- Lower productivity can translate into slower development cycles and iterations, which stands the risk of losing project momentum
- Trade-off between having adequately sized hardware and the cost associated with purchasing that hardware
- One way to mitigate undersized hardware is to use smaller subsets of data during the prototyping phase

Conclusion

Treat the Warehousing as a process, not a project



Conclusion

This means:

- focusing on iterative releases and rollouts that follow in quick succession
- keeping the warehouse in line with the ever changing needs of the business, instead of treating it as a one-time project
- In order to achieve this, a change in the development approach and tools utilized for building the data warehouse must be adopted

Conclusion

The key factors to creating a successful data warehouse are:

- ✓ Implementing the True Development Approach
- ✓ Choosing a Rapid Development Product
- ✓ Ensuring Data Availability
- ✓ Involving Key Users throughout the whole project
- ✓ Relying on a Pragmatic Governance Framework
- ✓ Utilizing experienced Team Members
- ✓ Selecting the right hardware and other related Infrastructure Technology

My personal information:

Raphael Klebanov, **Analyst at WhereScape**

Office Phone: **303.968.0703**

Email address: **rklebanov@wherescape.com**

Public Profile:

http://www.linkedin.com/in/raphaelklebanov