Agile Data Warehousing

Christina Knotts
Associate Consultant
Eli Lilly & Company

Answers That Matter.

Overview

- Defining Agile Data Warehousing
- Reasons for Agile Data Warehousing
- Walk-Thru with Case Study
- Key Learnings
- Additional Information

Manifesto for Agile Software Development

We are uncovering better ways of developing software by doing it and helping others do it.

Through this work we have come to value:

Individuals and interactions over processes and tools
Working software over comprehensive documentation
Customer collaboration over contract negotiation
Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

Source: Beedle et al. (2001). *Manifesto for Agile Software Development.* Retrieved from http://agilemanifesto.org/

Characteristics of an Agile DW Project

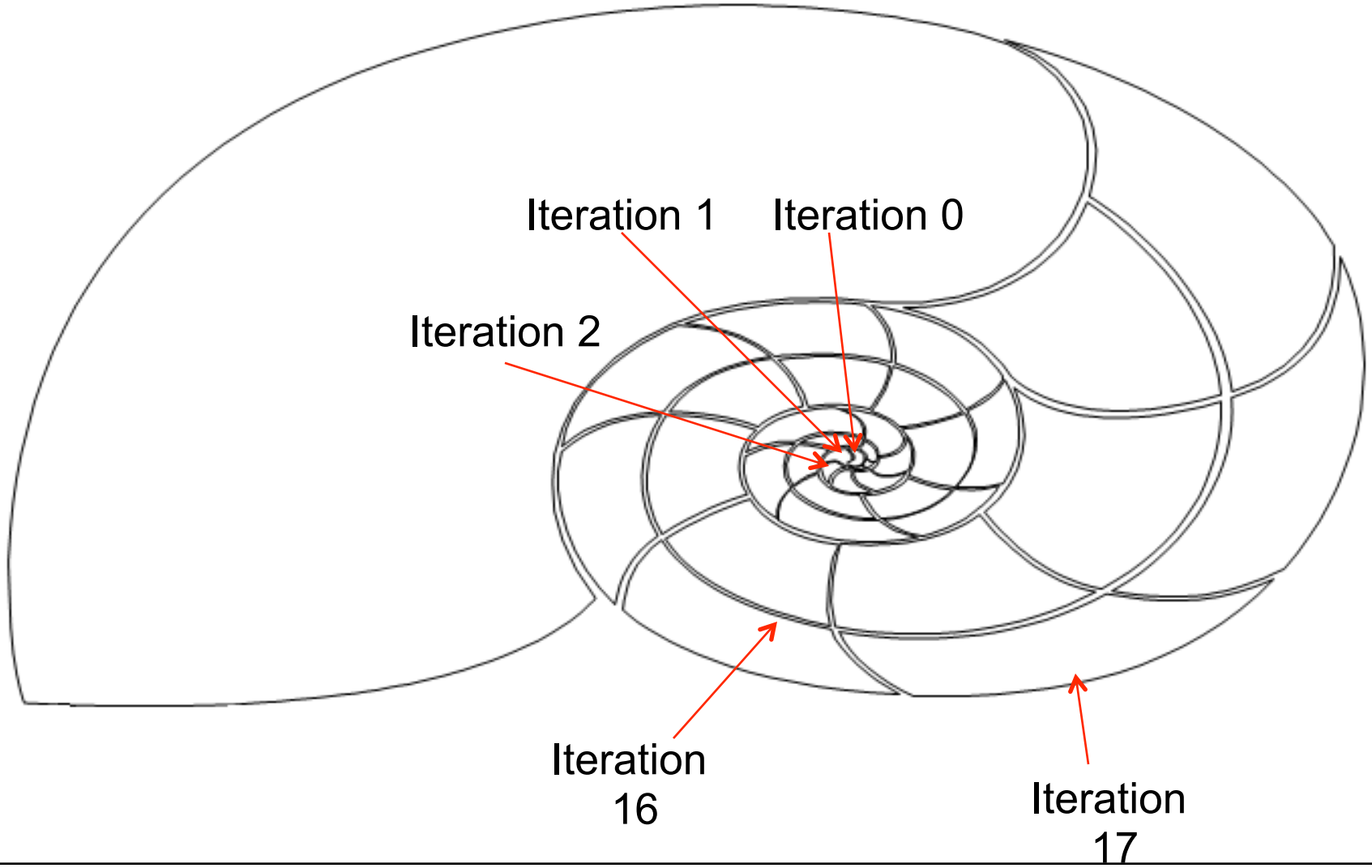
- Incremental, Iterative, Collaborative Development
- Cross-Functional Teams with IT and Business
- Regularly scheduled stand-up meetings
- Deliverables modular and additive towards program
- Specification documentation lags behind development

Source: Kobielus, James. (2010). *Agile Data Warehousing: Do You Scrum?.* Retrieved from http://intelligent-enterprise.informationweek.com/blog/archives/2010/07/agile_data_ware.html

Why Use Agile for DW Projects?

- Increased level of stakeholder ownership and involvement
- Demonstrate value quickly
- Proof of concept activities
- Increased adaptability to changing requirements

Iterative Natural Development



Establishing an Agile DW Program – Iteration 0

- Determine high-level program scope and sponsors
- Research industry best practices and patterns
- Construct teams
- High-level technology decisions
- Establish code repository and deployment environments
- Determine iteration length
- Choose applicable Agile Methodology

Case Study: Establishing Orion

- Chose Agile Scrum methodology
  - 4 week sprints
  - Product Owners, Scrum Masters, Scrums, Product Backlog
- Built on Microsoft SQL Server stack
  - SQL Server for DB, SSIS for ETL, SSAS for cubes, SSRS for reports
- Uses Kimball approach to data warehousing
- Modeled after Microsoft's Project REAL
- Multiple function level teams

Building a Product Backlog

Goal: Obtain high-level requirements for current and future iterations

Case Study: Interview Process

- Conducted extensive Kimball requirements interviews
  - Interview team consisted of 2 interviewers and 1 scribe
  - Maximum of 3 interviewees at a time
  - Broad sampling of interviewees from across groups
  - Sampling of interviewees to include all levels of users, including executives
  - Request sample reports/visualizations in advance
  - Open-ended, non-scripted questions
  - How does success for the program look for each user?

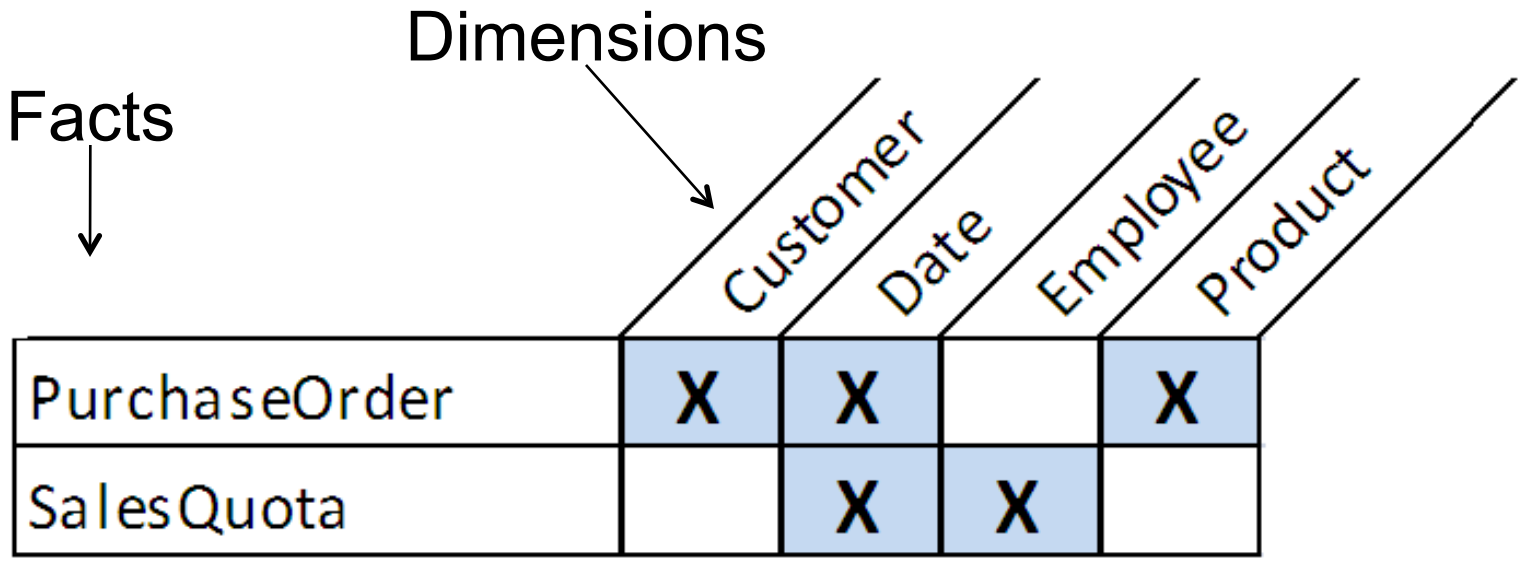
Lessons Learned: Interview Process

- Do not schedule back to back interviews
- Schedule no more than 4 interviews per day
- Allow time immediately following interview sessions to regroup and review outcomes and themes
- Don’t schedule users with their line management in same session
- Analyze outcomes and report examples for backlog items

Iteration 1: Rubber Meets the Road

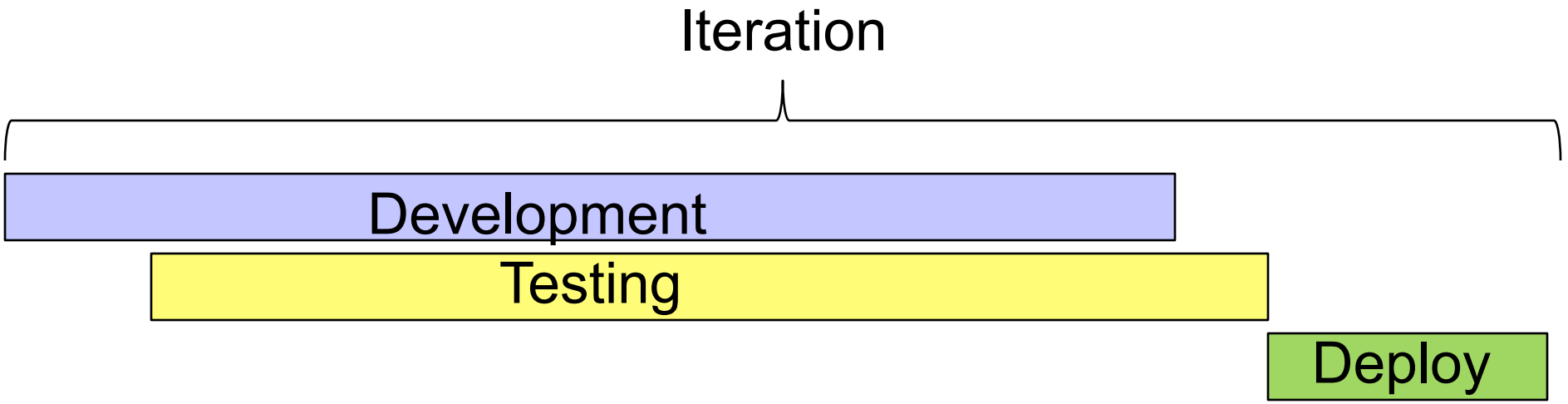
- Create and analyze Product Backlog
- Estimate size and complexity of Product Backlog Items
- Focus on data domains
- Assign work for 60-80% allocation per sprint
- Start data warehouse models
  - Bus matrix
  - Dimensional models

Bus Matrix

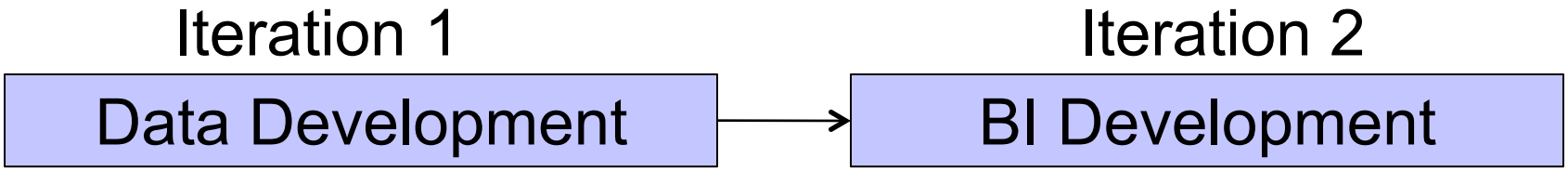


- Defines grain of fact tables
- Ease discussions with customers

Sprint Breakdown



Handling Development Dependencies



Recommendation: Keep data warehouse and ETL development at least a sprint in advance of OLAP cube or report development

Lessons Learned

- Negotiate scope post-sprint planning
- Adjust deploy timelines as needed
- Involve support staff early
  - Ease hand-off at end of development
  - Keep developers focused on development
- Adapt, Adapt, Adapt
  - Be willing to learn new technologies and methodologies
  - Step outside of your traditional, rigid role definition

For Further Information

- Microsoft Project REAL
  - Microsoft stack BI best practices
  - http://www.microsoft.com/sqlserver/2005/en/us/project-real.aspx
- Eli Lilly & Company's Orion System Overview
  - http://www.informationweek.com/news/business_intelligence/perf_management/showArticle.jhtml?articleID=210600831
- Agile Manifesto
  - http://agilemanifesto.org/
- Agile Data
  - http://www.agiledata.org/