

O'REILLY®

Sponsored by  
**TERADATA**

# How To Be Agile with Your Big Data



Mike Barlow

---

# How to Be Agile with Your Big Data

*Mike Barlow*

## **How to Be Agile with Your Big Data**

by Mike Barlow

Copyright © 2014 O'Reilly Media, Inc. All rights reserved.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

This article originally appeared on the *O'Reilly Data blog* and is part of a collaboration between O'Reilly and Teradata. See O'Reilly's *statement of editorial independence*.

May 2014:                      First Edition

### **Revision History for the First Edition:**

2014-05-29:    First release

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. *How to Be Agile with Your Big Data* and related trade dress are trademarks of O'Reilly Media, Inc.

While the publisher and the author have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the author disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

---

# Agile, Big Data, and the Enterprise Data Warehouse

Data analysis, like other pursuits, is a balancing act. The rise of big data ratchets up the pressure on the traditional enterprise data warehouse (EDW) and associated software tools to handle rapidly evolving sets of new demands posed by the business. Companies want their EDW systems to be more flexible and more user-friendly—without sacrificing processing speeds, data integrity, or overall reliability.

“The more data you give the business, the more questions they will ask,” says **José Carlos Eiras**, who has served as CIO at Kraft Foods, Philip Morris, General Motors, and DHL. “When you have big data, you have lots of different questions, and suddenly you need an enterprise data warehouse that is very flexible.”

EDWs are remarkably powerful, but it takes considerable expertise and creativity to modify them on the fly. Adding new capabilities to the EDW generally requires significant investments of time and money. You can develop your own tools internally or purchase them from a vendor, but either way, it’s a hard slog.

“You wind up with an endless proliferation of tools, each slightly different, designed to handle a specific type of query,” says Eiras. “Every time you have a different question, you have to use a different tool. It isn’t a good situation.”

In today’s ultra-competitive markets, business operates far too quickly for traditional software development timetables, and most companies are rightfully wary of relying on outside vendors to solve ongoing business challenges. Luckily, agile and other “lean” software develop-

ment methods offer repeatable processes for harnessing creativity and expertise, at a pace that's swift enough to keep the business happy.

## Achieving Concrete Results Faster

Unlike traditional **ITIL**-based software development lifecycle (SDLC) or “**waterfall**” methods, Agile does not begin with a finely detailed compendium of abstract requirements. Instead of striving for perfection, Agile aims for the minimum viable product (MVP), which means that Agile produces concrete results faster than traditional development methods.

The difference between the “abstractedness” of traditional methods and the “concreteness” of Agile might seem like a matter of semantics, but it's highly relevant in a world of hyper-turbulent markets and fickle customers. Abstractions are tolerable when you have plenty of time and money, but when time and money are tight, you need concrete results in a hurry. The whole point of Agile is finding out quickly whether your idea works or not. Agile embodies the concept of “failing fast,” a phrase that sums up the ethos of modern innovation.

But here's the rub: adopting Agile methodologies is not the same as being agile. Agile is both a process and a mindset. It embraces both discipline and flexibility. Agile is *not* some kind of free-form anarchy; it's a structured way of producing usable software without having a pre-written script, like improvisational theatre.

The first rule of improv is always to say, “Yes, and...” Everything an actor does on stage is complementary; nothing is exclusionary. New information emerges unexpectedly, and continual change is a given, but established premises remain in place.

## Cultural Challenges

“The biggest challenge is cultural,” says **Oliver Ratzesberger**, Senior Vice President/Software at **Teradata Labs**. “Executives tell me they want agility, and then in the next sentence they ask me for a project plan, a roadmap, and a product release date. They switch back to waterfall without even realizing it.”

For modern software executives like Ratzesberger, explaining the difference between Agile and “agility” is not an abstract dilemma. “Agility, especially in the context of big data, takes a lot of effort and really hard

work. Creating an Agile environment doesn't mean turning off governance, eliminating documentation, and giving developers a sandbox to work in. That's not agility—that's the Wild West," says Ratzesberger.

The problem with taking a Wild West approach to software development is that, "you end up with results that are not reproducible, which very quickly erodes confidence in your ability to handle the data," he says. If the business executives who depend on the EDW don't trust you with their data, they will look elsewhere for answers to their questions.

"Without methodology, Agile quickly runs away on you," says Ratzesberger. "You need to train people and you need to change the culture of your organization. If you're applying Agile to big data, you need to work on the Agile piece first, and then overlay the big data on top of it. You can't start with big data and then try to make it Agile. Then it becomes the Wild West and you're in trouble."

## Train the Business Too

Explaining Agile to business users is also important, says Ratzesberger. "You need to train the business. If you only train the IT people, the people in the business units won't understand what you're doing. They will assume they need to submit a requirements document, which they dread, and then there will be no agility between IT and the business." He suggests co-locating technology and business people to improve communication and collaboration during Agile projects.

Agile is not necessarily the answer to every software development challenge, says Ratzesberger. "There are certain tasks where accuracy is paramount. In production, for example, you're not going to upgrade a critical process using Agile. If it's a customer-facing process that the company depends on for revenue, then Agile might not be the best path."

Agile is optimally suited for situations in which the ability to generate lightning-fast results can produce genuine competitive advantages. Experimenting with subsets of customer data, testing new product categories, evaluating the usability of a web page design, gauging the appeal of a smart phone app—those are the best scenarios for getting the most value from Agile.

## Complementary Technologies

**Jim Tosone**, former Director, Team Lead of the Healthcare Informatics Group at Pfizer Pharmaceuticals, uses the principles and process of improvisation in his leadership development practice to help clients improve business performance.

Tosone sees Agile, big data, and the EDW as natural partners. “Big data technology such as Hadoop, which is based on open-source code, is inherently Agile,” he says.

“Now that we’re seeing more SQL-on-Hadoop integrations, we have a global community of open-source developers working to solve EDW challenges. That phenomenon is likely to accelerate,” says Tosone. “The relationships between Agile, big data, and EDW are complementary rather than antagonistic.”

## Where to Begin

Companies looking for ways to bring Agile into their EDW operations should begin with simple steps. Tosone recommends outlining key use cases (e.g., hypothesis generation, hypothesis testing, business plan execution) and forming developer teams with diverse backgrounds. Diversity will assure that a broad range of possible solutions are considered.

Then map the potential solutions to each use case. Look for solutions that will enable users to run queries and analyses across multiple platforms. Favor solutions that can be modified, iterated, and replaced easily.

Tosone agrees with Ratzesberger that changing the culture is critical to success. “You need to educate and train people, and help them feel comfortable working across multiple functions and disciplines,” he says. “True agility requires a mindset that perceives change as a positive instead of a negative. You need a bias toward exploration and patience to resist the desire for closure.”

## Hybrid Environments

It seems clear that Agile offers a reasonably quick and cost-effective way to bring flexibility to the EDW and to integrate newer open-source technologies such as Hadoop with existing database systems. Flexi-

bility translates into greater speed and cost savings; integration promises a significantly wider range of analytic capabilities, which creates more opportunities for the business to pursue.

As we move forward into an era of bigger, faster, and more effective data analytics, it seems logical to assume that database systems architectures will include both traditional and open-source components. In hybrid computing environments, the real challenge is optimizing the relationships between all of the various people, processes, and technologies required to get the job done quickly and efficiently.

Instead of picking fights over platforms, we should search for win-win scenarios in which we use available resources to help the business gain advantages in competitive markets. Or as my friends in improvisational theater would say: “Yes, and...”



## About the Author

---

**Mike Barlow** is an award-winning journalist, author, and communications-strategy consultant. Since launching his own firm, Cumulus Partners, he has represented major organizations in numerous industries.

Mike is coauthor of *The Executive's Guide to Enterprise Social Media Strategy* (Wiley 2011) and *Partnering with the CIO: The Future of IT Sales Seen Through the Eyes of Key Decision Makers* (Wiley 2007). Mike has also written many articles, reports, and white papers on marketing strategy, marketing automation, customer intelligence, business performance management, collaborative social networking, cloud computing, and big data analytics.

Over the course of a long career, Mike was a reporter and editor at several respected suburban daily newspapers, including *The Journal News* and the *Stamford Advocate*. His feature stories and columns appeared regularly in *The Los Angeles Times*, *Chicago Tribune*, *Miami Herald*, *Newsday*, and other major US dailies.