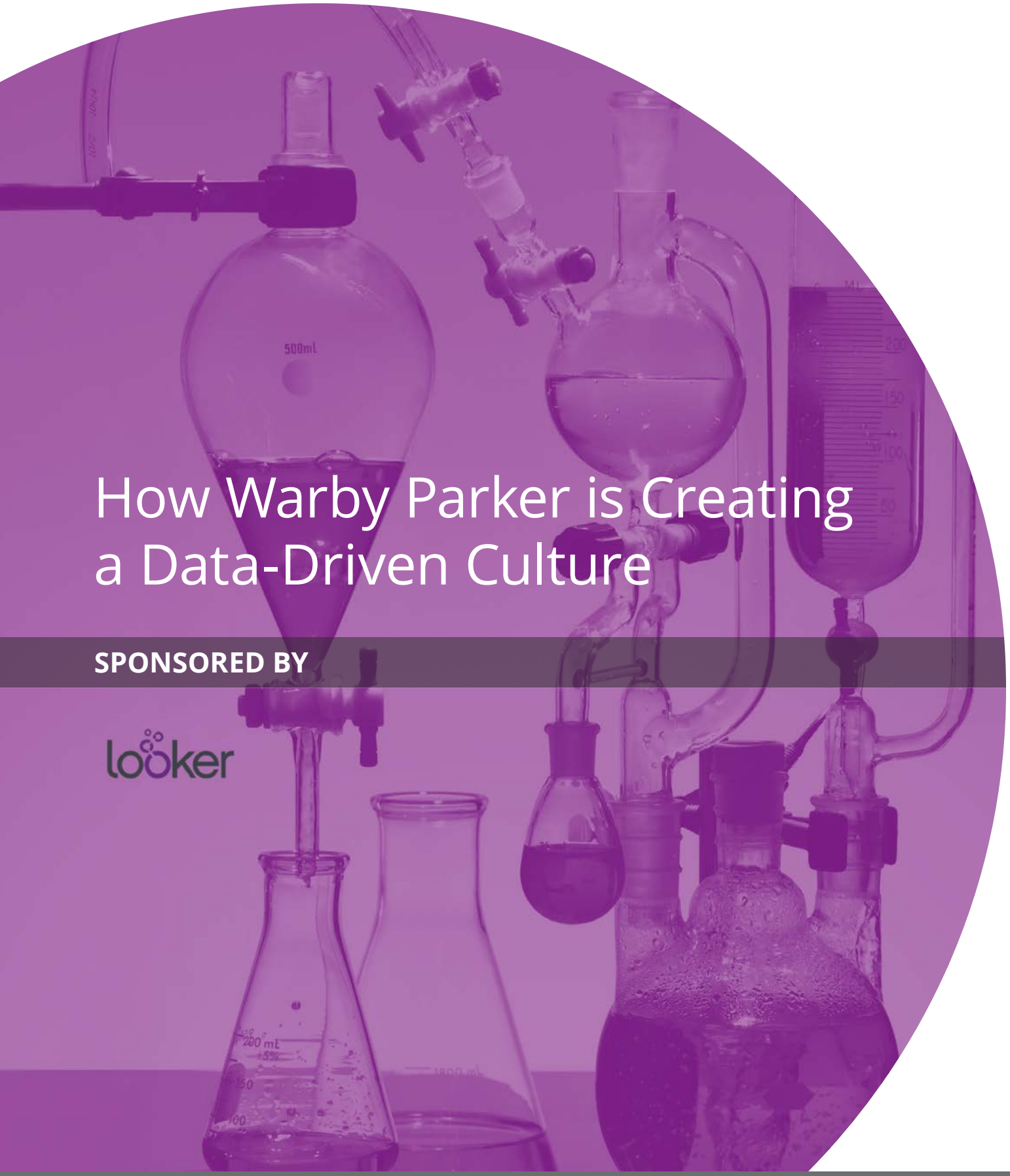
CITO Research

Advancing the craft of technology leadership



How Warby Parker is Creating a Data-Driven Culture

SPONSORED BY

looker

CONTENTS

Introduction 1

Escape from ERP 2

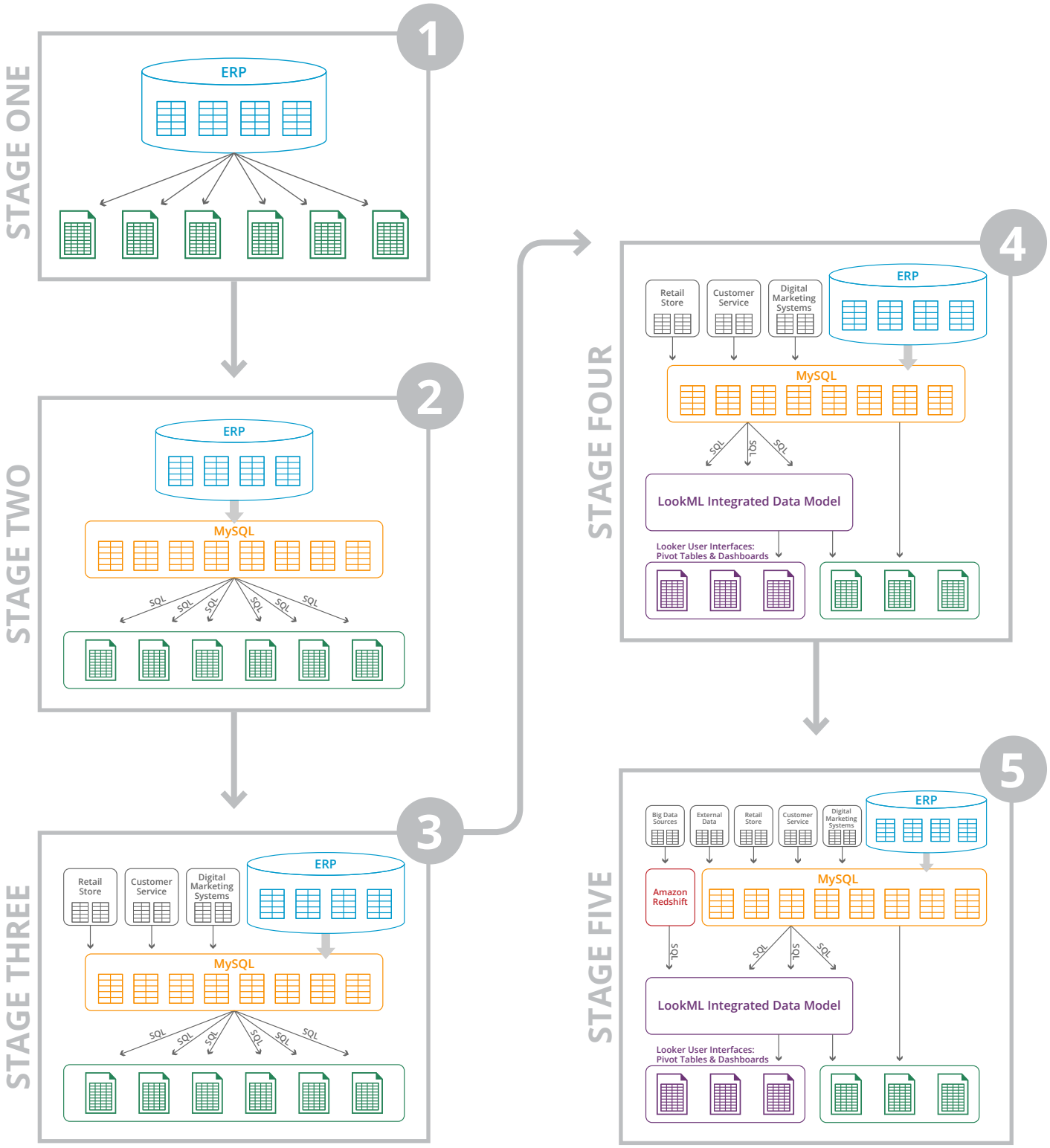
Expand Automation and the Scope of the Data Warehouse 4

Create a Unified Data Model of the Business 5

Expand Data Use Throughout the Company 8

The Future of Warby Parker's Data-Driven Culture 10

A Snapshot of How Warby Parker Built a Data-Driven Culture



Introduction

Warby Parker's journey from an early stage startup to a thriving, dynamic, lifestyle brand contains important lessons for companies large and small about how to create a pervasive data-driven culture, something that is often discussed but rarely seen in practice.

Warby Parker's decisions are powered by a precisely-defined and integrated data model

Warby Parker's co-founders Neil Blumenthal and Dave Gilboa have strived for, and inspired, a data-driven culture since the company launched in 2010. However, as the company grew, a clear need for increasingly mature data practices arose. They brought on Lon Binder, now CTO, and Carl Anderson, Director of Data Science at Warby Parker, to spearhead the data maturation process; in a way, this reflects a microcosm of the development of the entire Business Intelligence industry. In a few short years, Warby Parker has gone from a company that was run entirely out of an ERP application, to a world in which the company's decisions are powered by a precisely-defined and integrated data model that is used for discovery and reporting and which has the potential to bring the power of data to bear on every process in the business.

Anderson, author of ***Creating a Data Driven Organization*** (O'Reilly, 2015), and Binder view the transition towards a more sophisticated data organization as a work in progress. The results have already had a huge impact on Warby Parker in the following ways:

- Complex spreadsheets that took hours to integrate vast amounts of data were replaced by scalable automation that executes in minutes.
- Technology and skill bottlenecks were removed so that a much larger group of people at the company can explore the data without help from internal experts.
- Analysts can drill down from high-level summaries to see individual transactions using the familiar pivot table user interface.
- Key definitions of business concepts were migrated from complex spreadsheet formulas to a shared curated model that improved transparency and accuracy at the business.
- Use of new sources of both external and internal data has vastly increased because bringing new data to bear has become much easier.
- Analysts spend a much larger percentage of their time understanding what data means rather than working to create clean, integrated data sets and servicing one-off data requests.
- The speed of decision-making has increased because questions can almost always be answered immediately without having to wait to for the data team to gather and present the appropriate data.

This CITO Research case study will explain how Warby Parker achieved these results. The story will cover the use of various technologies to create a data warehouse and the use of Looker to create an integrated model that supports discovery, analysis, and automation as well as the propagation of data to every corner of the company. As part of this story we will cover how to execute the key, high risk stages of this journey and outline Warby Parker's vision for the future.

Escape from ERP

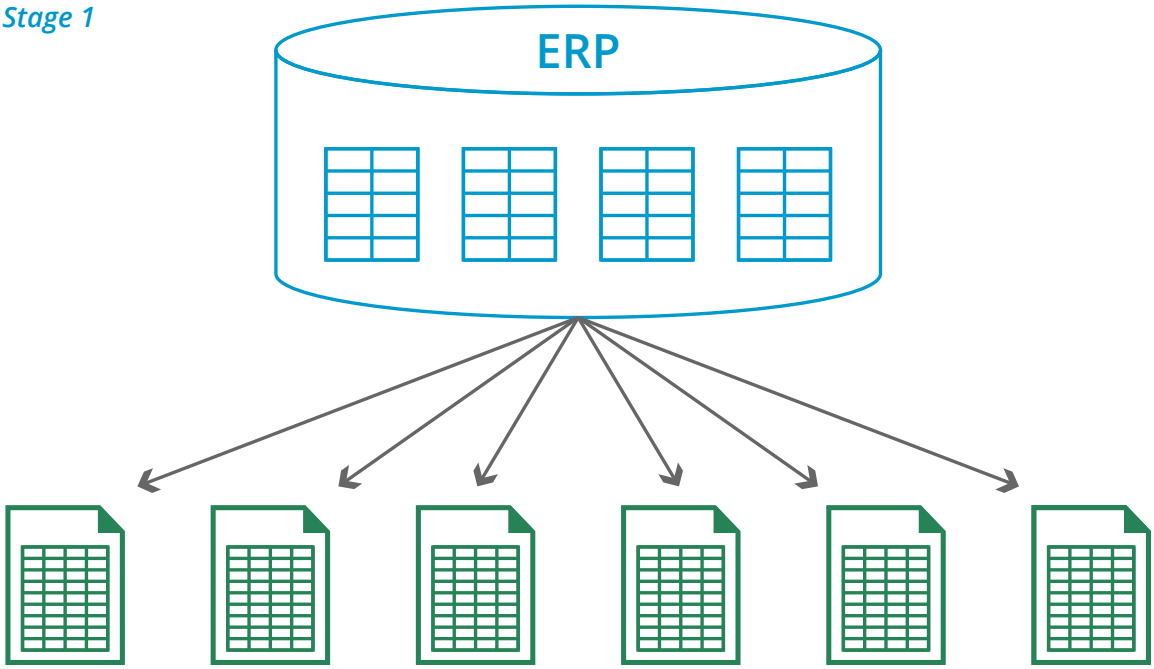
As a first priority Warby Parker focused on understanding and pleasing customers. To get the business going quickly, Warby Parker relied on an off the shelf SaaS-based ERP system that was created to support ecommerce. As a result of choosing this strategy, the company was able to quickly get to market, which was vital to its success.

The solution was to extract data from the ERP system into spreadsheets

As the company grew, however, the need for analyzing business results became crucial. The solution, which again was expedient, was to extract data from the ERP system into spreadsheets.

The flow of data looked like this:

Stage 1



Spreadsheets were used as the way to capture the data, integrate it, clean it up, and define the key metrics.

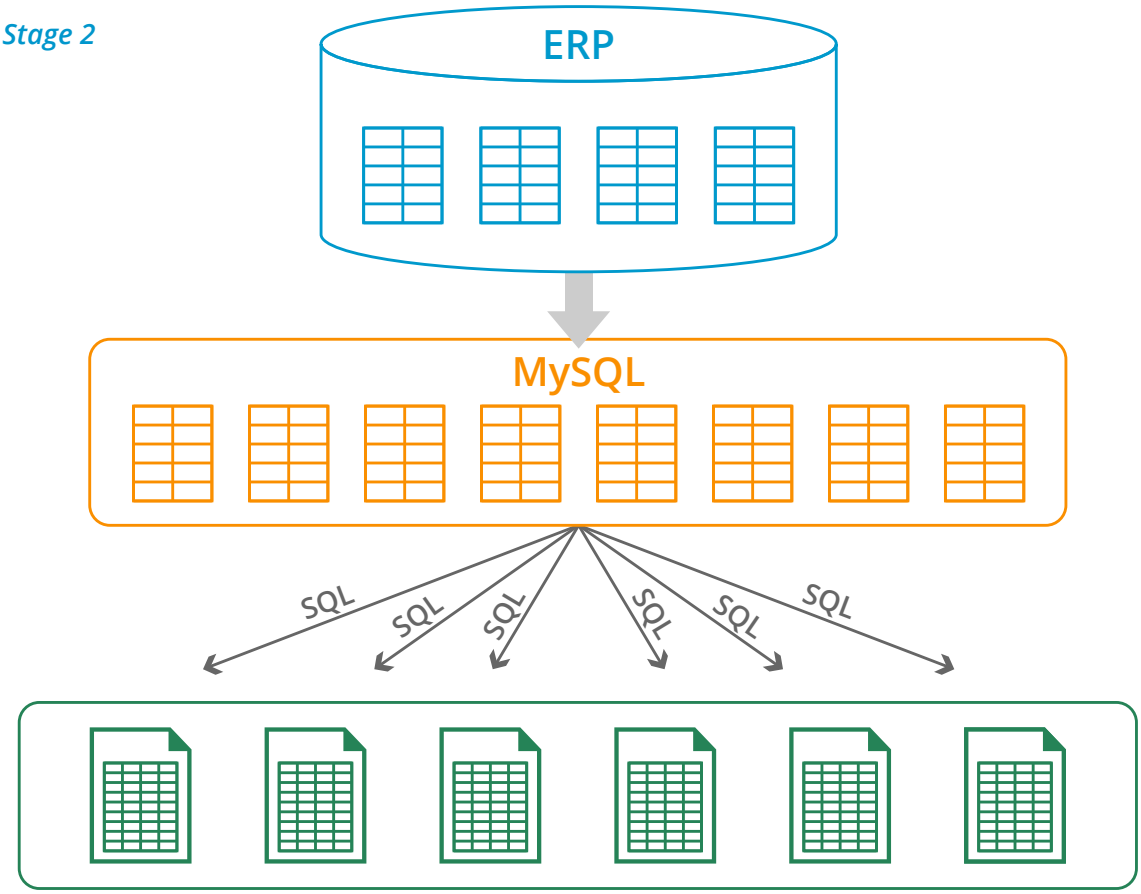
This approach worked well until the data volumes and number of different tables used in each spreadsheet got too large. Spreadsheets that made heavy use of VLOOKUPs and other functions sometimes would take 10 hours to run. If something failed, the spreadsheet would have to be restarted. This architecture could not support widespread use of timely data.

This new architecture drastically decreased the amount of time required to pull analysis

In March 2013, Carl Anderson joined Warby Parker and rapidly brainstormed a solution. Anderson, Lon Binder and their teams charted an analytics escape from ERP by instituting an ETL data integration tool and syncing key tables from the ERP system into a MySQL data warehouse. Analysts were taught to migrate from hacking spreadsheets to using SQL for data integration and cleanup.

This created the following flow of data:

Stage 2



This new architecture drastically decreased the amount of time required to pull analysis. The spreadsheets that were running in hours now could be run in minutes because SQL worked so much faster than the data integration code in the spreadsheets. The analysts who were able to use SQL were able to do a lot more than they could before.

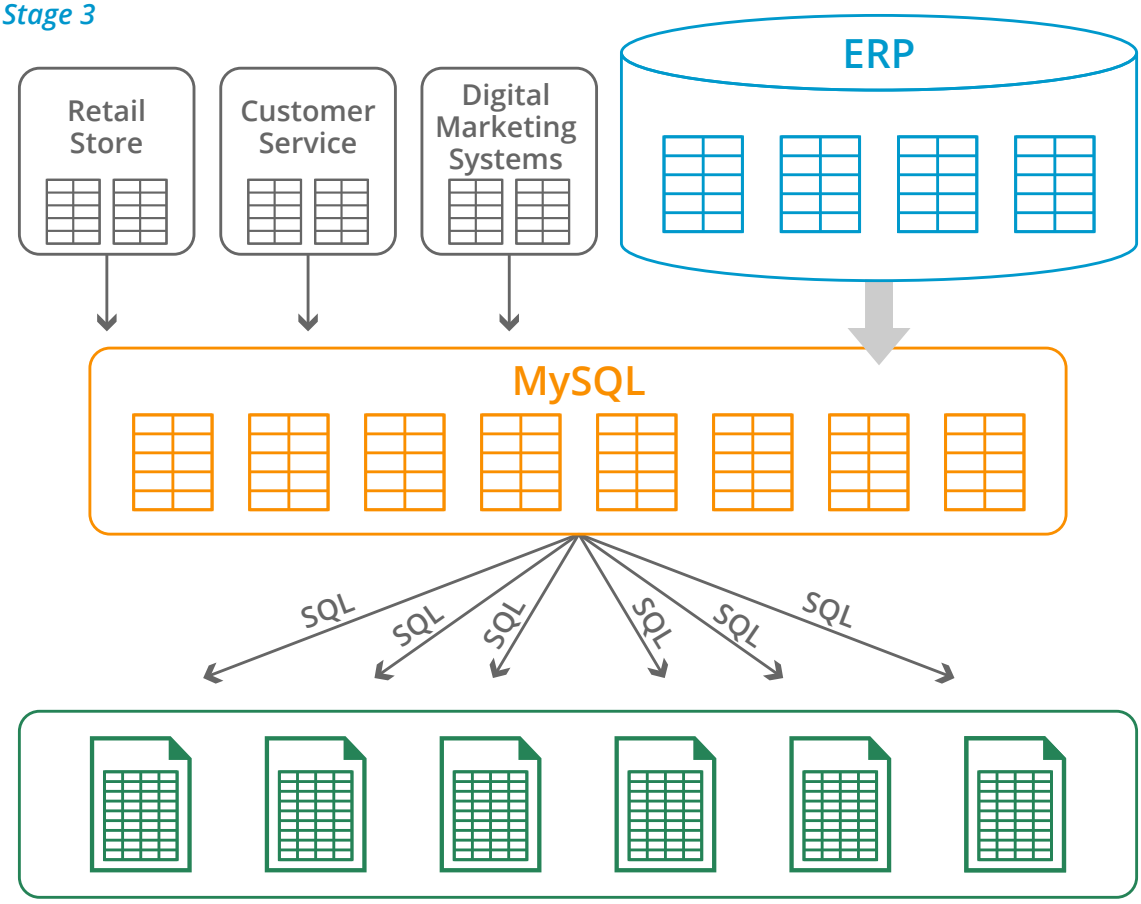
Expand Automation and the Scope of the Data Warehouse

In the next stage, the data warehouse grew in scope. As Warby Parker grew, the data from systems like Google Analytics and other customer service and digital marketing systems provided more information about customer interactions and the state of the business. At first, data from these systems was used via dashboards or by extracting data into spreadsheets. The data from these systems was replicated into the data warehouse for the same reasons that drove the replication of data from ERP.

This enhanced data warehouse helped expand the number of reports that could be created

The result was a flow of data that looked like this:

Stage 3

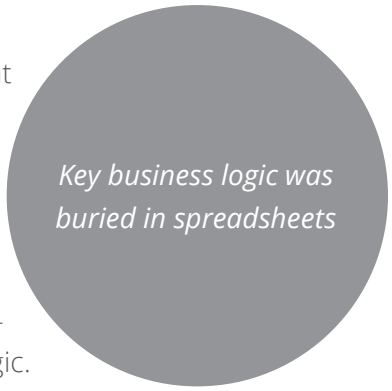


As a result of adding more data, a higher level of skill was required to write SQL to integrate everything. The analysts, who could write SQL to integrate data from a few tables, now had access to many more tables. This enhanced data warehouse helped expand the number of reports that could be created.

Create a Unified Data Model of the Business

At this point, Anderson, Binder, and their teams realized that spreadsheets were still playing a central role in transforming the data from its raw form into the metrics and measures that drove the business.

Key business logic—such as how sales channels are defined, what constitutes a customer, how gift cards are considered in bookings—was buried in spreadsheets, sometimes inconsistent with certain edge cases, and not fully understood.



There were dozens of such business concepts locked away inside complicated spreadsheet formulas. Furthermore, each line of business had its own unique definition of some aspect of the business logic.

Anderson and Binder then realized that the next stage was going to be the most crucial. The business staff was deeply intimate with the data they were using. They invested a great deal of time and effort creating speadsheets that reflected the business.

"We had a choice. If we just put a tool in place that made more data accessible—without a clear, coherent data dictionary—to everyone without some sort of alignment, where would that lead? We were skeptical that it would help and suspected confusion would result and people would stick with their spreadsheets. We realized we had to do the hard work of coming to agreement about the key concepts and metrics we wanted to use to run the business."

Carl Anderson, Director of Data Science

Anderson and Binder looked around for the right tool. They explored tools that made visualizations easy to create, but found that the analysts had a strong appetite for looking at raw data. The visualization technology allowed the visualization to be explored but only provided data at a summary level and did not offer sufficient flexibility. The SQL interface already in house was too difficult to use (it required SQL expertise, which few on staff had), and legacy BI tools were not under consideration because the time-to-value and older architecture did not match Warby's needs.

Anderson and Binder experimented with Looker and found that its ability to model data and provide data discovery capabilities—that is, to present aggregate data in tabular and graphical form while still allowing drill down to the detail records—was crucial for analysts.

In addition, the pivot table interface supported exploration using a familiar user interface, which allowed analysts and end users the flexibility to leverage their advanced Excel skills and perform complex queries without SQL.

Then the hard work began. Using Looker's LookML language for modeling data, Anderson and the Data Science team worked with each department to examine the business logic embedded in their spreadsheets and move that logic, where appropriate, into Looker.

It was during this process that many fascinating things were discovered, such as:

- The spreadsheet definitions often varied from definitions of the metrics.
- Spreadsheets often didn't account for edge cases and complexities in the data.

So, over many months, the Data Science team used LookML to model the improved definitions and to build confidence in the new model among the business teams. Binder called this process "unification," and a large part of that unification involved creating a data dictionary.

"We knew that our unified model of the business with the new definitions of metrics would only have the transformative effect toward a data-driven culture if it were voluntarily adopted by the analysts. But to get that adoption, they had to trust the unified model, and trust is not something you can dictate; it has to be earned."

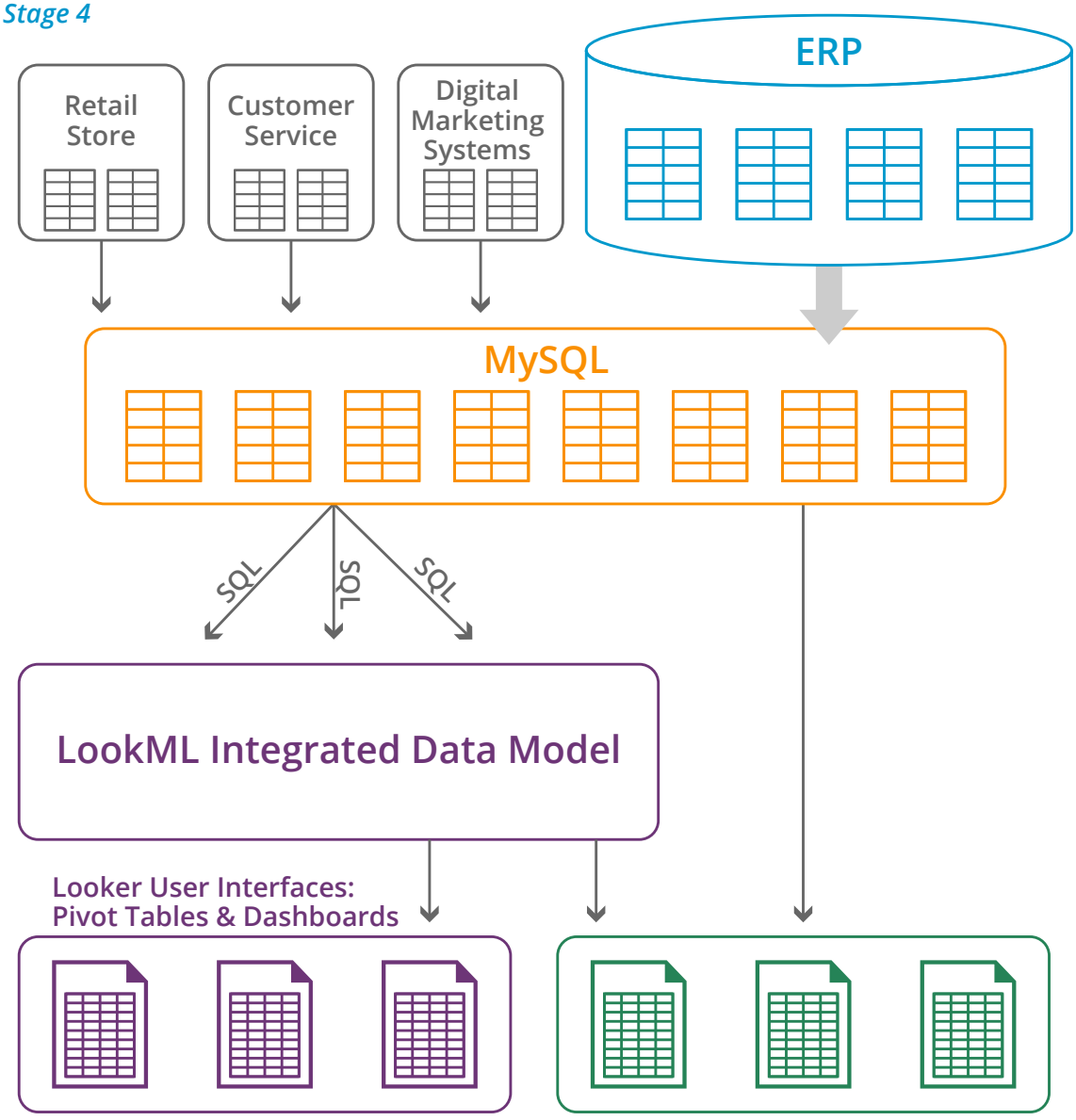
-Carl Anderson, Director of Data Science

To gain that trust, they adopted an iterative process working very closely with the business teams. They discussed the core logic for each metric, thought through edge cases, and got alignment from other teams if definitions differed. Next, they implemented those dimensions and measures in LookML and generated an output dataset to validate against the reference Excel dataset. That meant digging into any differences, working out if it was because of a difference in implementation (a bug) or, instead, if there was some edge case that had not been accounted for (missing logic).

Finally, they reached a state where the values either matched perfectly or had a small set of differences of which the root cause was clearly understood. The team was confident in the superior logic implemented in LookML. Working so closely with the business owners and really understanding any differences in the numbers developed trust so that the business owners were confident to sign off on LookML as valid and ready for use by all the analysts.

As a result the data flow looked like this:

Stage 4



The unification process moved the key definitions from spreadsheet code, which had not been vetted or debugged, into LookML. Using Looker, analysts could now spend far more time exploring the data and understanding what it meant rather than writing SQL or preparing datasets for analysis.

Using Looker, analysts could now spend far more time exploring the data and understanding what it meant

In addition, analysts who could never write SQL could now traverse the entire data model.

"This new common understanding not only expanded the number of people who could explore and become intimate with data. But it also had a powerful effect on helping departments understand each other. The unified model created a common language that improved the speed of decision making."

-Carl Anderson, Director of Data Science

As the unified model became adopted and used via Looker, many delays associated with assembling data disappeared. When employees asked a new question, analysts didn't have to prepare more data and set up a new meeting. They were usually able to pivot and drill their way down to an answer.

"By putting the unified model into Looker, we were able to dramatically increase the amount of time our analysts spend on gaining a deeper understanding rather than performing complex data prep tasks."

-Carl Anderson, Director of Data Science

Spreadsheets are still used, but most of the heavy lifting happens in the unified model using LookML.

Expand Data Use Throughout the Company

Because data was no longer so difficult to use and incorporate into decision-making, the company became hungry for more data.

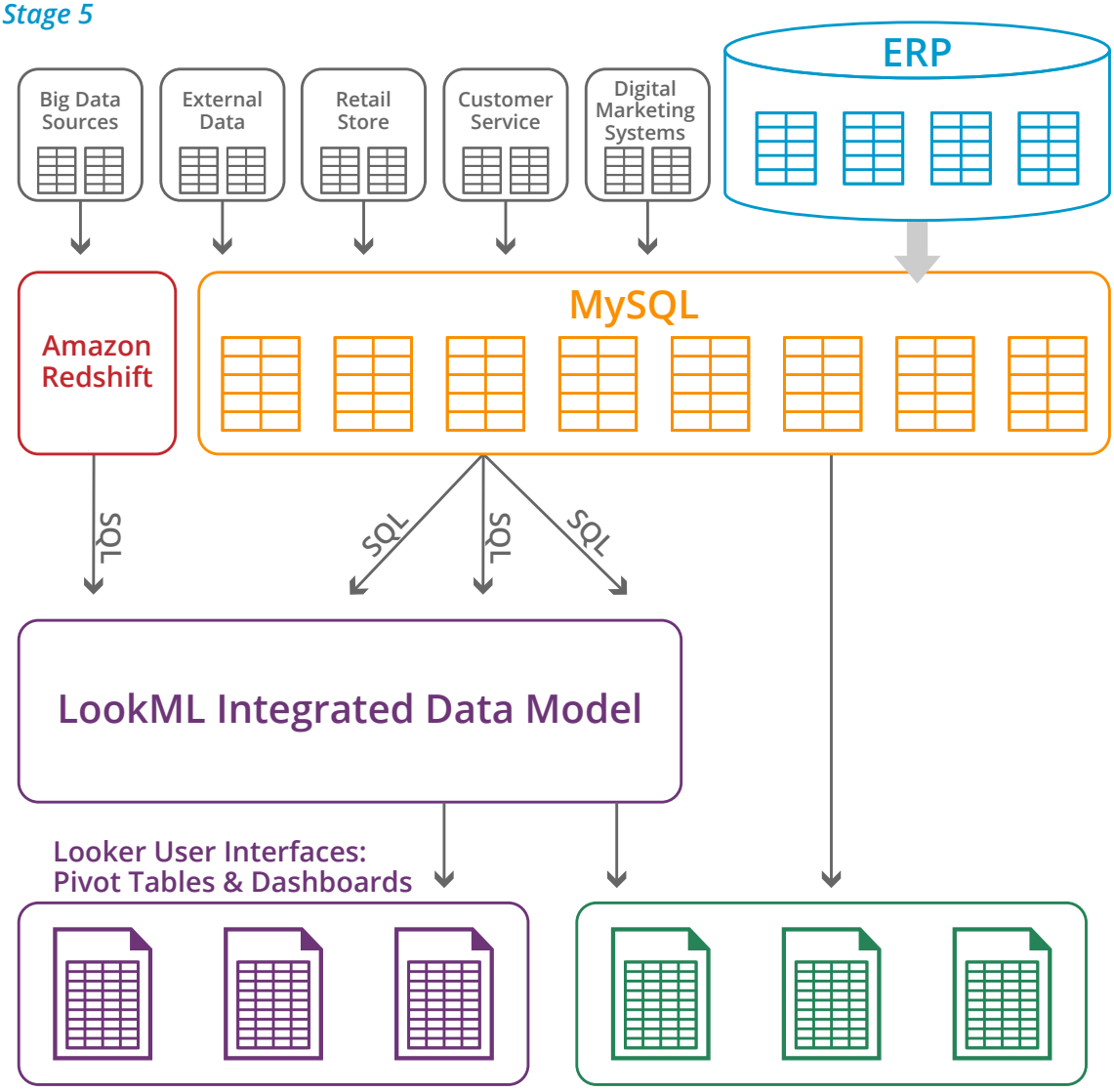
"After Looker became the center of our universe, analysts started asking for demographic data, geospatial data, and other datasets from outside the company. The scope of our data model expanded by extending our unified model using LookML. Once the data arrived, people were quickly able to put it to use."

-Carl Anderson, Director of Data Science

The team also started adding big data by making web log data available to the unified model via an Amazon Redshift database. The company had developed a new ERP system internally and that data, which was in another database, also became available to the unified model.

The resulting data flow looked like this:

Stage 5



In addition, the automation capabilities of Looker are used to send reports to each store. At the beginning of the week, each store manager has a detailed picture of sales and customer interactions. The same is true for key roles in the company such as product and brand managers.

"By expanding the awareness of the unified model, now every team has a better awareness of how their efforts make Warby Parker successful. We have a link between the efforts of individual contributors, who have a better understanding of how their contribution moves crucial KPIs."

-Carl Anderson, Director of Data Science

The Future of Warby Parker's Data-Driven Culture

The future of Warby Parker's data-driven culture continues to look up. Recently, the team developed a vision statement about how to continuously embody a data-driven mindset.

"Our goal is to have data embedded into all of our business processes as much as possible. A data-driven culture is objective, tests assumptions, and challenges new ideas to prove themselves. We want people to know they must always have the data to back up what they are trying to do."

-Carl Anderson, Director of Data Science

Anderson feels that self-service is the key to full realization of a data-driven culture.

"By using Looker we took away the need to learn SQL to enable people to do analysis. Now they can meet most of their needs for data on their own. Analysts can really do as much as they need to do, create their own dashboards, do their own analyses, and share them by themselves to engage in dialogue and interaction with all the other parts of the business. The key thing is that this self service creates a much richer context for decisions, a deeper understanding of the customer."

-Carl Anderson, Director of Data Science

The next frontier for Warby Parker is to add statistical inference and predictive modeling as widespread practices.

"We have created a triple-win concept for our customers, shareholders, and the community by leveraging data from day one. Without the right tools, this required a lot of manual effort, which doesn't scale. To bring the benefit to the masses, our future fate relies on the right data toolchain."

-Lon Binder, CTO

"We need more testing and more statistical inference so we can be more objective about why something is happening. We want more predictive models, more forecasting models, more optimization. This should provide a deeper causal understanding of the business that will be of immense strategic value. For example, we will know how to change our operations and supply chain because our models have made predictions we can rely on."

-Carl Anderson, Director of Data Science

Warby Parker's journey shows the power of data and the steps needed to unlock that power. The journey has multiple stages, none of which are easy. Many companies will seek to create a data-driven culture. The ones that succeed will do so because they have the right people in place, choose the right tools, and do the hard work of creating a unified understanding that is the foundation of a data-driven culture.

Build your data-driven culture. Get started with Looker here: Looker.com/trial

This paper was created by CITO Research and sponsored by Looker

CITO Research

CITO Research is a source of news, analysis, research and knowledge for CIOs, CTOs and other IT and business professionals. CITO Research engages in a dialogue with its audience to capture technology trends that are harvested, analyzed and communicated in a sophisticated way to help practitioners solve difficult business problems.

Visit us at http://www.citoresearch.com