Data Warehouse Trends Report 2019

Executive Summary

As data warehousing technology evolves to keep up with the ever-increasing demands of an engaged audience, it's critically important that we listen closely to those who work in the digital trenches - day in and day out - to meet this demand.

At Panoply, one of the ways we do this is by administering an annual industry survey of data engineers, scientists, analysts - across a wide range of industries, and from companies big and small. We've collected survey responses from live event attendees, like those at Amazon's re:Invent conference, and from online surveys of people from technical, business, and analyst backgrounds.

Panoply has been capturing feedback and releasing this, our annual Data Warehouse Trends Report, for 3 years running. This year, we were fortunate to receive over 800 responses to our survey, which we've designed to both identify and strengthen our understanding of critical data warehouse attributes that drive delight or frustration, and to identify new opportunities within the industry.

Respondents vary significantly not only their roles but in the ways that they utilize and depend upon different data warehousing tools. This report is based on their candid feedback.

Key Findings

Here are some of the major takeaways learned from this year's survey.

Opportunities Abound: Migration to cloud-based data warehouse solutions continues at a rapid pace, estimated recently to be around 10x the adoption rate of on-premise solutions. Despite this, and despite macro-level trends toward machine learning, artificial intelligence, and other data-heavy endeavors, nearly a quarter (23%) of respondents reported that they're not currently using any data warehouse solution at all. This is most common in smaller organizations of between 1 and 50 employees - where a full third of companies have yet to choose a data warehousing partner. When we combine this with the 29% of respondents who are using an on-premise data warehouse offering, it becomes clear that migration to cloud-based solutions is far from complete. This presents enormous opportunity for innovators and market disruptors that offer powerful, affordable, and simple-to-use data warehousing solutions.

Redshift is losing ground: While still the clear leader across the spectrums of both industry and company size, Amazon's Redshift has ceded substantial market-share to other offerings. In 2017, 60% of respondents reported using Redshift. Recent polling suggests that number has fallen to around 40%. Interestingly, other perceived industry leaders (BigQuery, Azure SQL Server, etc.) are not picking up that slack in any significant way, making only modest gains year over year. Instead, offerings like Panoply, Snowflake, and Oracle are emerging as legitimate market contenders.

Complexity remains a significant 'sore spot' for data warehouse users: For the third year in a row, more than half of data warehouse users characterize their data warehouse solution as "difficult" or "very difficult" to use. This is true, albeit to varying degrees, of every brand we asked about. In the case of Redshift, 62% of respondents found it difficult to use, due largely to issues around complexity. Others fared marginally worse.

It's also worth pointing out that those who reported using an on-premise solution were by far the most likely (81%) to call out complexity as a key pain point. Clearly, a strong case can be made for the virtues of cloud-based data solutions, as opposed to on-site, siloed offerings. However, there's little evidence to suggest that tangible progress is being made toward driving ease-of-use for those who've already embraced cloud solutions. There's never been a better time for solutions like Panoply, that work to simplify the complex and automate labor-intensive, redundant tasks.

Performance issues are also persistent: When asked: "How frustrated are you with your data warehouse's hanging or slow queries?" - more than half of respondents chose either "somewhat frustrated" or "very frustrated". While it's difficult to ascertain the root causes of this frustration, we can surmise that the boundless appetite for more complex queries, larger data sets, and deeper analysis has outpaced advances in technology for the majority of data warehouse users. Additional factors like the integration of multiple data sources, lack of database normalization, and bandwidth (both human and digital) limits may also come into play here, adding to the perceived

urgency of each interaction. As it takes more and more to 'feed the beast', we anticipate a growing need to simplify, automate, and streamline core processes wherever possible.

Survey Results

Fig.1 What are you using for a Data Warehouse?

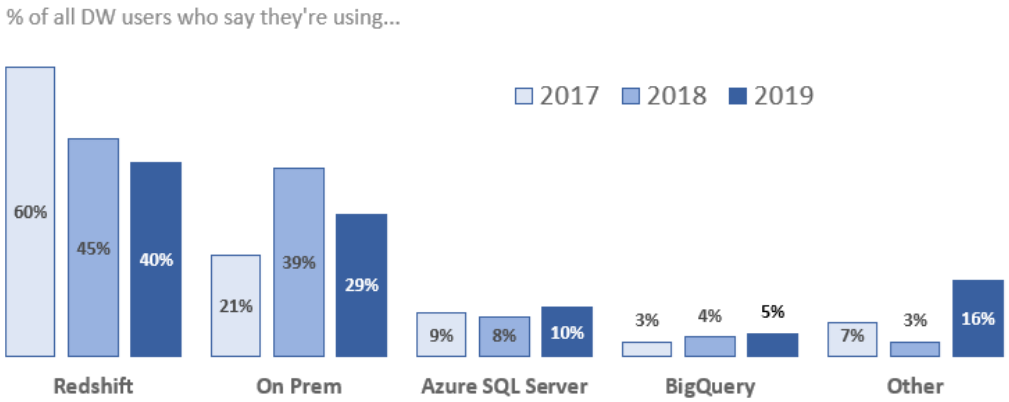


Fig. 2 How would you rate the complexity of operating your data warehouse?

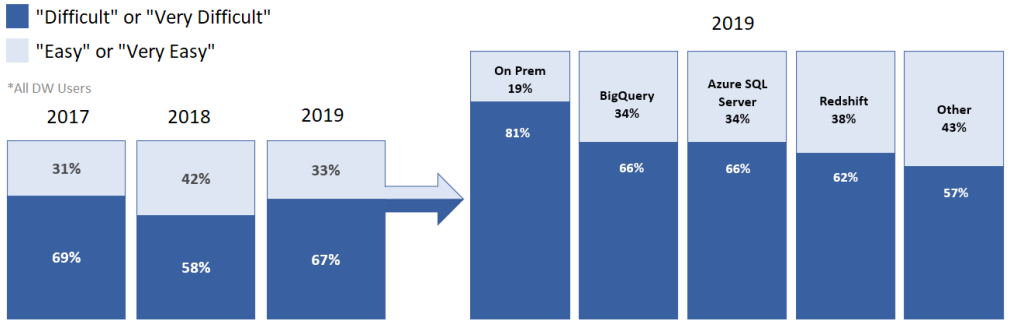


Fig. 3 Is data warehousing too complex in your industry? (% Respondents who rated their Data Warehouse solution as “Difficult” or “Very Difficult”)

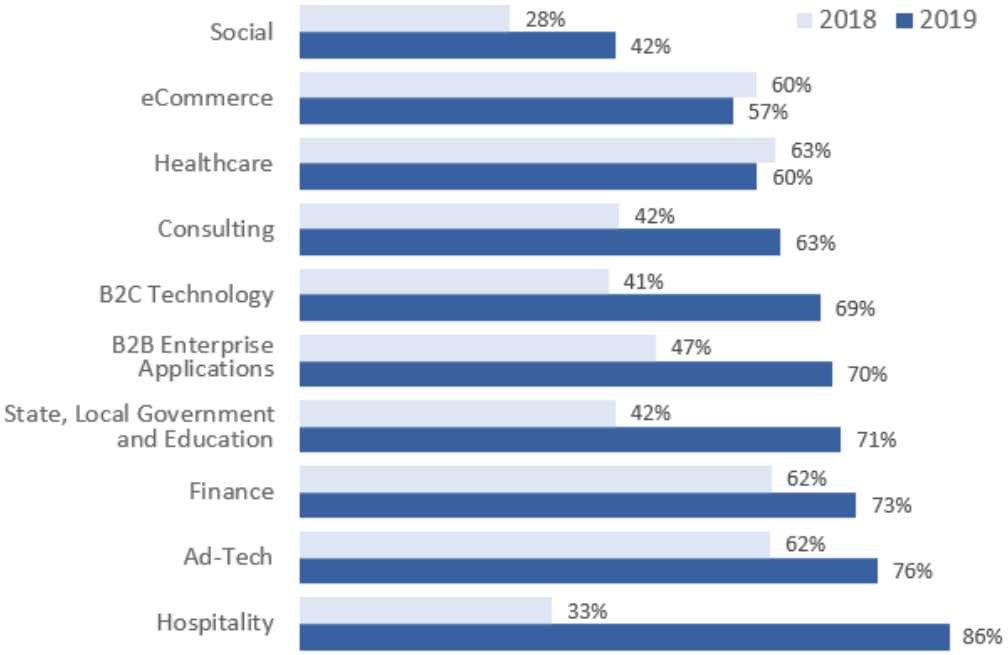
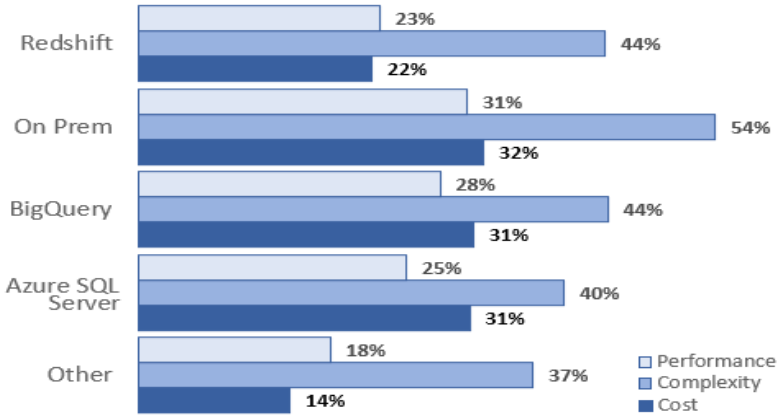
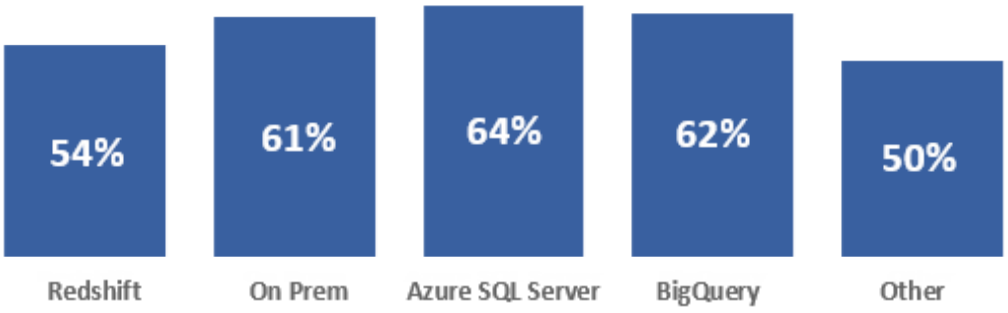


Fig. 4 Why did you rate your Data Warehouse solution as “Difficult” or “Very Difficult”?

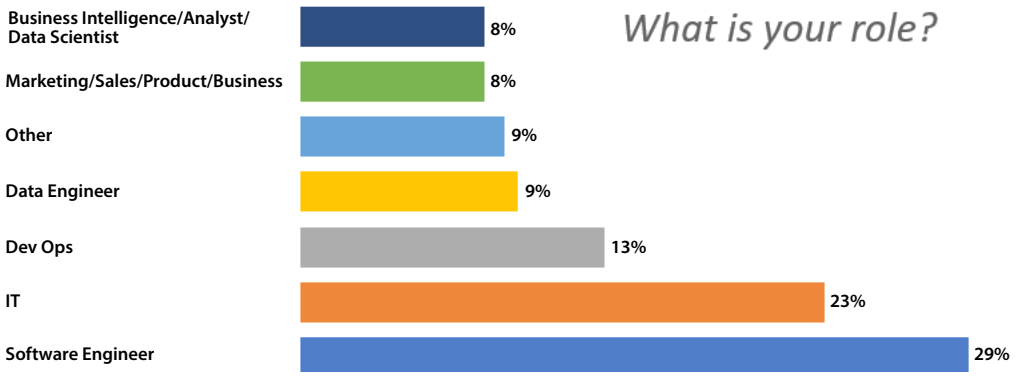


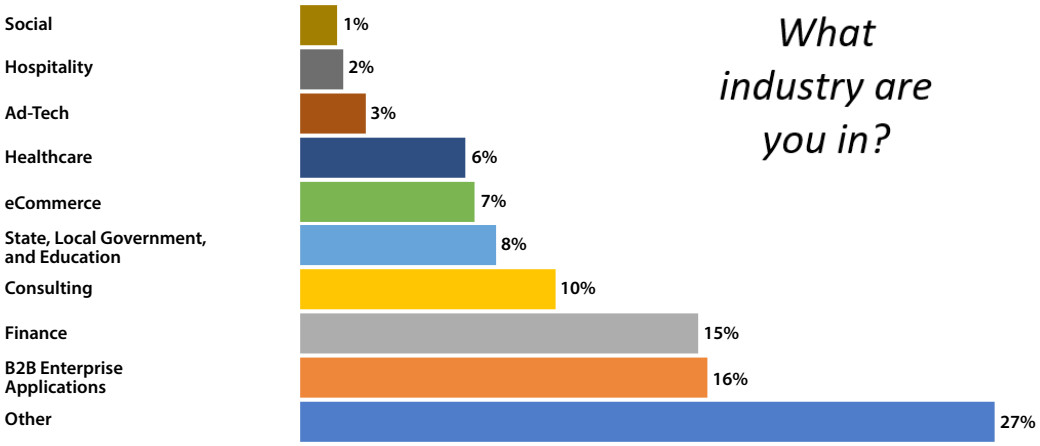
***Note:** Respondents were allowed to choose more than one answer.

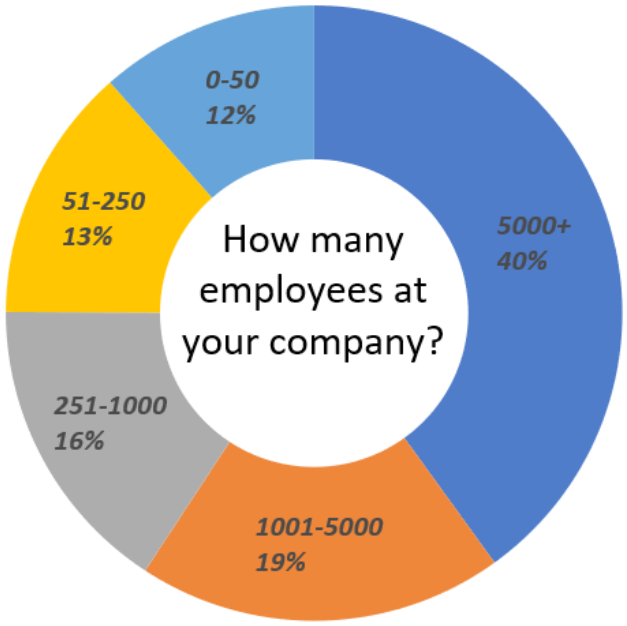
Fig. 5 How frustrated are you with your data warehouse’s hanging or slow queries?



About the Survey Sample







Final Notes

To summarize, data warehousing continues its brisk migration toward cloud-based solutions that are more powerful, customizable, and dependable. While Redshift leads the pack, that lead is shrinking, and there are myriad opportunities for others to gain market share, by addressing some key pain points - simplicity, performance, and cost. Winners in this space will need to deliver a product that caters to the technical and non-technical alike, offering a simple UI, robust functionality, and the processing power to keep pace with the ever-growing need for more complex queries, larger and more disparate data sets, and real-time, actionable insights. The simplification of complex operations and the automation of redundant or labor-intensive tasks are critical to the evolution of data warehousing, as demands rise and resources struggle to keep pace. That's where Panoply and others make a real difference - by streamlining the effort required to turn data into insights, and insights into sound business decisions. In today's hyper-competitive atmosphere, fast, data-driven decision-making is often the fuel that drives rapid growth and sparks market disruption. From SMBs to Enterprise-level organizations, the capacity to ingest and interpret big data, and to respond accordingly, has become the 'must-have' advantage of modern industry.

hello@panoply.io | www.panoply.io | 188 King St. San Francisco, CA 94107