# panoply

# Data Warehouse Trends Report
# 2018

# Table of Contents

# Executive Summary

The recent 6th annual AWS re:Invent conference offered a unique opportunity to connect with data and cloud experts, collaborate at AWS bootcamps and learn how AWS can improve productivity, security and performance. It brought us face-to-face with the individuals who handle the day-to-day work in the digital data trenches. We gathered feedback from these professionals to better understand cloud-industry trends and the market's ever-increasing needs. Their feedback is especially important as the volume, variety and velocity of data continue to increase at a blistering pace.

re:Invent was also the setting for our annual survey, from which this report is crafted. Our findings offer compelling insights into today's use of data warehouse systems, such as Redshift and Azure, while providing us with a better understanding of data professionals' biggest challenges, their satisfaction with current solutions and the ways in which they use BI solutions in concert with their data warehouses. The survey includes input from over 1600 attendees, across a wide variety of industries, and holding roles that range from Data Analysts to C-level positions.

## The Data Warehouse

Expectations of data warehousing have evolved with the advances in technology and greater focus on data-driven business strategies. No longer thought of as simply a repository for disparate data sets, today's data warehousing solutions are tasked with facilitating a more holistic, less siloed approach to data, that unifies and aligns an entire organization toward shared goals and business outcomes. Solutions are now routinely expected to handle structured and unstructured data, offer easy access to the data, quickly integrate new data sources and most importantly, provide a platform that can quickly adapt to constant change, as organizations' needs evolve. In today's hyper-competitive environment, there is always a need for a data-driven 'edge', in the form of new dashboards and insights, deeper analyses and must-have tools that present data

in new formats to an ever-widening group of stakeholders. Marketers, for example, are increasingly eager to get their hands dirty, and work directly with the data. Data professionals, similarly, are becoming more involved in the formation of their company's strategy, for which data is a key input.

To meet these evolving needs, data warehouse technologies require leading-edge data virtualization, processing, and transformation methods. There are several delivery platforms to make this happen, including physical appliances, such as dedicated traditional storage subsystems built to support analytics and BI performance. With the addition and ongoing evolution of the cloud, today's cloud-based solutions, such as Amazon RedShift and Google BigQuery, aim to simplify both the hosting and analysis of data in an increasingly complicated environment. It's clear that data warehousing is an industry on the rise, poised to become as integral to an organization's daily operations such as email, Slack, or coffee.

Marketers and data practitioners are faced with many data-related challenges, many of which are outlined in this white paper. Here, we highlight our key findings from the data warehouse community, as well as key areas to address in order to accelerate data adoption and keep up with ever-increasing data demands.

**1** https://aws.amazon.com/redshift/
**2** https://cloud.google.com/bigquery/

# Key Findings

**Many companies still do not have a data warehouse**

Surprisingly, 21% of the people we interviewed stated they currently do not have a data warehouse solution. This seems relatively high, considering the fact that we surveyed people at re:Invent 2017, an event specifically designed for infrastructure and data professionals. One of the obstacles to embracing a data warehouse may be the belief that they still have to denormalize and aggregate their data, unaware that data warehouse tools such as Panoply can address these issues. Of those companies currently leveraging a data warehouse, 45% use Redshift. This number, while impressive, shifted downwards from the 2016 results, which showed 60% using Redshift. We should note, however, that the survey was conducted at an Amazon event. So this drop may be more indicative of wider interest in the event, rather than a shrinking market share. Following Amazon (at a distance) is Azure and Big Query, who make up 8% and 4% of the market respectively. Despite Redshift's continued dominance in the space, Microsoft and Google are making inroads into this market with aggressive marketing. In some respects, Redshift opened the door to cloud-based data warehousing, and companies such Panoply are reducing the complexity.

**Complexity is still a challenge**

The majority of survey respondents, over 62% of them, indicated the management of their data warehouse solution is difficult or very difficult. When asking respondents to tell us why they were unsatisfied, they mentioned complexity, cost and performance. Specifically for Redshift some respondents were unsatisfied because of Complexity and Performance, with 38% of respondents saying it's too complicated. For SMBs with 250 employees or less, 71% of them had a warehouse, and of those 58% found managing their warehouse too difficult. The bottom line is whether a company is big or small and even if they are already in the cloud, managing a data warehouse is often far too complicated.

### Automated tasks desired

To understand the complexity issue better, we asked if infrastructure teams would prefer if data scientists and analysts should manage their own data flow, with automated tools. 81% of respondents said yes, they would prefer data professionals to have these tools. When asked about what processes they would want to automate in their data warehouses, respondents mentioned four areas: ingesting different data sources, transforming data, managing data and query optimization. 'Transforming Data' was mentioned most often by Redshift and Azure users, while BigQuery users showed a bigger appetite for 'Ingesting Different Data Sources'.

### BI tools

Among respondents to the survey, almost 56% use Tableau, which is flat, relative to last year's findings. The next group was Microsoft PowerBI (20%). Some key trends accelerating usage of BI Tools include the desire for self-service, de-silofication, better data visualization and having a cloud-based solution. Other key trends that continue to drive adoption are AI and Machine Learning, which help to automate data-driven processes and accessibility.
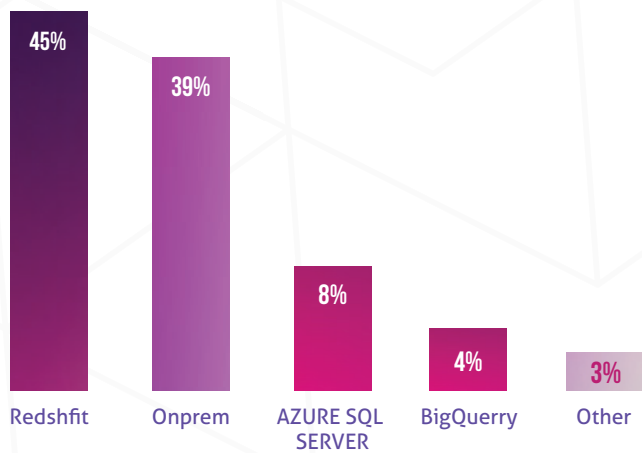
### Other insights

Most respondents indicated that they learn about analytics and data management from LinkedIn, and surprisingly, on Reddit. Facebook and Twitter lagged behind as educational resources. Also noteworthy — a person's job level (Analyst, Director, C-Level, etc.) had little impact on warehouse satisfaction rates, nor did company size.
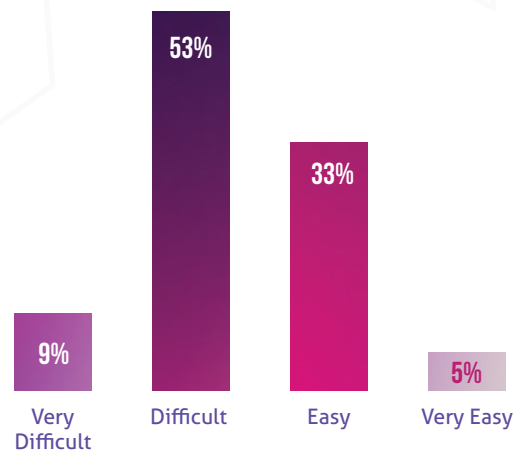
# Survey Results

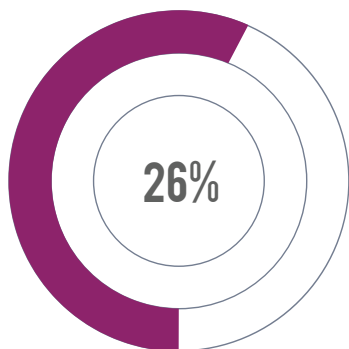**At the AWS re:Invent conference, we surveyed over 1,600 data professionals.**
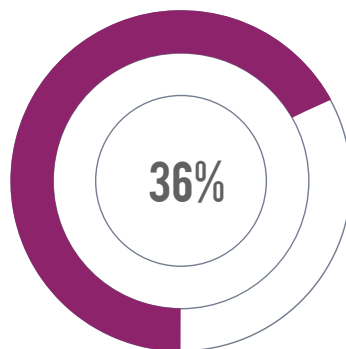
What are you using for a data warehouse?

| Redshfit | Onprem | AZURE SQL SERVER | BigQuerry | Other |
|----------|--------|------------------|-----------|-------|
| 45% | 39% | 8% | 4% | 3% |

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

| Very Difficult | Difficult | Easy | Very Easy |
|----------------|-----------|------|-----------|
| 9% | 53% | 33% | 5% |

What are Redshift users unsatisfied with?

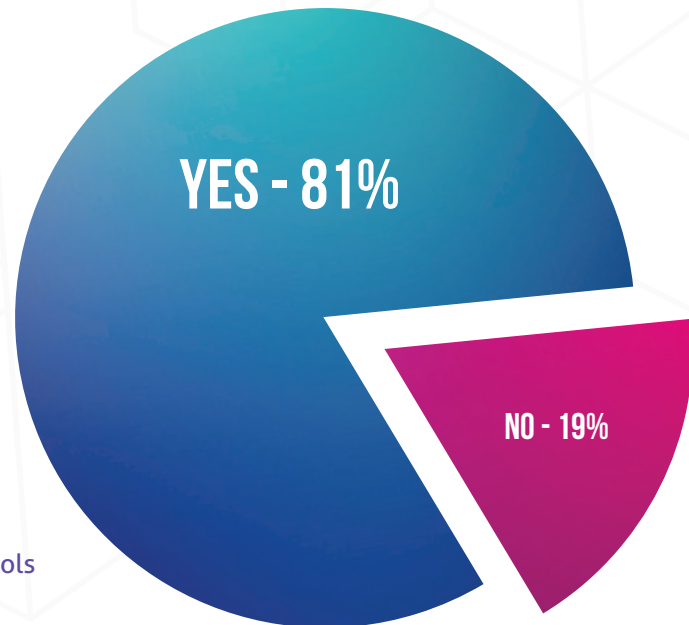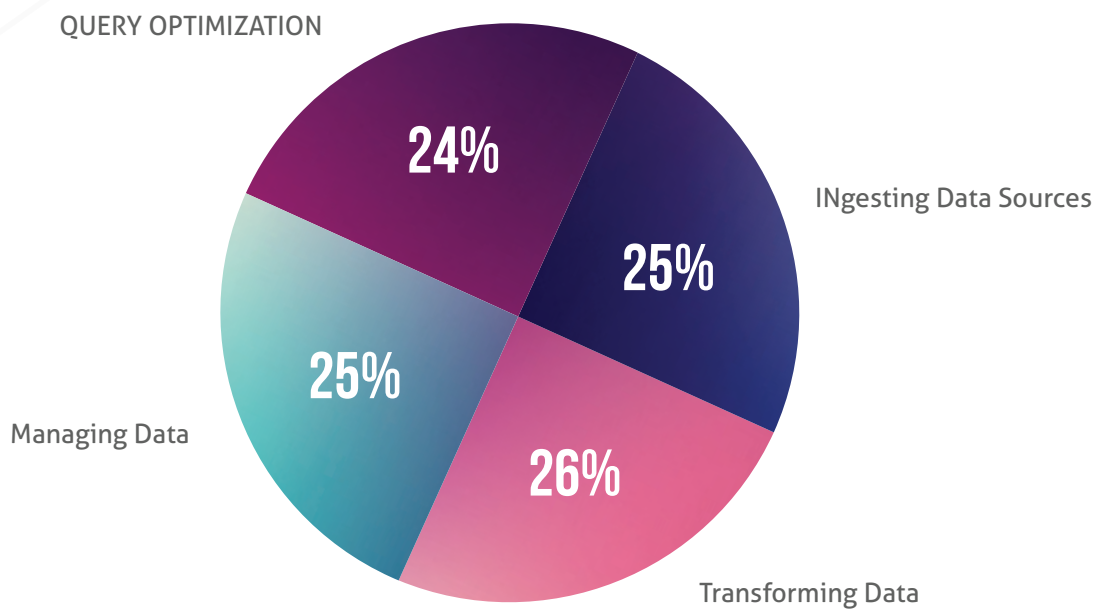| 26% | 36% | 38% |
|-----|-----|-----|
| COST | PERFORMANCE | COMPLEXITY |
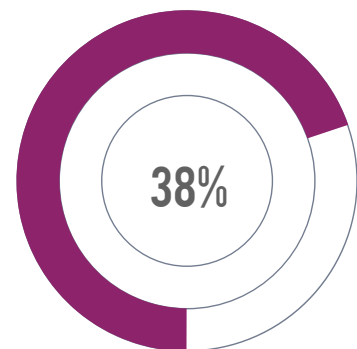
panoply

Would you prefer if analysts and data scientists could manage the entire data flow with automated tools?
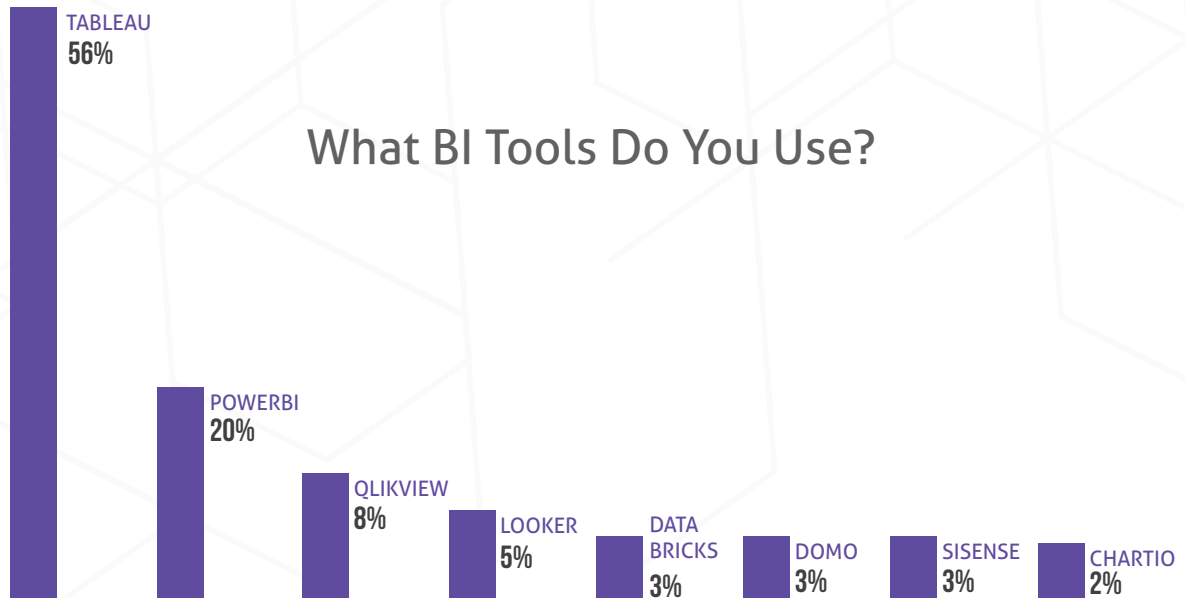
YES - 81%

NO - 19%

* 80% of respondents thought it makes sense to allow analyst and data scientists self manage their data with automated tools

What do you want automated in your Data Warehouse?

QUERY OPTIMIZATION

24%

INgesting Data Sources

25%

Managing Data

25%

26%

Transforming Data

panoply

## What BI Tools Do You Use?

TABLEAU
**56%**

POWERBI
**20%**

QLIKVIEW
**8%**

LOOKER
**5%**

DATA
BRICKS
**3%**

DOMO
**3%**

SISENSE
**3%**

CHARTIO
**2%**

## Is data warehousing too complicated in your industry?

Healthcare
**63%**

Ad-Tech
**62%**

Finance
**62%**

eCommerce
**60%**

B2B Enterprise
Applications
**47%**

Consulting
**42%**

State, Local
Government
and Education
**42%**

B2C
Technology
**41%**

Hospitality
**39%**

Social
**29%**

# About the Survey Sample

Software Engineers — 34%

IT — 24%

Dev-ops — 17%

Data Engineers — 9%

Marketing and Sales — 7%

Product Leaders — 3%

Business Intelligence — 2%

Analysts — 2 %

Data Scientists — 2%
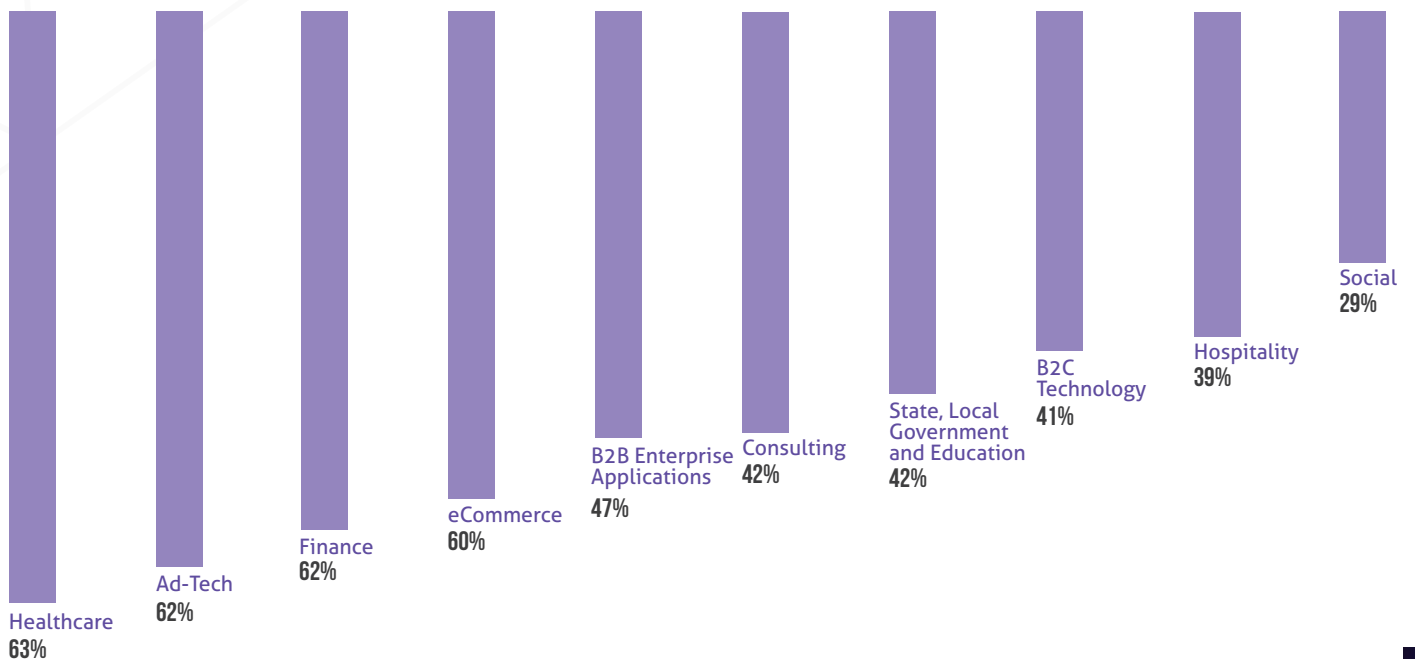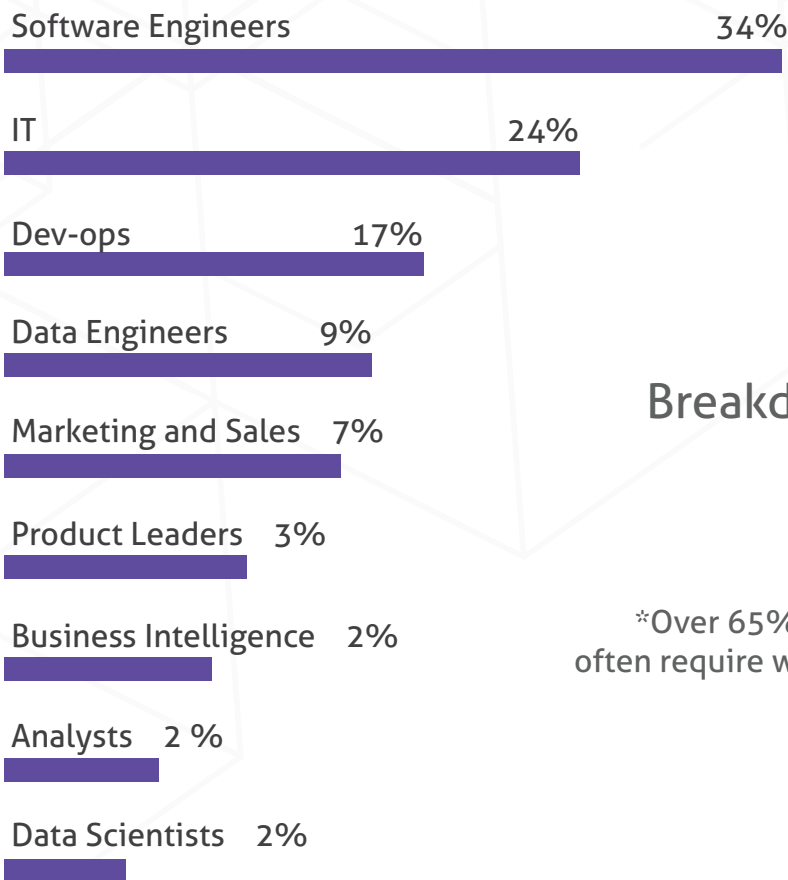
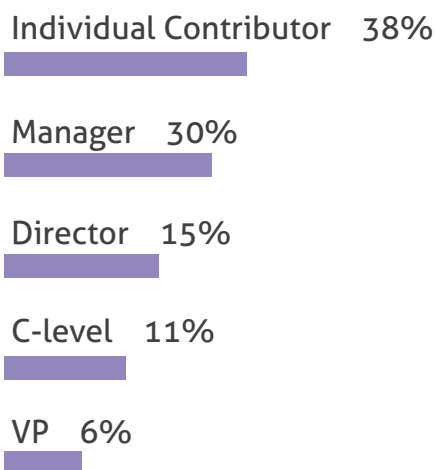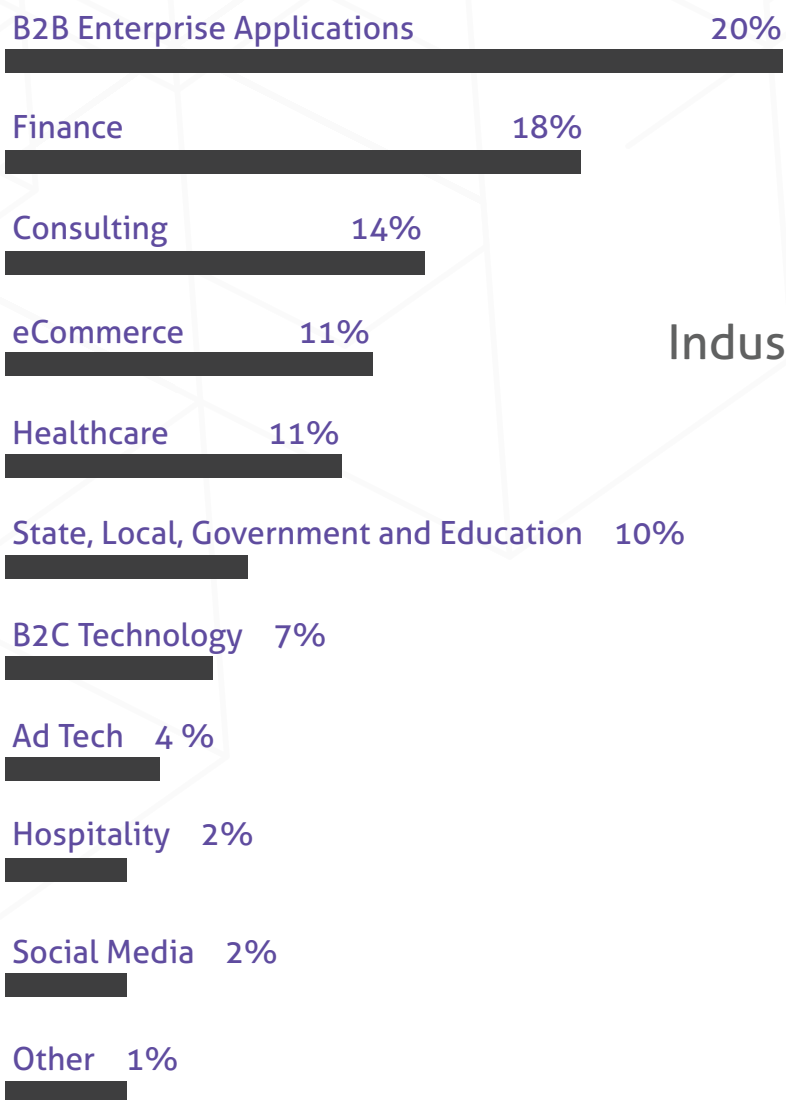## Breakdown of their role in the company

*Over 65% worked in positions that often require working with data directly.

Individual Contributor — 38%

Manager — 30%

Director — 15%

C-level — 11%

VP — 6%

## Breakdown of their job level

## Industry breakdown

**B2B Enterprise Applications** — 20%

**Finance** — 18%

**Consulting** — 14%

**eCommerce** — 11%

**Healthcare** — 11%

**State, Local, Government and Education** — 10%

**B2C Technology** — 7%

**Ad Tech** — 4 %

**Hospitality** — 2%

**Social Media** — 2%

**Other** — 1%

# Final Notes

In sum, data warehousing continues to move beyond simple storage, calling for new ways to ingest data, extract information and analyze data. Companies, however, are faced with tight timelines, limited resources, the desire to respond to insights real-time and ever-changing needs, so data warehouse providers need to automate more tasks and make data more accessible for both their SMBs and Enterprise clients. It is also necessary to address the complexity issue, and provide an easy-to-use and seamless end-to-end data management service, which can streamline the data journey from source to analysis, thus reducing the time for data to become actionable. Automated data management solutions, like Panoply, address these gaps — automating many data related tasks, freeing up time for data gurus and everyday data-driven decision makers, enabling them to work with analytics more efficiently, so they can make faster and better business decisions.

**ABOUT PANOPLY**

Our story begins with an idea: In the Big Data era, free up your data engineers and scientists, and you create value for your customers and your business. It's simple, right? We believe in taking the load off the IT and data engineers that have long been mired in time-intensive tasks like schema building, data mining, complex modelling, performance tuning... Our easy-to-use platform gives small and medium businesses the tools to harness Big Data and get analytics quickly, so they can make faster and better business decisions.

**panoply**