To paraphrase an old saying, if you torture data long enough, it'll tell you what you want to hear. And putting big data through that torture only lets us tell bigger lies. Marketing can justify crazy ad campaigns: "Sentiment analytics shows our latest campaign is actually a huge hit with the under-25-urban-vegan demo!" The supply chain team can use it to get more funding: "Our geolocation analysis shows if we invest in robotic warehouse automation, we'll reduce costs by 15%." Sales can explain why it missed its numbers: "We don't have an iOS app, and smartphone data shows that's what 87.4% of customers use. It's not our fault."

Don't get us wrong. The ability to collect and analyze data is a core IT value proposition. Companies such as Wal-Mart, FedEx, and Southwest Airlines gained strategic advantage by digging into their core business data long before it was labeled "big." And there's no question that more data is available than ever before, especially information from the Web and smart mobile devices. Our beef, though, is that most businesses aren't good at using the data they have now. What are the odds they'll get better at analysis by adding volumes without changing their strategies?

Our InformationWeek 2013 Big Data Survey shows that some companies are making progress. For example, most have built the required infrastructure and support various roles, in terms of primary data users; about one-third say they encourage wide access to information for business users. However, when it comes to data acquisition and use models, the wheels start to fall off. There are major gaps in data analysis, even for the most common types of information: transaction data, system logs, email, CRM, Web analytics.

Worse, fewer than 10% of the respondents to our survey say that ideas for promising new data points are primarily driven by a collaborative or cross-functional team within their companies. The stats we gleaned from our survey suggest this percentage should be much higher: Nearly half of respondents have 500 terabytes of data or more under management; 13% have more than 10 petabytes. Surely there are untapped riches.

IT organizations clearly know there's a problem, as only 9% of respondents rate their companies as extremely effective users of the data they have. However, just 4% admit they stink at putting their data to its best use. Fact is, many organizations are deluding themselves into thinking they're empowering their businesses. So
before you buy more storage, upgrade your warehouse platform, or spin up a massive Hadoop instance, let's take a reality check. Here are six big data lies organizations tell themselves. How many have you heard lately?

Lie 1: We understand how much data we have today

We asked in our survey which of seven key data sources are actively managed, hoping to see respondents widen their view beyond servers, storage arrays, and archives. Unfortunately, only 30% of respondents factor in their organization's cloud data, and just 11% include supply chain information. All that information zipping around on mobile devices? Considered by just 35% of survey respondents.

If you don’t include dynamic data sets, you’re setting your analysis up for failure. How can you do vendor performance reviews without details on how well suppliers do getting the right goods to you at the right time for a competitive price? Likewise, if you’re studying customer behavior, how can you get a true picture without Web or cloud-based CRM data?

Lie 2: The data we have is good

We’ll bet money that most respondents’ data sets are inaccurate, incomplete, and/or misaligned with one another. Do you really have a single source of truth? Do different groups slice data in different ways? Are you making decisions based on inaccurate or incomplete data?

Case in point: 19% of companies in our survey use geolocation as part of their analysis strategy, pulling information from smart devices and Web visitors to understand behavior. However, Web location tracking is notoriously inaccurate when it comes to enterprise and institutional traffic. That’s because most companies and government agencies work in private clouds with a limited number of egress points. If you’re using Web location data to track the success of your sales and marketing programs by region, you’re likely basing decisions on bad information. That big block of traffic from Boston may actually come from an enterprise with offices in the Midwest.

Who’s checking data quality? Just one in four respondents identified a dedicated business analyst group as one of the top two users of data within their company. It’s simply amazing how many reports and graphs we see without sampling or accuracy notes. For example, almost every company does customer surveys, yet very few indicate confidence levels or bias results. Got 25,000 customers? Your customer service survey should have 1,843 respondents if you want a 99% confidence level with a plus or minus 3% margin of error. Furthermore, results should be biased by revenue level. Reality is, we just don’t see that done with any type of data.

Lie 3: Everything will be OK if we can just get more tools

A quarter of respondents plan to use more big data tools over the next 12 months. Now, we like Hadoop, NoSQL, Splunk, and the plethora of other big data options out there, but we recommend looking at what data sets are sitting idle before cutting a check. Given the low levels of use of the 20 internal and external data sets we asked about, it’s clear the problem is related more to staffing than systems.

Unfortunately, fewer respondents plan to invest in big data staff versus spending money on technology. Only 33% plan to grow their training and development programs; 9% are cutting back. Net new hiring ranked at the bottom of our list, with 17% growing staffing.

Data: InformationWeek 2013 Big Data Survey of 257 business technology professionals at companies with 50 or more employees, September 2012.

Get This And All Our Reports

Our full report on our 2013 Big Data Survey is free with registration at informationweek.com/reports/2013bigdata

This report includes 32 pages of action-oriented analysis, packed with 23 charts.

What you’ll find:
> Outlook on use of cloud services for big data management
> Top 14 big data tools in use, from Excel to Exadata to EDW
levels compared with 14% cutting.

Nowhere is this “tools over people” focus more evident than in healthcare. The federal government’s electronic records incentives have driven the industry to a new level of data collection and reporting. But now that healthcare providers have all this data, they’re trying to figure out how to use it. “There is big money to be made in healthcare big data, so everyone and their brother was throwing up solutions,” says Bill Gillis, CIO of Beth Israel Deaconess Physician Organization. But it’s important to work with people who understand healthcare organizations and the complexity of their data, Gillis says. “The business need should drive the process,” he says. “The tool alone will not change much. Finding a skilled hand that can effectively wield that tool will.”

**Lie 4: There’s an expertise shortage**

Speaking of staff, an oft-quoted McKinsey & Co. study estimates a shortfall of 140,000 to 190,000 people in “big data staffing” by 2018. Our own InformationWeek Staffing Survey shows that 18% of respondents focused on big data want to increase staff in this area by more than 30% in the next two years, but 53% say it will be difficult to find people with the right skills.

Roll the clock back a few years and substitute the words “virtualization engineer” or “Cobol programmer” or even “webmaster” for “big data specialist” and you’ll similarly find people predicting doom. Don’t get sucked in again. You already have much of this talent within your organization; you just need to set it free. Consider that 39% of respondent organizations have department-level analysts as the main users of their information. Break those people out of their department silos and move them toward a more holistic view of data.

For example, a U.S. retailer we worked with always had separate data teams aligned with various departments. The strongest team was within the catalog group, banging out circ plans, catalog yields, conversion rates, even profit per page. Great stuff, but that team was limited to using catalog and financial data. Siloed from the Web team, they were missing the transformation happening within the customer base. Separate departments, separate views of the truth.

Don’t blame the analysts. The company built the structure, and IT wasn’t involved enough to identify the information gaps between departments. Our point is, IT not only has to understand the data itself, but it must also become an integral part of identifying and growing the centralized talent pool. That means putting more emphasis on training and talent development. Competitors will steal some of your more talented big data pros if you don’t give them a reason to stick around.

We scanned the online listings of the major hiring sites and found that the salary levels for data analysts still range from about $55,000 to low six figures. However, if you add “big data” to the standard titles, the average salary doubles. Expect everyone’s LinkedIn titles to change in the next 12 months.

**Lie 5: We know what data we need**

In our survey, we asked about 10 internal and nine external data types.
Internal sources include financial accounting applications, detailed sales and product data, CRM data, unstructured network data such as Office files and images, and unstructured data stored on end user devices. External sources include government statistics and other public records, geolocation data, data collected from sensors on company products and services, social network data (Facebook, Twitter), and unstructured data stored in the cloud (Office365, Google Docs).

Clearly, there’s a lot of information out there. But when we asked who’s driving data analysis ideas, we were surprised to find that only 5% of respondents have a centralized team to drive big data strategy; an additional 3% use a looser collaborative effort.

We’re not the biggest fans of committees, but given the fact that the users of your data are likely spread far and wide, it makes sense to create a cross-functional group to identify new sources or elevate the importance of an existing stream. It’s staggering to see some of the great data that’s all but untouched.

Take CRM, phone, email, and Web analytics. These four data points cover most of the communications relationship with your clients. Tying them together isn’t rocket science, especially if you have decent baseline customer data to start with. Not only can you determine the number of conversations your company typically has with customers, but you can also understand how email relates to phone calls and Web traffic. If you have an outside sales force looping in, your CRM data gives you a profiling capability to model everything from product rollouts to customer service problems.

All that intelligence exists today, yet few companies have this level of analysis integrated into their big data strategies. While 35% of survey respondents say their IT organizations include CRM in their integrated plans, only 29% include email, 22% Web analytics, and 14% phone logs.

**Lie 6: We do something with our analysis**

There’s nothing more frustrating for an analyst than to work for days or weeks on a project, present the findings, have a great meeting with execs, then watch those recommendations die on the vine. Everyone focuses on the positive aspects of data analysis—helping find new customers or discover more productive logistics routes. But the reality is that big data analysis will find some negative things—about your sales team’s effectiveness, your online presence, your true costs of operations.

The slow economy of the last four years has weakened multiple parts of most companies. Adding data sources and a more holistic analysis will help find and prioritize the problems you need to fix.

**IT Truth Tellers**

Want to raise IT’s profile as a business enabling? Step in and assume responsibility for data quality across the company. Here’s a quick check of items IT should review today:

1. Is there a centralized data quality team? If not, set one up ASAP.
2. Does the team do regular or, at minimum, spot audits of various analyses? Does it regularly look to add new data sources?
3. Are critical external events annotated within your data warehouse or as part of your reporting process? For example, think about major system upgrades that would change the underlying data related to order flow.
4. Do you require statistical notes, including sampling statements?
5. When it comes to customer or vendor surveys, are sample sizes validated against your base customers or total market size to ensure accuracy?
6. Do you run regular “stress tests” for current data sets with cross-functional teams, challenging assumptions and sacred cows?
7. Do you look outward? Most respondents, 75%, report some public cloud use. Yet many companies aren’t capturing associated data—think WebEx conferencing for customer behavior analysis or Google analytics for sales tracking.

Michael Healey is the president of Yeoman Technology Group. Write to us at iwletters@techweb.com.