What is Data Decay, and How Much Does it Cost Your Business?

by James Miller - Tuesday, November 12, 2019

https://www.searchbug.com/info/what-is-data-decay/

If data decay sounds big and scary, that's because it is.

The reality is that no data remains correct forever. And, some data degrades faster than others. Furthermore, personal data is some of the most unstable data out there.

On average, Americans move 11.7 times in their lifetime. Over 37 million phone numbers are reassigned each year; some are reassigned in as little as a few days. And 60% of customers intentionally give false information—like fake email addresses—when they submit personal information online.

That's why customer data goes out of date so fast.

This is especially bad news for businesses because they rely on customer data for just about everything—marketing, lead generation, sales, customer support and service, and almost every other stage of business operation requires accurate customer data.

So, what exactly is data decay? How much does it cost companies? And, most importantly, how do you stop it?

What is data decay?

Data decay is also called data degradation, data rot, or bit rot.

Data decay simply the gradual loss of accurate data in a database.

There are two types of data decay: mechanical and logical.

Mechanical data decay

Mechanical data decay is the most straight forward. It happens when there's a hardware failure in the data storage system that causes data to be lost or corrupted. Hard drive failures are the most common form of mechanical data decay. But, corruption in data management software or failures in supporting hardware can also cause mechanical data decay.

Logical data decay

Logical data decay is harder to spot. It occurs when information in the database goes out of date or becomes incorrect because that data point has changed, but the database was not updated. This is the most common form of customer data decay, because customers usually won't tell you when they change their phone number or move.

There's something called the [1-10-100 rule](#) that estimates how much data decay costs businesses. On average, it costs $1 to validate data as it's entered, $10 to cleanse a bad record from your database, and $100 per incorrect customer that is never corrected.

The fact that there are three levels of cost tells you that both types of data decay are preventable with proper data cleaning and management.

Here's how to do that.

How to prevent data decay

Mechanical data decay is the easiest to prevent, and you're probably already doing it: backup your data. That way you can restore your data if it's lost or damaged.

Additionally, use solid state storage whenever possible. And, monitor your storage hardware. Replace mechanical drives well before they reach a critical failure point.

That's the easy part.

Preventing logical data decay is tougher, because you must identify entries that have decayed and update them, and remove entries that can't be corrected.

Logical data decay is an ongoing process. So, it must be prevented with ongoing data management and cleaning.

Data cleaning steps

At a high level, the data cleaning process looks like this:

1. Validate data as it enters your databases.
2. Perform regular maintenance data cleansing and updates while data is in storage.
3. Verify data before use.

The most important step is the first step—validating data as it's entered—because that's the most cost-efficient way to quality control your data.

Prioritize the second two steps based on how you use your data. If you often reference your stored data without deploying it directly, you may need more frequent maintenance cleaning cycles.

On the other hand, if you immediately use your data once it's collected, or if you know the data will sit, untouched, in storage for a long time before you use it, you can dial back your data maintenance processes in favor of on-demand data validation.

There are two ways to build a data verification process that prevents data decay: batch processing and data integration.

Batch processing

Batch processing is best for verifying data in bulk. If you need to validate single entries on demand, batch processing is too cumbersome. So, this is best for maintaining the integrity of your databases and updating existing data that hasn't been cleaned in a long time.

The batch data cleansing process is simple, but requires a bit of manual work:

1. Export your data to a .csv, .txt, or Excel file.
2. Remove duplicate entries and separate personal information into separate columns (i.e. a separate column for first name, last name, address, city, state, zip, etc).
3. Upload your data to your batch processing service. You can select which data you need cleaned and updated in the online user interface. Your results will usually be returned in minutes, depending on the file size.

ARE YOU MISSING PERSONAL CONTACT INFORMATION FOR YOUR CONSUMER OR BUSINESS CUSTOMER RECORDS?

This do it yourself data append tool makes it quick and easy to update your list of database records. Find names, addresses, phone numbers and email addresses.

THERE ARE NO MINIMUM OR MAXIMUM NUMBER OF RECORDS. PROCESS JUST A FEW OR A MILLION RECORDS.

UPLOAD TXT, CSV OR EXCEL FILE TO GET STARTED

| CHOOSE FILE | No file chosen | Enter E-mail Address | | CONTINUE |

4. Reload your data into your database, ready for use.

Batch data cleansing is easy. However, data integration is better for real-time data validation, since it removes the manual labor.

Data integration

Data integration uses an API to add data validation to your data pipeline, and automates the process of uploading data for verification.

Typically, a data verification API integrates through a unique URL. So, there are no API keys to worry about. And, most CRM and data management systems have an interface for adding API integrations. However, if you need your tech team to do it, it will take them a matter of minutes to add API connectivity.

Then, your data management systems will automatically send data out for validation on-demand, with no

effort from you or your teams.

There are two places where data integration will deliver the most benefit:

1. Data entry forms.
2. Systems where the data is used (such as a CRM system, where sales teams retrieve contacts for customer outreach, or a shipping system).

Adding data validation to your data entry forms where customers enter their information will filter out almost all invalid data.

So, why validate the data again when you use it? For a couple of reasons:

- Data entry errors are the number one source of bad data. New errors get introduced whenever your data is manually transferred in transit from collection to deployment. It's unavoidable.
- You may need to append additional information to build complete profiles for the teams that actually use the data.For example, you may collect only a name and phone number on a lead generation form to improve form completion rates. But, the mailing address may be valuable for sales teams that can use geographic data for more personalized sales outreach.Or, email addresses can be appended to contacts for more customer communication channels.

Data integration is both a database quality control measure and a conversion optimization. And, it enables you to validate data at the point of entry, where it will cost you a mere $1 to verify, according to the 1-10-100 rule.

But, whether you use batch data processing or data integration, you'll avoid the $100 price tag of each inaccurate entry in your database, which is a significant reduction in operating costs.

Really, you can't afford *not* to validate your data.

If you're ready to start working with dependable data, check out our batch processing services and data validation APIs that can verify any combination of personal information.

Searchbug.com