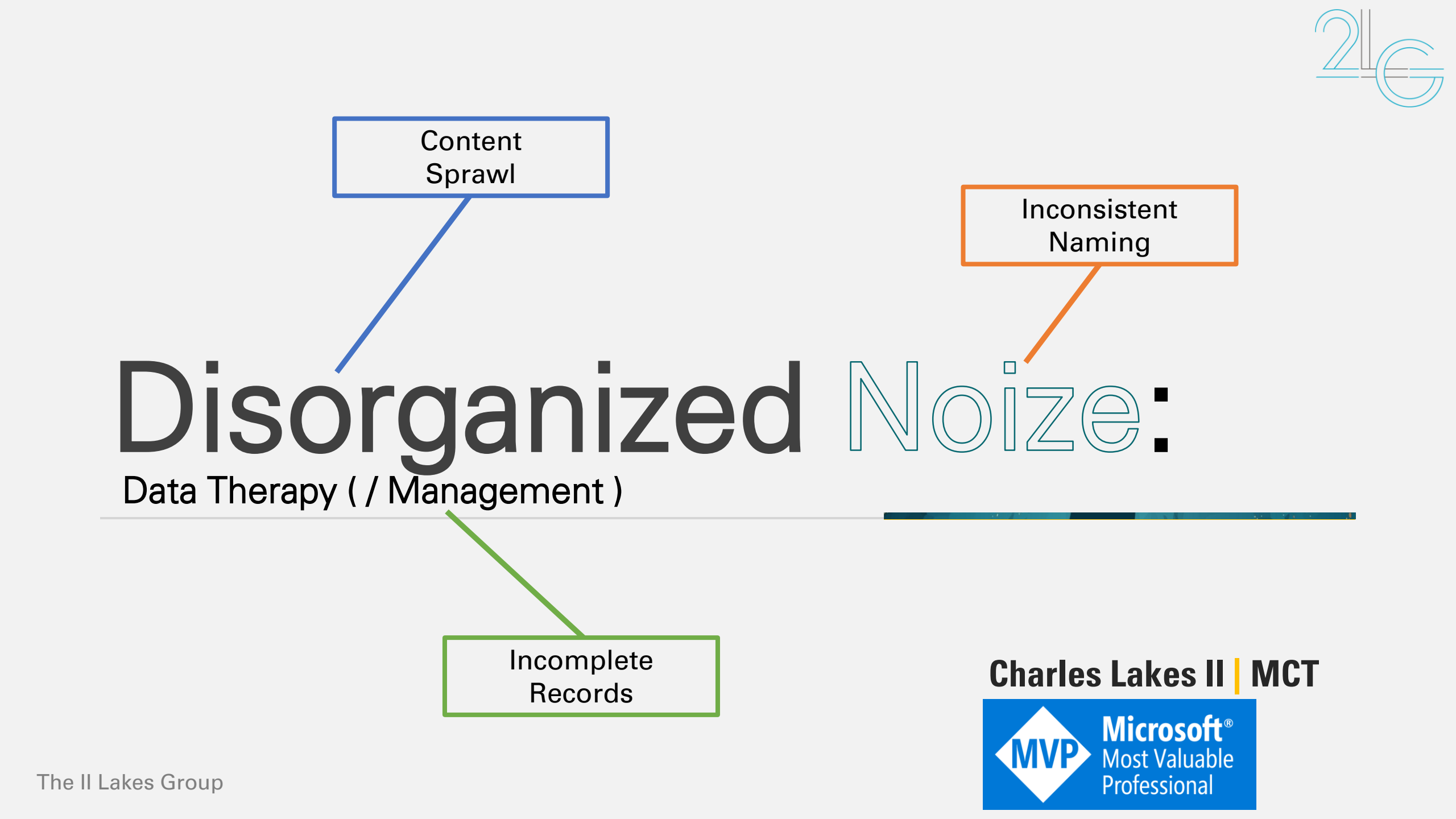
Disorganized Noize:

Data Therapy (/ Management)

Charles Lakes II | MCT



Charles Lakes II | MCT

Microsoft® MVP Most Valuable Professional

About Me:

- Software Engineer.
- Community Speaker.
- M365 Blogger.
- Middle-child.

Agenda:

Leadership.

Power users.

Developers.

Admins.

Data Management:

“. . . is the process of ingesting, storing, organizing and maintaining the data created and collected by an organization. Effective data management in IT systems is crucial to running business operations and delivering information that helps drive decision-making by corporate executives, business managers and other end users.

The data management process includes different functions that collectively aim to make data accurate, available and accessible.”

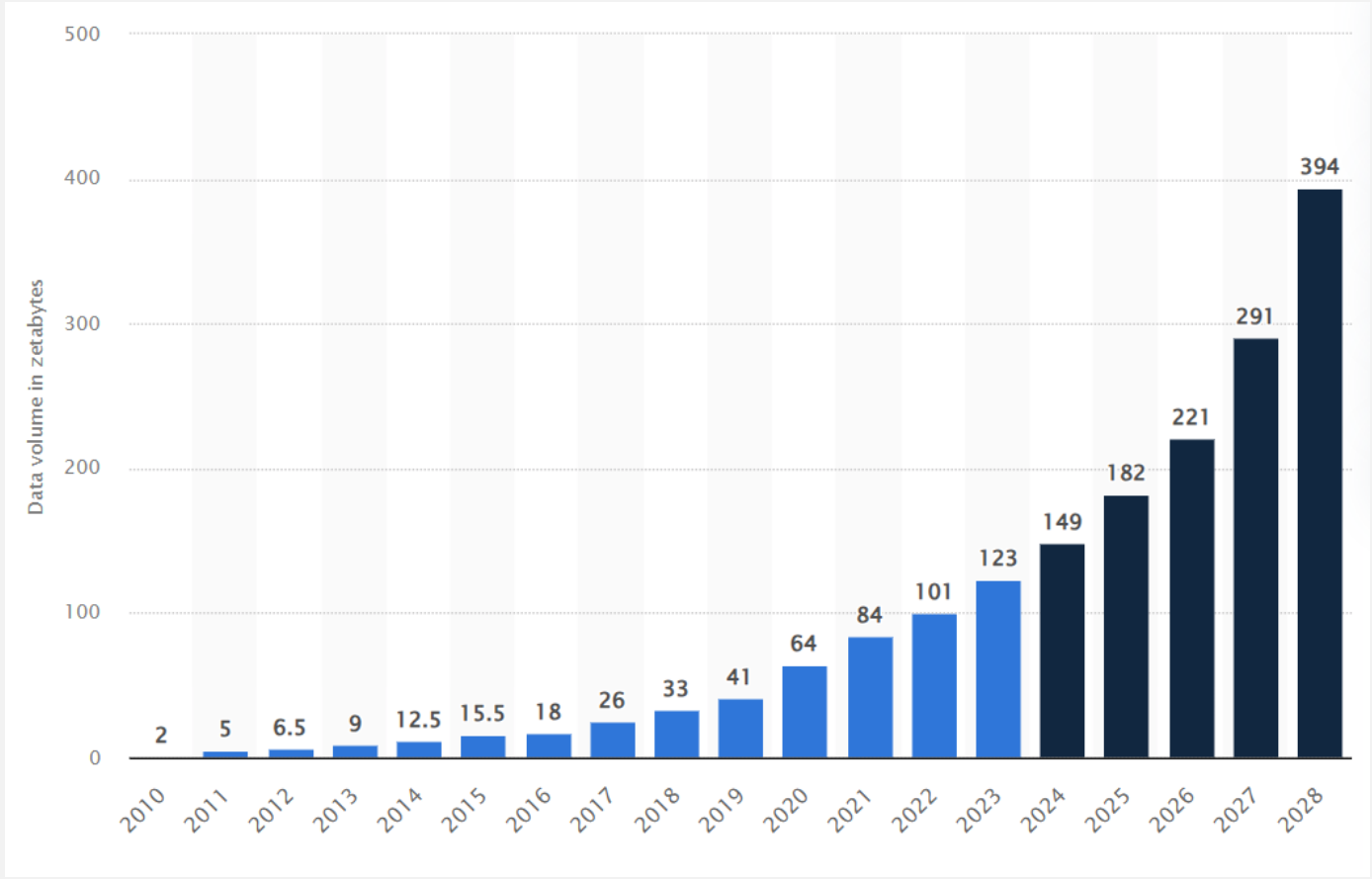
- TechTarget

Admins

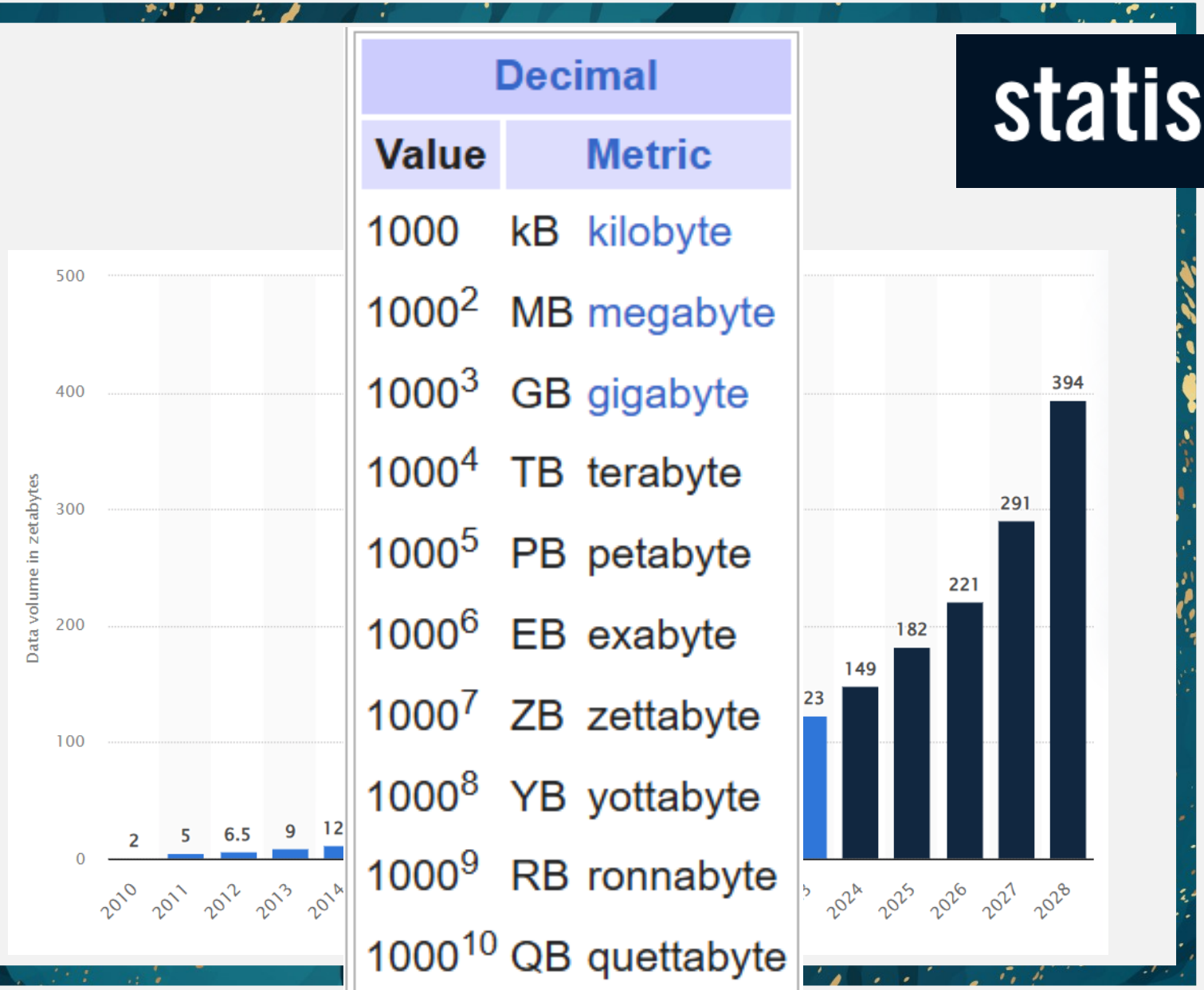
- DBAs
- Architects
- Engineers

Admins

statista



- ❑ DBAs
- ❑ Architects
- ❑ Engineers



Decimal		
Value	**Metric**	
1000	kB	kilobyte
1000^2	MB	megabyte
1000^3	GB	gigabyte
1000^4	TB	terabyte
1000^5	PB	petabyte
1000^6	EB	exabyte
1000^7	ZB	zettabyte
1000^8	YB	yottabyte
1000^9	RB	ronnabyte
1000^{10}	QB	quettabyte

statista

- ❑ DBAs
- ❑ Architects
- ❑ Engineers

Choose where to apply this policy

The policy will apply to content that's stored in the locations you choose.

You can set up data connectors to import content from non-Microsoft apps like Slack, WhatsApp and many more, for use with this solution. Set up now

Status	Location	Applicable Content	Included	Excluded
On	Exchange mailboxes	Items in user, shared, and resource mailboxes: emails, calendar items with an end date, notes, tasks with an end date, and Public folders. Doesn't apply to items in Microsoft 365 Group mailboxes. More details	All mailboxes Edit	None Edit
On	SharePoint classic and communication sites	Files in classic sites or communication sites or team sites that aren't connected to a Microsoft 365 group, and files in all document libraries (including default ones like Site Assets). More details	All sites Edit	None Edit
On	OneDrive accounts	All files in users' OneDrive accounts. More details	All user accounts Edit	None Edit

Manage archived sites

Archived sites can be reactivated or deleted. Deletion of archived sites follows the same behavior as that of active sites; that is, a site doesn't need to be reactivated before being deleted. However, sites in the "Reactivating" state can't be deleted until reactivation completes.

Admins can view details of the site, such as the URL, Archive Status, or Storage, from the **Archived sites** page.

Developers

- ❑ Frontend
- ❑ Backend
- ❑ Desktop
- ❑ Mobile

Developers

Medium

In the era of big data, reliable and accurate data is the backbone of any successful application. Whether it's powering machine learning models, driving business insights, or enabling seamless user experiences, the quality of data plays a critical role. However, data engineers often face numerous challenges related to data quality, including missing, incomplete, incorrect, or outdated data. These issues can significantly impact the performance and reliability of data-driven applications. In this article, we'll explore these common data challenges and discuss effective strategies to address them, ensuring robust and trustworthy data pipelines.

- ❑ Frontend
- ❑ Backend
- ❑ Desktop
- ❑ Mobile

Solutions and Best Practices

To address missing data, data engineers can implement several strategies:

Imputation Techniques:

- **Mean/Median Imputation:** Replace missing values with the mean or median of the column.
- **Mode Imputation:** Use the most frequent value in the column for categorical data.
- **Predictive Imputation:** Utilize algorithms to predict and fill in missing values based on other data points.

Type	Required
Single line of text	Hidden
File	Yes
Multiple lines of text	No
Choice	No
Single line of text	No
Single line of text	No
Single line of text	No
Single line of text	No

Power Users

- Analysts
- Accountants
- Auditors

Power Users

Forbes

Numerous studies of "knowledge worker" productivity have shown that we spend too much time gathering information instead of analyzing it. In 2001, IDC published its venerable white paper, "The High Cost of Not Finding Information," noting that knowledge workers were spending two and a half hours a day searching for information.

Since then, we have seen the rise of the cloud, ubiquitous computing, connectivity and everything else that was science fiction when we were kids becoming a reality — including the imminent emergence of AI. Yet in 2012, a decade after the IDC report, a study conducted by McKinsey found that knowledge workers still spend 19% of their time searching for and gathering information, and a 2018 IDC study found that "data professionals are losing 50% of their time every week" — 30% searching for, governing and preparing data plus 20% duplicating work.

Clearly, all the technology advances have not flipped the productivity paradigm; it seems like we still spend more time searching for information that exists rather than analyzing and creating new knowledge.

- ❑ Analysts
- ❑ Accountants
- ❑ Auditors

Create content type

Search content types

All Items

Site content type	Parent	Category
Announcement	Item	List Content Types
Basic Page	Document	Document Content Types
Circulation	Item	Group Work Content Types
Comment	Item	List Content Types
Contact	Item	List Content Types
Content Freshness	Site Page	Document Content Types
Discussion	Folder	Folder Content Types
Document	Item	Document Content Types
Document Set	Document Collection Folder	Document Set Content Types
Dublin Core Columns	Document	Document Content Types
East Asia Contact	Item	List Content Types
Event	Item	List Content Types
Folder	Item	Folder Content Types

Power Users

Database normalization is the process of structuring a relational database in accordance with a series of so-called ***normal forms*** in order to reduce data redundancy and improve data integrity. It was first proposed by British computer scientist Edgar F. Codd as part of his relational model.

Normalization entails organizing the columns (attributes) and tables (relations) of a database to ensure that their dependencies are properly enforced by database integrity constraints. It is accomplished by applying some formal rules either by a process of *synthesis* (creating a new database design) or *decomposition* (improving an existing database design).

Leadership

- ❑ C-Suite
- ❑ Directors
- ❑ Managers
- ❑ Supervisors

Leadership

Harvard Business School Online

As data becomes increasingly available, businesses that use it to drive decision-making are reaping the benefits. According to McKinsey (pdf), data-driven organizations are 23 times more likely to outperform competitors in customer acquisition, nine times more likely to retain customers, and up to 19 times more profitable.

With so much hinging on the power of data, the importance of data integrity can't be overstated. One error in a dataset can have a ripple effect and impact your business's most vital decisions. So, what is data integrity? What can you do to combat threats and maintain datasets' integrity for the benefit of your organization, data subjects, and customers?

- ❑ C-Suite
- ❑ Directors
- ❑ Managers
- ❑ Supervisors

Recap:

Leadership.

Power users.

Developers.

Admins.

Contact:

/charles-e-lakes-ii

@The2Lakes

/losodamus

Charles is a technical speaker and low-code advocate with a pro-code background, specializing in transforming business data architecture. Passionate about innovation, he shares his expertise to solve complex business problems and guide organizations in reimagining their legacy processes, advancing automation initiatives, streamlining operations, and reducing infrastructure costs.