



How Many of These Data Management Industry Buzzwords Do You Know?

DATA LAKE — A centralized repository designed to store, process and secure large amounts of structured, semi-structured and unstructured data. It can store data in its native format and process any variety of it, ignoring size limits. –Source: Google

What's the buzz about a data lake (sometimes called a data swamp)? Data lakes are implemented everywhere. But without proper governance and management, they're very costly, can pose a privacy risk and ultimately fail to meet expectations.

DATA LAKEHOUSE — A hybrid architecture approach that leverages the best from the data lake, and the best from the data warehouse, while attempting to eliminate the worst. A lakehouse handles structured and unstructured data and leverages intelligent metadata layers to categorize and classify the data so that unstructured data can appear structured. –Bernard Marr

What's the buzz about a data lakehouse? They're almost at the top of the "hype cycle" and will/should be a conversation in your company. Data lakehouses are a new capability that will stretch your organization's thinking around how to govern it. And even a lakehouse needs to be cleaned, so it's crucial to truly understand what's automated.

DATA MESH — An architectural concept that addresses the challenges of the next-generation data lake, founded on four principles — distributed, interoperable, atomic and domain-centric. An infrastructure that enables the idea of Data as a Product by reinforcing the idea of an ecosystem of atomic parts and processes. –Zhamak Dehghani

What's the buzz about data mesh? Although a data mesh is created to serve analytics, the construct is also relevant to operational reporting and data sharing. It could add confusion because it stakes a position on data domain management, data sharing, data governance, data enablement and other areas. A data mesh focuses on decentralization and distribution of responsibility to people who are closest to the data to support continuous change and scalability. This means that accountability for the business process, the data and the data about the process are all the same.

DATA FABRIC — A design concept that serves as an integrated layer (fabric) of data and connecting processes. A data fabric utilizes continuous analytics over existing, discoverable and inferred metadata assets to support the design, deployment and utilization of integrated and reusable data across all environments, including hybrid and multi-cloud platforms. –Gartner

What's the buzz about data fabric? The quality of the metadata — both business and technical — will determine the effectiveness of implementing a data fabric.

DATA AS A PRODUCT — A solution delivery concept that seeks to deliver a reusable data asset, engineered to deliver a trusted dataset for a specific purpose. It integrates data from relevant source systems, processes it, ensures it's compliant, and makes it instantly accessible to anyone with the right credentials. –FSFP

What's the buzz about data as a product? It's shifting the data design and delivery approach to a product management mindset, and there are best practices to adopt when making this shift.

DATA MARKETPLACE — Ecosystems centered around data assets that provide infrastructure, transactional capabilities and services for participants. Marketplaces can be either internally or externally facing to facilitate data sharing and monetization. –*Gartner*

What's the buzz about a data marketplace? They're a great way to enable self-service and measurably improve organizational efficiency. Most common in large organizations, data marketplaces are vital mechanisms to distribute data as a product and make finding and consuming organizational data as easy as shopping on Amazon.

DATA DISCOVERY — The collection and pattern analysis of data from multiple sources with the goal of identifying previously unknown patterns. Commonly used to identify data holding specific classifications, such as personally identifiable information (PII) or protected health information (PHI). –*FSFP*

What's the buzz about data discovery? It can accelerate understanding of “what data is where,” rather than attempting to do a manual data inventory.

DATA ENABLEMENT — The practice of empowering individuals in a business with the support and tools they need to responsibly leverage trusted data to achieve real business outcomes. –*DATAVERSITY*

What's the buzz about data enablement? It's a capability specifically focused on data usage and what is needed to achieve business outcomes. It can be confusing determining the difference between data enablement, data literacy, data governance and organizational change management (OCM).

DATA LITERACY — The capability that assists individuals in making data-driven decisions by enabling and empowering everyone to create, locate, understand, use and explain high-quality, governed data. –*FSFP*

What's the buzz about data literacy? It pulls together training, communication, governance, metadata management and data quality in a way that caters to and supports data users to accomplish business goals. Many governance programs have been “doing data literacy” without calling it that.

ORGANIZATIONAL CHANGE MANAGEMENT (OCM) — The discipline of driving business results by changing behavior. This includes helping people through new programs, processes or paradigms quickly and successfully to achieve business value. –*FSFP*

What's the buzz about OCM? Success rates for complicated programs are directly proportional to the quality of the change management efforts expended (Prosci Best Practices Reports since the 1990s). Proactive OCM should be a part of any data initiative.

SYNTHETIC DATA — Information that is artificially manufactured rather than generated by real-world events. Synthetic data is created algorithmically, and it is used as a stand-in for test datasets of production or operational data to validate mathematical models and, increasingly, to train machine learning models. –*Tech Target*

What's the buzz about synthetic data? It's a cost-efficient, inexpensive alternative to real-world data that can be optimized by industry, geography or other criteria. Ensuring ethical and governance standards are used to create synthetic data is essential.

DATA-DRIVEN — Overall need to use the organization's data to drive effective decision-making. –*FSFP*

What's the buzz about data-driven? There's a tremendous amount of work to do to enable a company to be data-driven — from architecture and technology to ownership and accountability, as well as a significant amount of organizational change management. Being data-driven has quantifiable business benefits.

DIGITAL TRANSFORMATION — An enterprise initiative to address an overall need to automate or digitize business processes throughout the organization. These strategic evolutions create exponentially more data. –*FSFP*

What's the buzz about digital transformation? It's an important driver for data governance and data management activities. If your company is launching a digital transformation and you have yet to stand up a data governance capability, now is the time to get started.



DATA TRASH — Seemingly important data that we keep collecting and supporting that then is never used, just takes up server space for no good. It tends to accumulate on the cloud when an organization has limitless storage capacity. –*Towards Data Science*

What's the buzz about data trash? This data could be a Pandora's box of problems if you don't have a handle on what it is, how it is protected and who should have access to it. Any wasted cost related to data is too much.

DATA DEMOCRATIZATION — The ongoing process of enabling everybody in an organization, irrespective of their technical know-how, to work with data comfortably, to feel confident talking about it, and as a result, make data-informed decisions and build customer experiences powered by data. –*Towards Data Science*

What's the buzz about data democratization? Seeking it influences the decisions around data access and sharing to be more generous than those organizations with a greater focus on data protection. Data literacy will improve the positive impact of data democratization through improved data understanding and more productive use of data.

DATA OPERATIONS — A group of support functions that manage the set of business and technical activities and processes that enable data systems to run effectively and efficiently. Although typically considered to be an IT function, data operations roles are also needed in the business. –*FSFP*

What's the buzz about data operations? Data ops is a critical capability to support data growth and effective management throughout an enterprise. Incorporating it into your overall data governance and management strategy will help to ensure sustainability.

DATA QUALITY — The practice of defining expectations of data, monitoring for conformance to expectations, and correcting non-conformance. Corrections will be both preventative and reactive. –*FSFP*

What's the buzz about data quality? It can be performed in a variety of systems across a data lifecycle, by a variety of roles, for a variety of reasons. Data quality as an enterprise capability will drive greater economies of scale, consistency and harmonization of data and overall trust.

DATA OBSERVABILITY — An extension of data quality that broadens the scope of monitoring to provide visibility into the flow of data through the organization's platform(s) and may incorporate machine learning for automation of issue detection and resolution as well as issue prevention. –*FSFP*

What's the buzz about data observability? The ability to assess and understand data's quality at rest and in motion is critical to ensure that data governance and management standards are followed. Leveraging greater automation will increase an enterprise's ability to scale best practices around data.

DATA PREPARATION — The process of preparing raw data so it is suitable for further processing and analysis. Key steps include collecting, cleaning and labeling raw data so it can be used for a variety of downstream purposes, including machine learning and data visualization. –*Amazon*

What's the buzz about data preparation? Because these steps could include data profiling, cleaning, validation, integration, enrichment, transformation and noise reduction, this is a great place to embed data standards and governance into the process. Data preparation is also a great opportunity to capture the results, transformations and metadata in a shared repository to advance broader data understanding and trust.

DATA WRANGLING — Occurring after data preparation and preprocessing, this practice employs a variety of techniques to transform the raw data into readily used formats. It is commonly employed when making the machine learning models and involves cleaning the raw dataset into a format compatible with the machine learning models. –*HBS Online*

What's the buzz about data wrangling? Like data preparation, this is an opportunity to embed standards and governance and glean necessary metadata that can enhance enterprise data understanding.



DATA ENGINEERING — The practice of designing and building systems for collecting, storing and analyzing data at scale. A critical component of creating and optimizing a data pipeline for analytic or operational purposes. –*Coursera*

What's the buzz about data engineering? As data volumes continue to grow, so do specializations to optimize data pipelines. Data engineers are strategic members of the data organization that work with data architects, scientists and stewards to ensure data availability.

DATA DEBT — The future cost of “bad data” resulting from the deference of correct behavior with data assets. –*John Ladley*

What's the buzz about data debt? This quantifiable amount is a valuable business case component for data governance and/or improved data management. This financial metric illuminates the need to govern and manage data well in the moment, not after the decision or project. Like any debt, this future obligation gets bigger with time until it is paid or written off.

DATA GRAVITY — The idea that data has mass and, like a planet, the larger that mass becomes, the stronger the gravitational pull of other systems and applications towards it. –*Dave McCrory*

What's the buzz about data gravity? The effect of data gravity impacts data location and storage to minimize workload performance issues. This, in turn, impacts your data and system architecture, your adoption of the cloud and decisions of which data resides where, how it is protected, how it is accessed, etc.

DATAT GOVERNANCE — The organizing framework to ensure that data assets support the strategy of the enterprise and are of high quality, meaning they are consistent across systems, maintain integrity, are auditable and complete, are unique and not duplicated, and are well defined, understood and appropriately available. –*FSFP*

What's the buzz about data governance? This organizational function sets the direction for how data will be managed and provides support for the development of better data management processes. Data governance is a foundational capability to ensure your data promotes organizational goals, such as growth, profitability and risk management.

B	I	N	G	O
Organizational Change Management	Data Quality	Data Preparation	Data Literacy	Data Governance
Data Trash	Data Fabric	Data Observability	Synthetic Data	Data Lake
Data Lakehouse	Data Enablement	FREE SPACE	Data as a Product	Data Wrangling
Data-Driven	Data Gravity	Data Discovery	Digital Transformation	Data Democratization
Data Marketplace	Data Engineering	Data Operations	Data Debt	Data Mesh

