Axis 4: Data management: Big Data everywhere

In the digital era, data management and the effective use of Big Data have become critical components of organizational success. The fourth axis of digital transformation focuses on how well an organization captures, stores, processes and utilizes vast amounts of data to enhance decision-making, improve operational efficiencies and drive innovation. This axis examines the infrastructure, strategies and competencies that organizations deploy to handle the complexities of Big Data.

Data Management

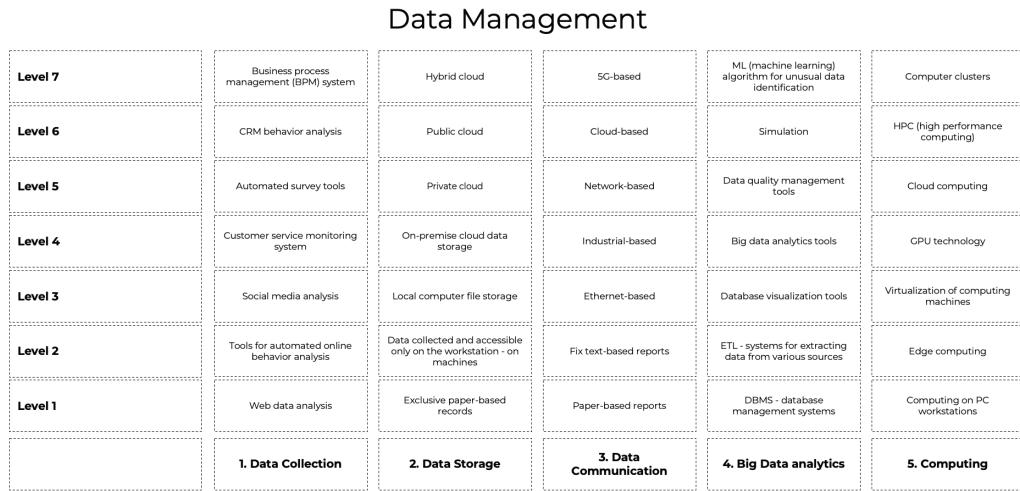|	1. Data Collection	2. Data Storage	3. Data Communication	4. Big Data analytics	5. Computing
Level 7	Business process management (BPM) system	Hybrid cloud	5G-based	ML (machine learning) algorithm for unusual data identification	Computer clusters
Level 6	CRM behavior analysis	Public cloud	Cloud-based	Simulation	HPC (high performance computing)
Level 5	Automated survey tools	Private cloud	Network-based	Data quality management tools	Cloud computing
Level 4	Customer service monitoring system	On-premise cloud data storage	Industrial-based	Big data analytics tools	GPU technology
Level 3	Social media analysis	Local computer file storage	Ethernet-based	Database visualization tools	Virtualization of computing machines
Level 2	Tools for automated online behavior analysis	Data collected and accessible only on the workstation - on machines	Fix text-based reports	ETL - systems for extracting data from various sources	Edge computing
Level 1	Web data analysis	Exclusive paper-based records	Paper-based reports	DBMS - database management systems	Computing on PC workstations

Figure 14

As we explore this axis, we investigate the various facets of data management that are essential for building a strong digital ecosystem. From collecting data and ensuring their secure storage to conducting advanced analytics and deriving actionable insights, each component is crucial in turning raw data into strategic assets. This axis encompasses not only the technical skills required to handle Big Data but also the cultural transformations needed to cultivate a data-centric approach throughout the organization. A seven-tier maturity model has been proposed to assess each area, reflecting the broad range of technical capabilities present in these fields.

Area 4A. Data collection

This is a complex process involving various stages such as identifying the sources of data and collecting, validating and evaluating those data. This requires the right tools and technologies to ensure not only the high quality and reliability of the process but also the security and confidentiality of the gathered data. Another important component of the data collection system is its assessment and control of data quality, including of accuracy, completeness and consistency. It is also valuable to provide suitable procedures and tools for monitoring and analysing the data to understand organizational processes better and make decisions more effectively.

Level 1. Manual data collection. This involves employees manually completing forms with information about their work or other activities related to the organization. These forms typically include information about tasks performed, working hours, equipment used, materials and tools employed by the employee, and work-related problems or incidents. The downside of manual declarations is that they are time-consuming and costly, requiring manual filling of documents by employees. They are also often prone to human error, which can degrade data quality and complicate analysis. An example of this level is a construction company using paper cards to track working hours and equipment usage. While a simple task, errors in records could lead to inaccuracies in project management.

Level 2. The use of graphic symbols. Such techniques as the use of barcodes or QR codes are popularly employed in warehouses, stores or production to identify goods, products or devices using a scanner. Each product or device is graphically labelled, and its data are stored in an IT system. Using a scanner, an employee can quickly and precisely gather information about products or devices to ensure their smooth operation or to assist customers. In a large supermarket network, each product on the shelf has a barcode. Employees use scanners for quick inventory counting and receiving deliveries, thereby increasing efficiency and reducing errors.

Level 3. Advanced technological solutions. Technologies such as Radio Frequency Identification (RFID) streamline operations and processes.

RFID enables the automatic identification and tracking of objects using radio waves. In manufacturing, this technology serves as a data collection system, significantly improving warehouse management, logistics and the monitoring of production processes. For example, in an automotive parts factory, the use of RFID tags is invaluable. The tags are used to track boxes containing car components. With this system, the organization can precisely monitor inventory availability, allowing for more efficient management. RFID technology not only improves efficiency but also enhances monitoring accuracy, eliminating human errors and providing better control over the organization`s resources.

Level 4. Data from manufacturing machines and devices. A key source of information is machines and devices equipped with various sensors that measure a range of process parameters, such as temperature, humidity, pressure, speed, rotations, vibrations and more. These data are collected in real-time and processed in IT systems, enabling continuous monitoring and analysis of production processes. For example, beverage factories may use sensors to monitor temperature and pressure in the pasteurization process. This data is analysed in real-time by the system, which can adjust pasteurization parameters on the fly. This ensures consistent product quality and minimizes the risk of failures.

Level 5. Mobile data collection devices integrated into the production management and optimization process. Crucial information can be gathered regarding employees` interactions with mobile applications and devices used to monitor their work. These data may include information about employees` location, behaviours and preferences, and other relevant factors. An example is the use of smartphones or specially adapted devices that allow employees to be located within the production plant and data related to their work to be collected. For instance, these tools can monitor time spent at specific machines, work efficiency and breaks taken. With these tools, manufacturing companies can refine their operations and effectively respond to changing production conditions.

Level 6. Production data collected from physical objects. Modern sensors are increasingly recording various environmental parameters such as

temperature, humidity, pressure, light intensity or motion. The data that are collected are extremely valuable, as they can be used for various purposes such as monitoring working conditions, performing machine diagnostics and optimizing production processes. An excellent example is that of a food production factory using temperature and humidity sensors in its warehouse. These sensors constantly monitor product storage conditions, and the data they collect are transmitted to the appropriate software. This allows the company to meticulously track the quality of stored products. In case of any irregularities, such as inappropriate storage conditions, the software enables swift intervention, minimizing losses and ensuring high product quality.

Level 7. Optical control. This advanced approach to collecting data from physical objects uses cameras and other optical devices to monitor and assess the state of physical objects. Optical devices capture images of objects, which are then meticulously analysed for potential damages or other irregularities. Thus, cameras might be installed on the ceiling of a production or warehouse hall for real-time monitoring and identification of the activities that equipment is performing. These advanced cameras, equipped with sensors and image-processing algorithms, track objects on the production line, identifying their position and actions. This intelligent solution integrates visualization, data analysis and monitoring to create a more efficient and secure production environment.

Area 4B. Data storage methodology

This assessment area focuses on how an organization manages its data, encompassing its collection, storage, organization and sharing processes. It involves understanding the types of data collected, their update frequency, and security needs. The selection of appropriate technologies such as databases, cloud solutions and storage networks is crucial, along with strategies for backup and archiving. The performance, availability and scalability of these systems must be defined to ensure that data are accessible and processed efficiently. Additionally, data reliability and security must be ensured to prevent loss and protect privacy. These considerations are critical in designing and implementing a robust data management system that delivers maximum organizational benefit.

Level 1. Traditional data storage methods. These rely on physically recording information on paper. Data are printed and archived as hard copies in offices or specialized archives. This approach is typical for smaller firms lacking extensive IT systems and databases. A typical example is a notary`s office, which stores the legal documents of its clients as hard copies in binders and cabinets. When clients need access to their records, office staff must manually search through the paper archives, which can be time-consuming. In this case, paper-based storage is dictated by legal requirements.

Level 2. Single local digital storage. Information is held on a single device or workstation, which limits data access to that device only, meaning the information is not available to other users or systems. This approach is often employed in machine control systems, where data generated by the machine are stored on a local control device. An example would be a manufacturing plant where a machine control system for metal-cutting stores all its data on a local computer only. Engineers working in different parts of the factory cannot easily access these data, hindering the analysis and optimization of the production process. When there is a need to share this information with other employees or systems, data must be painstakingly copied manually.

Level 3. Dispersed local digital storage. Various computers or devices are commonly used for data storage in small businesses or by individual users.

Data are saved on the hard drives of individual computers, resulting in a lack of centralization in data storage. This approach carries some risk related to data protection, as the failure or loss of one computer can lead to the loss of important information. Additionally, it complicates collaboration and data sharing among employees and can be an obstacle to the analysis and utilization of data for business purposes. An example is a small accounting firm where each employee stores client data on their personal computer. If one of these computers fails or is stolen, there is a risk of valuable data being lost. Furthermore, the lack of a central database hinders effective collaboration among employees, as they do not have easy access to shared information.

Level 4. Local cloud storage. A company maintains and manages cloud infrastructure in its own data centre. This allows the organization to retain full control over its data and ensure a high level of security, as access to data is limited to authorized personnel. However, this method requires significant investments in infrastructure and human resources for managing and maintaining the local cloud. The example here is a manufacturing company that uses its server space to store and manage data related to quality control and safety. Although the company has full control over its data and protects it from unauthorized access, it bears the substantial costs of purchasing, maintaining and updating the infrastructure.

Level 5. Public cloud storage. This involves using the data-processing services of external cloud service providers. These providers maintain cloud infrastructure, which their clients access via the Internet. Public clouds offer a wide range of services, including data processing, data storage, virtual machines and tools for data analysis. A company that uses the services of providers such as Google Cloud, Dropbox or AWS is operating at this level. Employing a public cloud reduces costs, allows resources to be scaled, and provides flexibility in managing data and applications.

Level 6. Private cloud storage. Cloud infrastructure is dedicated to a single organization. This means that cloud resources such as servers, storage and network are available only to the employees of that specific organization, and not to external users. An example is a financial corporation such as a bank

using a private cloud to securely store customer data. This ensures full control over data security, a critical aspect in the financial sector.

Level 7. Hybrid approaches. Companies strategically manage their data storage by utilizing a mix of solutions tailored to the specific needs of different types of data. For instance, operational data critical for production processes are stored on edge devices for immediate access, while sensitive information like technical drawings is kept in private clouds to enhance security. Other data, such as financial and material reports, benefit from the scalability and accessibility of public clouds, allowing for resource adjustments in response to fluctuating demands. This approach not only optimizes performance and cost-efficiency but also ensures robust security and control, exemplified by an automotive component manufacturer who effectively segregates data storage based on function and sensitivity.

Area 4C. Data communication

A data communication system comprises methods and tools for transmitting information between different devices, applications and individuals. Fundamental components of the data communication system include communication protocols, computer networks, network devices, network management software and tools for monitoring and analysing network traffic.

Level 1. Hard-copy reporting. Traditional paper reports are the primary means of communication and data transmission. This classic method involves printing data out and passing them between employees or sending them outside the company. An example might be a financial industry company still using traditional paper reports for internal and external communication. This is a costly and time-consuming approach with negative environmental impacts.

Level 2. Reporting via email. This is a fast and convenient method of information transfer, facilitating easy document sharing among employees and ensuring easy access. However, it comes with data security risks, such as the possibility of data interception by unauthorized individuals or email system failures. It is thus essential to ensure adequate security measures, such as data encryption, confidentiality regulations and backup procedures. Let us take the example of a consulting company that relies on email to ensure the rapid transfer of essential reports and documents between its teams. The company, acutely aware of the security risks associated with data, implements advanced data encryption technologies and strictly adheres to confidentiality regulations.

Level 3. An Ethernet-based architecture. This is one of the most popular ways of building computer networks, allowing data transmission between devices using an Ethernet cable. Ethernet-based architecture is applied in Local Area Networks (LANs), which are known for their high performance and operational reliability.

Level 4. An industrial Ethernet-based architecture. This specialized network configuration is designed for industrial applications and has higher reliability, security and data transmission speed than traditional Ethernet networks. Industrial networks employ special network devices such as

switches, routers and gateways that are resistant to industrial conditions like vibrations, dust and humidity. Additionally, specific communication protocols like Modbus, Profibus and DeviceNet enable communication between various industrial devices.

Level 5. A wireless communication architecture. Devices in this network connect to one another via a central access point, facilitating centralized network management. Wireless networks find applications in various industries, including manufacturing, transportation, medicine and services.

Level 6. A WAN/LAN architecture. Here, devices like routers, switches and servers connect multiple computers and devices into a network. WAN/LAN networks provide a stable connection between company branches and offices, enabling access to data and applications from different devices and locations. Various protocols and technologies such as TCP/IP, VPN and MPLS ensure fast and secure data exchange within the organization. In particular, VPN (Virtual Private Network) technology ensures secure communication between locations, which is crucial for safeguarding sensitive company data.

Level 7. A cloud-based architecture. Applications, data and resources are stored and processed in the cloud, which is usually provided by an external vendor. A cloud-based architecture allows access to applications and data from anywhere and any Internet-enabled device, allowing resources to be scaled flexibly according to needs. This enables organizations to adapt more easily to changing business requirements and reduce the costs of maintaining infrastructure. An example of cloud-based architecture is a microservices architecture, where applications are built as a set of smaller, independent services that can be easily scaled and updated.

Area 4D. Big Data analysis

Big Data analysis is a critical component of digital transformation, as it relates to leveraging vast datasets to extract valuable insights. This process employs advanced technologies and methodologies designed to efficiently analyse large volumes of data. By utilizing Big Data analysis tools, organizations can unlock profound insights into customer behaviour, emerging market trends, and intricate details of internal operations. These insights are pivotal in making informed decisions that can significantly enhance business strategies, operational efficiency and competitive advantage. Big Data analysis not only informs tactical moves but also supports strategic planning by providing a clearer picture of the market landscape and organizational performance.

Level 1. DBMSs. DataBase Management Systems facilitate effective data management by allowing large datasets to be created, stored, updated and analysed, giving them a crucial role in data collection and organization. DBMSs have applications in various fields, from business to science and engineering. MySQL, a popular solution in this category, is frequently used in web development, serving as an engine for storing and managing user data on products, transactions, and other elements that contribute to the smooth operation of websites and online applications.

Level 2. ETL systems. Extract, Transform, Load systems perform a key process in data management to enable the collection, processing and loading of information from various sources into a single database. ETL systems allow efficient data acquisition from diverse sources such as databases, CSV files, spreadsheets or XML files. Subsequently, these data are processed to convert them to a consistent format and eliminate any errors or inconsistencies. Finally, the data are loaded into the target database, where they are ready for analysis and use.

Level 3. Tools for visualizing database information. Data are far better comprehended when presented in an effective graphical form. These tools enable users, including managers and business analysts, to quickly create various visualizations like charts, tables and heatmaps that improve our understanding of the information stored in databases. These visualizations

help users detect patterns, analyse trends and identify significant relationships in data. For instance, a marketing department manager may use a database visualization tool to create an interactive chart presenting the effectiveness of different advertising campaigns in various regions. This allows for a quick assessment of which campaign works best so that the marketing strategy can be adjusted.

Level 4. Tools for analysing large databases. A crucial element in Big Data analysis, these are advanced applications and platforms, such as Apache, that enable the processing and analysis of massive amounts of data in a distributed manner. This means that data are processed on multiple servers simultaneously, ensuring scalability and significant computing performance. An example might be an e-commerce company employing Apache Spark to analyse vast sets of transactional data. The tools help a company identify customer purchasing patterns, analyse trends and optimize marketing strategies. These tools provide a deeper understanding of customer behaviour and better adaptation of the product offering.

Level 5. Data quality management software. Data quality can be checked, improved and maintained through the automated detection and elimination of data errors, standardization of data, verification of information from external sources, and completion of missing data. Often, these tools can manage the data entry and modification process, continuously monitor data quality and then generate data quality reports. For example, an insurance company might use such tools to automatically detect and correct errors in customer data.

Level 6. A data simulation methodology. Data simulation is an incredibly useful tool in various fields and involves the creation of data that mimic real-world information. This is particularly valuable for testing different scenarios and variations of system operations, algorithms or mathematical models. Data simulation can be used for safe, controlled testing without using real data, which is a crucial facility when dealing with sensitive information. In the medical field, for example, data simulation can create artificial datasets of patient information. The simulated data can be used to test and evaluate the performance of new diagnostic or therapeutic algorithms, eliminating the risk of making real-world errors or endangering patients.

Level 7. Machine learning algorithms. These powerful tools can identify and classify atypical data in ways that are often hard or impossible using traditional analytical methods. The algorithms learn from existing data in order to classify new data, making them extremely versatile. In banks and payment processing companies, for example, machine learning algorithms can be used to identify anomalous customer behaviours. By analysing extensive datasets, machine learning algorithms can detect suspect behaviours and block such operations. Thus, by working with a database of banking transactions, machine learning algorithms further protect customers against, for example, fraudulent activity.

Area 4E. Computing

In the realm of digital transformation, the area of computing focuses on the essential tasks of processing and analysing data within an organization. Utilizing a variety of modern technologies, data computing aims to dramatically enhance operational efficiency and effectiveness. By effectively harnessing these technological tools, organizations can extract and leverage valuable insights from their accumulated data. This not only aids in more informed decision-making but also optimizes business processes, thereby ensuring that the organization remains agile and responsive.

Level 1. Computation by PC. Personal Computers are used for data-processing tasks. These stations are equipped with efficient data-processing software and hardware. Data processing on PC stations is suited to various fields, ranging from financial and marketing analysis to CAD/CAM design. The drawback of this method is its lack of scalability, meaning the amount of processed data is limited by the computational power of the individual machine. In the case of CAD/CAM design, PCs can be utilized for 3D modelling and simulation, but with more complex projects they may not meet the computational load requirements.

Level 2. Computation by edge devices. Edge computing is an approach to data processing in which computational operations are executed on devices located near the data source, rather than in distant data-processing centres or the cloud. This approach allows for faster, more efficient, more effective data processing since data do not need to be transmitted to distant data centres, thereby reducing delays and network loads. Additionally, computations on edge devices improve decision-making speeds, as data are processed in real-time, even before being sent to data-processing centres or the cloud. For example, in medicine, computations on edge devices enable real-time monitoring of a patient`s condition, and crucial responses to changes can be made more quickly than if the data were to be transmitted to a remote centre for calculation.

Level 3. Virtual computing machines. The virtualization of computing machines is a technology that allows multiple virtual machines to be run

on a single physical server or computer. Each of these machines operates as a separate system, with its own hardware resources and software, and their operation is isolated from other virtual machines and the host. Virtual computing machines in manufacturing companies, for instance, can speed up simulations used to optimize production scheduling by performing them in parallel, rather than in series. This enables near-real-time adjustments to production plans in cases of disruptions to production lines.

Level 4. GPU technology. Graphics Processing Units are a type of processor specifically designed for graphics processing and parallel computing. In recent years, they have become a popular tool for accelerating computations in areas such as machine learning, data analysis, numerical simulations and more. By leveraging hundreds or thousands of parallel cores, GPUs can accelerate computations by several orders of magnitude compared to traditional CPU processors.

Level 5. Cloud computing. This is a model of delivering computational, networking and application services over the Internet. In cloud computing, computational resources such as servers, memory, networks and databases are provided as a service, allowing them to be used flexibly, based on user needs. Various service models are available in cloud computing, including Software-as-a-Service, Platform-as-a-Service and Infrastructure-as-a-Service.

Level 6. HPC. High-Performance Computing is a technology that can execute extraordinarily complex computations on multiple processors simultaneously. This allows large amounts of data to be processed quickly, which is essential in various fields such as science, finance, industry and data analysis. HPC systems consist of multiple computing nodes connected by a high-throughput network. In the field of science, HPC technology is used for simulating complex physical phenomena to process vast amounts of data and obtain results in a short time.

Level 7. Computer clusters. These complex systems consist of many interconnected computers operating as a single computing unit. In the aviation industry, computer clusters are used for aerodynamics simulations, enabling the rapid processing of large amounts of data and shortening analysis times for more efficient designing of new aircraft models.

Axis 5: Competence levels and digital culture

Axis 5, focusing on the digital culture of an organization, describes softer, yet equally crucial aspects of a company`s digital transformation. Unlike the first three axes, which directly influence the financial performance of an organization, the last three, especially Axis 5, emphasize the organizational behaviours critical for successful transformation. Experience in implementing digital transformations with client companies indicates that failures are often attributable to a lack of attention to organizational culture.

A transformational culture within an organization encompasses values, attitudes and practices that promote agility in the face of changing market and technological conditions. Such a culture values openness to change, innovativeness and a propensity for experimentation and judicious risk-taking. In organizations endowed with a strong transformation culture, employees are actively engaged in the change process and encouraged by leaders to be creative and inventive, which enhances agility.

Culture of Transformation

	1. Leadership Styles		2. Readiness for Change	3. Continuous Improvement	4. Culture of Innovation	5. Resource Availability
Type 6	Transformative	Level 6	Consolidation of Changes and Integration into Organizational Culture	Cross-Organizational Project Teams	Collaboration with External Entities	Availability of Partners
Type 5	Passive	Level 5	Implementation of Organizational Change	Mentoring	R&D Process in the Organization	Access to Experts
Type 4	Autocratic	Level 4	Communication of Change Vision	Self-Education	Permission for Making Mistakes	Access to Training
Type 3	Directive	Level 3	Formulation of Vision and Change Strategy	External Trainings	Market Trend Analysis	Access to Technology
Type 2	Supporter	Level 2	Creation of a Coalition for Change	Internal Trainings	Experimentation	Access to Data
Type 1	Innovator	Level 1	Awareness of the Need for Change	Trade Shows and Conferences	Promotion of New Ideas in the Organization	Access to Capital
	1. Leadership Styles		**2. Readiness for Change**	**3. Continuous Improvement**	**4. Culture of Innovation**	**5. Resource Availability**

Figure 15

For digital transformation strategies to succeed, they must align with the organizational culture. New technologies and processes require shifts in both thought and action across all levels of the organization. Without a culture that supports innovation, flexibility and openness to change, even the most