CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

Data Geek

BUSINESS ANALYTICS Specialization

Volume: 1/ Issue 6/ August 2020

Data Governance in Analytics

From the Editors desk,

"Without a systematic way to start and keep data clean, bad data will happen."

– Donato Diorio, Founder & CEO, DataZ and RingLead Inc.

Data Governance is expected to play a key role in future Data Science and Data Analytics as it offers a means to perform key validity checks at multiple points in analysis. It enables teams to monitor before, during, and after the data analysis process to prevent data misuse and usage of corrupt methods of analysis that do not have a scientific backing.

Data Governance policies and procedures usually relate to the usability, integrity, security measures, and continuous availability of knowledge employed in any enterprise. Thus, Data Governance principles and practices are critical for all business-process functions like regulatory compliance, legacy upgradation, business intelligence systems, risk management, data warehouses, Cloud usage regulations and etc.

With the increasing concerns of data exploitation and the abuse of computational techniques in Data Science, it is becoming crucial that clear Data Protection policies and Procedures be placed in place to prevent the degeneration of data and the analytical methodology used to arrive at data-driven conclusions.

The key driving factors leading to the synergy between Data Governance and Data Analytics in the industry as of today are broadly generalised as:

Necessity of Maintenance of Ethics- Governance in Data Science and analytics would primarily play a role in ensuring that the techniques that would be applied for development of predictive models are properly validated and their originality is kept intact.

Data Governance as an Expanding Market Segment- In various market corridors, with increasing data leak controversies such as those relating to FB or Cambridge Analytica, data control, access, transparency, and data security are seeing heightened importance. This is resulting in the emergence of manifold platforms and solutoins. Some of them include sophisticated solutions for policy enforcement, policy monitoring, Data-Governance stewardship, and data discovery technologies.

Data Lakes- As one of the 5 Predictions for 2019: 'Business Value from Data', Forrester Research has identified the adoption of "data fabric technology" with data lakes. Data lakes are used by organizations who are serious about getting the maximum benefit from their development investments. Throughout 2019, the introduction of data infrastructure technologies would ensure efficient data governance and the execution of policies relevant to scalable application. Balancing Information Governance and Knowledge Protection provides a far-reaching approach as individual market consumers are encouraged to make critical choices, backed up by superior technologies. We encourage the reader to dive into the vast world of data, keeping in mind a moral frame that would ensure a holistic and sustained growth as a data scientist.

With this, we present this year, yet another enthralling edition of Data Geeks, centered on a pivotal theme of 'Data Governance and Analytics'.

Once again, I would like to congratulate the entire team that worked tirelessly and contributed in the making of this interesting and insightful newsletter.

Please feel free to reach out to us at any time for any queries or suggestions.

We can be contacted at our university email at- datageek@mba.christuniversity.in

With regards,

Dr.Kannika Nirai Vaani

CONTENTS

DATA GOVERNANCE – WHAT, WHY, HOW, WHO AND BEST PRACTICES

What is Data Governance?

Data governance refers to the overall management of aspects of data pertaining to the availability, usability, integrity and security of data used by an organization. Since the initial phases of development when the function of data, started becoming an important and fundamental issue for organizations, several industry experts have stressed the need for effective data governance. Data governance contains many aspects, which have been formulated from a combination of procedures, processes, responsibilities, policies, technologies, and decision-making rights for the use of data in organizations at all levels.

Data Governance includes the human beings, methods and technologies had to control and shield the organization's records property to guarantee generally comprehensible, accurate, entire, truthful, comfortable and discoverable company data.

Where is Data Governance used?

Data governance is a prerequisite for several tasks or projects and has many clear advantages:

- regular, uniform records and processes across the company are a prerequisite for higher and greater complete decision assist;
- growing the scalability of the IT landscape at a technical, business and organizational level via clear policies for changing strategies and facts;
- imperative manipulate mechanisms offer potential to optimize the price of facts control (more and more crucial within the age of exploding records sets);
- multiplied efficiency via the use of synergies (e.g. by reusing procedures and statistics);
- Better self-belief in records via first class-confident and authorized statistics in addition to complete documentation of information approaches.

The Outline of Data Governance

Effective Data Governance would entail the following basic aspects-

1. **Data Life Cycle**- By understanding how data is being employed in organizations, and the way long it must be retained, organizations can develop and deploy ways to create patterns in data acceptable the optimal storage media, thereby minimizing the full cost of storing data over its life.
2. **Compliance**- Data Governance should be deployed in a manner that leads to compliance with set regulatory requirements. Data governance with compliance would work towards achieving variety of objectives including benefits such as offering better visibility to internal and external customers and compliance with regulatory laws set in the industry.
3. **Data quality management**- Data quality is seen as the most vital aspect influencing usability of information for business processes and reporting. Data quality largely determines the effectiveness of the business processes, and also influences the reporting quality.
4. **Privacy and Security of Data**- Privacy of information, relates to the protection of knowledge stored via secure computer servers, or the other sort of electronic media such as Secure Cloud options. Describing in the simplest terms, it is expected that data will always be kept confidential, have integrity or accuracy, and be available when needed to produce them with service and records.

Best practices that will follow in general:

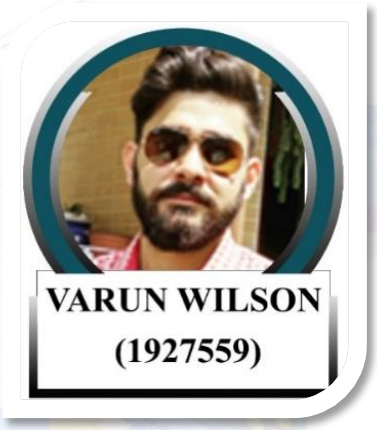
- Begin small- As in all elements of business, do no longer attempt to boil the ocean. Attempt for brief wins and build up pursuits over the years.
- Set clear, measurable, and specific goals- You can't manipulate what you cannot degree. Celebrate whilst dreams are met and use this to move for the subsequent win.
- Define ownership- Without commercial enterprise ownership an information governance framework cannot prevail.
- Become aware of related roles and responsibilities- Records governance is a teamwork with deliverables from all parts of the commercial enterprise.
- Educate stakeholders- Anywhere viable use commercial enterprise phrases and translate the instructional elements of the information governance field into significant content material within the commercial enterprise context.

REFERENCES

1. Kooper, M.N., R. Maes & E.E.O.R. Lindgreen (2011). On the governance of information: Introducing a new concept of governance to support the management of information. International Journal of Information Management. 31(3), 195-200.

2. Ladley J. (2012). Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program. Newnes.

3. Khatri, V. & Brown, C.V. (2010). Designing data governance. Communications of the ACM, 53(1), 148–152.



DATA GOVERNANCE VS. DATA MANAGEMENT AND GOVERNANCE IN INDIA

Data governance is the process of managing the data availability, integrity, relevancy, usability and security in an organization. It is a subset of Data Management. It helps the organizations manage their information knowledge and answer to questions such as

i. What is the source of this data? ii. Does this data obey to our companies rules?

iii. What do we know about our information?

A predominant analyst at Forrester defines information governance as "a strategic commercial enterprise application that determines and prioritizes the economic benefit statistics brings to agencies as well as mitigates the business hazard of terrible information practices and nice." SAS, a worldwide chief in analytics, defined the predominant manner information governance fails in many companies. Some groups view information governance because the primary way to organize records management initiatives and priorities. Every enterprise can have extraordinary primary tasks to accomplish inside its enterprise. That's why growing a custom designed facts governance framework can assist organizations' leaders develop motion plans primarily based on their enterprise's goals and ache points. Some of the capabilities that make up a records governance framework include:

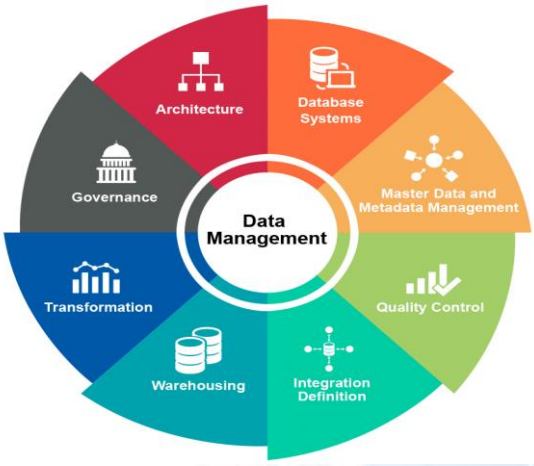
i. Statistics first-class measurement ii. Threat control iii. Compliance

iv. Value minimization v. Records security vi. Facts warehousing

Data Management:

Data Management (DM) is a program which consists of the operational processes which are executed and governed on a set of people in an organization and it also includes various types of technology to preserve and deliver information that is trustworthy, understandable and controlled. Data management is a set of technology that execute on numerous business rules and policies at the same time as contributing to the statistics- and compliance-primarily based requirements of customers and shareholders.

To get a complete draw close at the idea of information control, we can explore one important statistics control strategy: data safety. This is a major precedence for all B2B and B2C businesses that cope with private information. Even though many agencies consider the protection of records to a prime precedence, many don't awareness on how they may be storing this record in the corporation. Proper control of information means regulating the records that enters a database at any given time. The results of powerful reporting and dealing with of information safety include high levels of compliance and accuracy. In other words, Data management involves the implementation of tools, architectures and various processes to achieve the objectives of the Data governance. Governance is one of the process involved in Data Management.

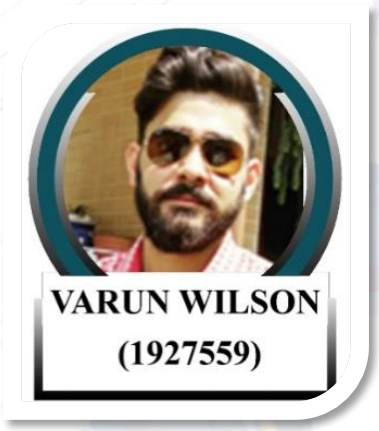


Data Governance in India:

After the launch of Digital India Program, the country has seen tremendous growth in digital infrastructure in innovating e-governance policies that can lead to digital empowerment of

citizens. Access to the internet across all areas has made India the country with the second-largest internet users in the world.

These technological advances have led to large volumes of data generated by various activities. India is proposing to regulate personal data from the approaches of China, European Union, United States through India's draft Personal Data Protection (PDP) Bill was passed in 2019. This Data Governance framework constitutes of three players which are Data Fiduciaries those who control the means of processing personal data, Data Processors who process data on behalf of DFs and Data Principal those whose personal data is processed.



CII GOVERNANCE

Mr M Sankaranarayanan, Managing Director, The Grand Technologies and Founder cum Chairman of Aathmalayaa School, Karaikal has been elected as the Chairman of Confederation of Indian Industry (CII), Puducherry State Council for the year 2020-21.

Income tax return: Processing all tax assessments in a technology driven and faceless manner is set to offer a big compliance relief to businesses while giving income tax department officials more time to specialize in complex areas of taxation, said industry representatives and experts.

While the tax department processes bulk of the returns in an automatic way, some may need clarifications and the degree of deep dive needed may vary. Some returns are chosen for scrutiny although these account for only a fraction of all the returns processed. Now, most assessments including scrutiny cases will be under the faceless scheme.

Tax reforms such as faceless assessment would promote ease of living, industry body Confederation of Indian Industry (CII) said quoting its president Uday Kotak. A policy-driven governance based on digital interface minimizes grey areas, eliminates discretion and promotes transparency and certainty, said Kotak. –Article from live mint (19th Aug 2020).

Corporate governance bar must be raised:

The government has prioritised creating a facilitative regulatory structure over the years. The Ministry of Corporate Affairs brought out the Corporate Governance Voluntary Guidelines in 2009 and, in 2012, it constituted an expert committee under the chairmanship of then Confederation of Indian Industry (CII) President, Adi Godrej. The committee's report prescribed 17 'Guiding Principles of Corporate Governance' which formed the basis for enhancing governance standards in the country. In October 2017, the report of the Committee on Corporate Governance under the leadership of Uday Kotak was released and several of the suggestions were incorporated into the SEBI (Listing Obligations and Disclosure Requirements) Regulations.

As a $3-trillion economy aspiring to enhance its global position, India is transforming its business environment and looking to entrepreneurs as wealth-creators. Ways of doing business are changing to align with global patterns. Today, governments and regulatory agencies expect companies to conform to global best practices for better functioning of markets and

convergence of business practices with societal needs. – Article from Business line (27th FEB 2020)

DATA GOVERNANCE FOR AI

Machine learning is the base of Artificial Intelligence applications that deciphers patterns and correlations in data using statistical analysis. Despite the increase in processing power and lower cost of infrastructure, the biggest challenge to for machine learning is not the speed of processing but in fact the data it is fed. Since machine learning has to be trained with data, any inconsistencies in the data can have reduce the quality of the AI and bring out inaccurate results. With AI becoming an integral aspect of businesses, data governance is also becoming an essential part of business operations. Data Governance is defined as the process of managing the availability, usability, consistency and security of the data stored in the enterprise systems based on internal policies on data standards and usage controls. An effective data governance policy ensures that the data available for analytics and further decision making is reliable for use and is not misused in any way. Without data governance, there would be inconsistencies in between data which could further reduce the effectiveness of Business Intelligence.

Business, AI and Data Governance

As more and more businesses are adopting AI with many more with plans to implement AI in the decision-making process, a large majority of these companies are yet to implement governance support for their AI.

Going forward, organisations will have to keep the following things in mind while building data governance policies:

1. Making sure that data is secure while still accessible to those who need it to draw insights
2. Preventing loss of data and ensuring that the data is not used for unethical reasons
3. Ensuring that the data is consistent across all systems in collection and storage
4. A complete tracking system should exist to trace where the data originated and who is using the data
5. All data being stored must be complete and without any errors

From these governance policies it can be seen that there are two opposing forces that organisations need to balance- data use and data protection. Extremely strict data protection policies within the organisation may lead to them missing out on important opportunities that could have been identified by analysing the existing data.

To ensure the quality of the data, checks must be done routinely at each step. By assessing the quality of the data, changes can be made on each level of data storage and transfer to ensure that all the governance policies are complied with.

Data as an Asset

As mentioned earlier, data and more importantly complete and accurate data is the essential building block of AI services that will be implemented by the organisations. Since AI is a long-term asset for the organisation and will play an important part in the formation of policies, data must be made a centrepiece during formulation of data related strategies. This help assure that the quality of data and results can be maintained in the long-run.

This would also help reduce the workload of data-scientists on cleaning the dataset before model building, which is the most time-consuming part of a data scientist or analysts' job.

In order to gain the accurate insights from the AI and to ensure that it happens for the long-run, organisations must start to build their data governance policies before implementing AI solutions in the business. This is so as to avoid the sunk cost of building and deploying the AI model with inaccurate predictions or incur additional costs in cleaning and integrating data properly.

MOBILE PHONE DATA DURING COVID-19: WHAT IT MEANS FOR THE FUTURE OF DATA GOVERNANCE

Tracking COVID-19-infected people has become an important weapon in global responses to combat the virus. Mobile technology offers an easy method for tracking people potentially exposed to COVID-19 with the use of geo-location. Big data analytics have the potential to track the spread of the pandemic and use analytics to forecast future contagion patterns.

The use of large data in the fight against coronavirus (COVID-19)—particularly the use of detailed data on mobile phones to track and monitor the pandemic — has fuelled serious privacy concerns. However, monitoring seems to be taking place at national level for now. This issue is important not only for the use of data in an emergency, but also because it sets a precedent that can and will be replicated across sectors.

Four areas of concern have largely been highlighted by phone-based contact tracing critics:

Accuracy: Critics point out that mobile phone tracking cannot accurately track dispersion and contagion. According to medical consent, the virus spreads when people are less than six feet apart. While call details, GPS, and Wi-Fi data can give us a general sense of where a person is, it cannot determine with sufficient precision whether two individuals are within six feet of each other. And while Bluetooth may help solve the problem, the technology is not yet ubiquitous.

Probability vs. reality: The bigger problem with data tracking for mobile phones is that it measures probabilistic risk rather than actual exposure. The provision of exposure information to people in many places can overwhelm health systems, as many people without symptoms may flock to health centres for fear of being exposed. Validating records of digital contact tracing would be a lengthy and labour-intensive task, resources that could be better used for other, established methods of response and recovery.

Representativeness: Another big concern is the fact that smartphone data is leveraged by digital contact tracing. For example, although India has a large tech industry, smartphones are available to only 28 per cent of people. And in Ethiopia, which has one of the lowest penetration rates for smartphones in the world, this figure is about 11 percent. Smartphones are also much skewed towards the male part of the population.

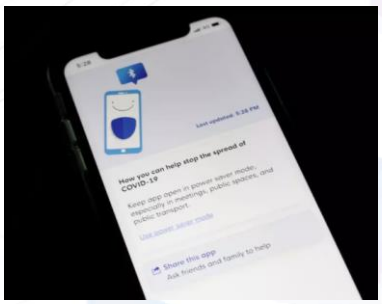
Privacy: Eventually, and perhaps most importantly, if left unchecked, data monitoring gives governments extra control. Despite these concerns, we acknowledge the enormous potential to harness big data. Its growth is an opportunity for development that is too large to be ignored — but it's important that we get the right use of data and put in place correct safeguards. We already know how critical it is to use data responsibly.

Based on validated outcomes: Electronic data has proved immensely helpful in tracking and managing aggregate-level results and service delivery. We use data from Call Detail Records (CDR) to optimize networks in the public transport. Such processes were previously developed and validated while retaining standards for privacy and representativeness. In the sense of COVID-19, we could easily see at aggregate rates how people travel to provide assistance more effectively after a disaster, or track patterns of migration on a large scale.

Supporting the creation of data trusts: these could be an efficient way to share and use data. Civic data trusts step beyond single trustees and create fiduciary governance structures for maintaining, accessing and exchanging data rights. Digital trusts, like code or a subset of digital rights, could hold digital assets. They can serve as clearing houses for rights or as management of processes. The digital data trust would be a private organizational structure, publicly and legally accountable, capable of balancing complex interests to serve a unifying purpose. **Develop an agenda for good data management**: the use of data is most successful in environments with strict criteria for data governance and accountability. This ensures that private data will also be used in a straightforward manner, being very clear about the aims and uses and ensuring that all data is private by design, as was our use of cell phone data for estimating the origin-destination. The collection of data is essential to COVID-19's response and wider development activities, but it must respond to a particular need and be regulated by institutional frameworks that ensure contextual interest, accountability and privacy.

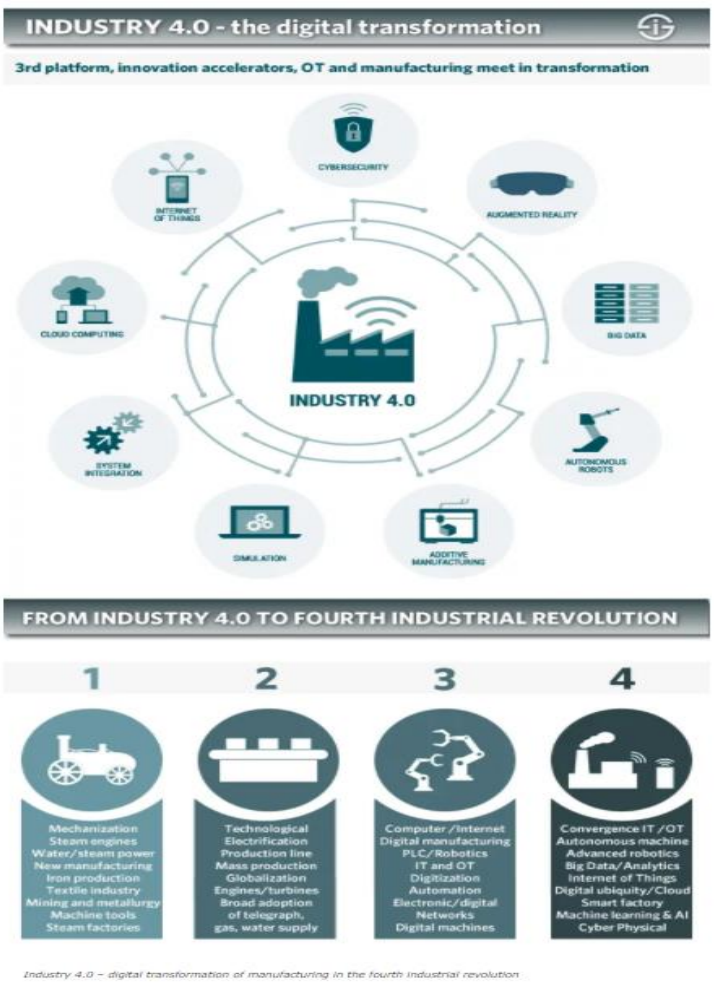
DATA GOVERNANCE MODEL FOR CONTACT TRACKING

The contact tracking app is the effective way of tracking the contact of the person. This app works in the background under power saving mode. Governance mechanisms should address three basic factors: individual rights such as privacy, the interests of data holders and the broader public interest. The use of contact tracking apps during the COVID-19 pandemic has ignited controversy about the effect on privacy and human rights of these devices. Data tracing apps can be effective anti-virus weapons – but they can also be state surveillance devices.



In response to government efforts, several private-sector initiatives have emerged that promise to be more conscious of privacy. Among them is an exposure-notification platform developed by Google and Apple. The platform shields users from government monitoring, but in doing so, effectively places the technology developers (the data holders) in the role of the state. This raises the stakes for governing how these private companies use the data they collect. Countries which have used personal data extensively have, generally, been more effective in suppressing the virus. The task, especially during a deadly pandemic, is to balance people's rights with the needs of society at large.

DATA GOVERNANCE FOR INDUSTRY 4.0 WHY DATA MANAGEMENT IS CRITICAL FOR INDUSTRY 4.0



Industry 4.0 – digital transformation of manufacturing in the fourth industrial revolution

Introduction to Industry 4.0

Industry 4.0 is the digital transformation of assembling or creation and related businesses and worth creation forms. Industry 4.0 is utilized reciprocally with the fourth modern revolution and speaks to another phase in the association and control of the industrial value chain. Cyber-physical frameworks structure the premise of Industry 4.0 (e.g., 'smart machines'). They utilize present modern control frameworks, have inserted programming frameworks, and discard an Internet address to associate and be tended to by means of IoT (the Internet of Things). Along these lines, items and methods for creation get organized and can 'impart', empowering better approaches for creation, esteem creation, and ongoing enhancement. Cyber-physical frameworks make the capacities required for shrewd industrial facilities.

Importance of Data Management in Industry 4.0

The targets of Industry 4.0 are to make all significant data of a modern area accessible continuously by associating every single pertinent substance and to have the ability to utilize the information that is produced to decide current procedure statuses consistently to infer the most ideal worth including choices. Industry 4.0 will convey a ceaseless wellspring of possibly important information.

Spanning the correspondence hole between lines of business, producing frameworks and mechanical gadgets, segments, and work pieces targets:

- Altered large scale manufacturing (low amounts, cluster size one)
- Reconfigurable creation frameworks
- The alternative to deliver a wide scope of various kinds of items on a similar creation lines.

Reviewing these goal-oriented destinations, second-level purposes like streamlined vitality the board, condition observing and finding or self-revision (where a procedure responds self-governing, quick and absolutely when there are deviations from the benchmarks) are "low-hanging fruits".

Thusly, Industry 4.0 spotlights on every one of the three of the accompanying measurements on coordination:

- **Horizontal integration** by value adding networks

Inside the setting of level reconciliation, interconnected organizations – producer, provider, and improvement and coordination administrations – routinely trade important data. This idea is to assess client explicit prerequisites all through all the various periods of an item's lifecycle – including plan, creation, conveyance, and use.

- **Vertical integration** within automation hierarchies

Vertical guidelines connect the various pecking orders inside the computerization innovation, for example at actuator and sensor, control, and arranging units.

- **Self-streamlining** of assets

Coordinating the assembling procedure is basic to self-improvement. The accessibility of interrelated information and the fitness to bridle canny devices and ideas make ready for esteem including improvements.

As a general specialized system for Industry 4.0, a Reference Architecture Model Industry4.0 (RAMI 4.0) has been indicated. Orchestrated on three tomahawks – functionalities inside production lines or offices, lifecycle and esteem stream and layer-based deterioration of a machine – specialized norms characterize a typical structure and "all inclusive arrangement of dialects" for explicit spaces.

References:-

- https://searchdatamanagement.techtarget.com/definition/data-governance#:~:text=Data%20governance%20(DG)%20is%20the,and%20doesn't%20get%20misused.
- https://www.i-scoop.eu/industry-4-0/#:~:text=Industry%204.0%20is%20the%20digital,of%20the%20industrial%20value%20chain.
- https://www.i-scoop.eu/industry-4-0/industrial-data-analytics/
- https://www.talend.com/blog/2017/05/08/data-management-critical-industry-4-0-success/#:~:text=The%20objectives%20of%20Industry%204.0,derive%20the%20best%20possible%20value

JOB OPPORTUNITIES IN DATA GOVERNANCE

A data governance analyst specializes in managing the information flow for organizations, including the integrity, efficiency, and availability of data. These analysts are needed in a variety of industries, including technical service firms or financial services agencies. They usually work full-time in an office environment.

General job duties for data governance analysts may include developing and implementing data standards, guaranteeing metadata is captured correctly, and creating methods for monitoring and reporting any data incidents. Data governance analysts may monitor application personnel to ensure that they are following data governance procedures appropriately. They may also create and maintain common data dictionaries and the tools or methods that support data standards for an organization.

Data governance analysts can analyse and oversee strategies designed to enhance data reliability and minimize redundancies. They could work with other information technology personnel to develop data quality standards. Data governance analysts can create key performance indicators and supply colleagues and organizational management with performance results. They can also assist with data monitoring and provide organizational management with guidance on streamlining data management.

Educational Requirements	Bachelor's degree
Job Skills	Excellent technical abilities, strong communication skills, effective analytical abilities, and strong project management skills
Median Salary (2020)*	$68,467
Job outlook (2018-2028)**	10% (Computer Occupations, All Other)

Sources: *PayScale.com, **U.S. Bureau of Labor Statistics'

Required Education.

Individuals will need a bachelor's degree in computer or information science to work as a data governance analyst, with some employers preferring a master's degree. A strong background in large-scale, shared data environments could be beneficial. Interested individuals can join a professional association like the Data Governance Professionals Organization. The organization provides governance professionals with networking and training opportunities.

Required Skills.

Excellent technical abilities are the most important asset for data governance analysts, as they are responsible for ensuring an organization's data is accessible. Relevant technical skills or knowledge may include database modelling principles, SQL, Security Master, and Microsoft Office. Data governance analysts should have strong communication skills in order to interact with internal and external stakeholders. They should also have effective analytical abilities in order to examine infrastructures and make recommendations on improvements. Data governance analysts will also need strong project management skills in order to manage daily operations or oversee information technology personnel.

Career Outlook and Salary.

The U.S. Bureau of Labor Statistics (BLS) does not collect statistics on data governance analysts; however, they reported a 10% growth for computer occupations, all other, during 2018-2028, which is about as fast as average for all occupations. In May 2020, PayScale.com reported a median annual salary of $68,467 for data governance analysts.

DATA GOVERNANCE IN THE AGE OF THE HOME OFFICE

The COVID-19 pandemic took many businesses by surprise. Companies that were vehemently opposed to remote work because of security and compliance concerns have found themselves forced to reconsider their stance at an alarmingly quick rate. The transition to remote work sent many companies into chaos, with data security practices becoming more lax. Cyber attackers have taken advantage of the situation to prey on vulnerable employees. Malicious attacks against corporate networks and servers also saw a surge as the pandemic slowed down response teams dealing with attacks. An increase in USB drive usage: Removable devices, and USB drives, in particular, have long been a thorn in the side of data protection strategies. **Adobe makes CDP data governance tools generally available**

Adobe announced the general availability of data governance tools in its Customer Data Platform (CDP). The platform is designed to help marketers combine known customer data with pseudonymous data (customer data that's been separated from the customer's identity).The data governance tools can help marketers ensure they're using that combination of data responsibly. The CDP includes a privacy console, from which a data steward can apply out-of-the-box labels or build their own custom labels to classify data. Real-time CDP is powered by Experience Platform capabilities and is packaged for the marketer," Skinner explained, "but if a customer had the Experience Platform and was using across marketing and IT teams, the data labels used and policies set within this framework would apply across the board

A new data governance model for contact tracing: Authorized Public Purpose Access

- Contact tracing apps are an effective way to track the virus, but proper data governance is essential.
- Governance mechanisms should address three basic factors: individual rights such as privacy, the interests of data holders, and the broader public interest.
- A balanced approach is the Authorized Public Purpose Access, a framework for governing the use of data that can be helpful in tackling serious problems such as pandemics.

The use of contact tracing apps during the COVID-19 pandemic has sparked debate about the impact of such tools on privacy and human rights. Contact tracing apps can be powerful weapons against the virus – but they can also be tools for state surveillance.

E-governance: Niti Aayog engages Oracle Cloud for data-driven transformation of aspirational districts

A key project of the government, its purpose is to help improve the quality of life of citizens in 112 of India's most backward districts, constituting 28% of India's population of nearly 1.4 billion. Government policy think-tank Niti Aayog has engaged Oracle to help it modernise vital IT infrastructure underpinning the Aspirational Districts Programme.

Information is collected across five key themes, on 49 key performance indicators. They include: education, health and nutrition, agriculture and water resources, financial inclusion and skill development, basic infrastructure including access to road, potable water, internet connectivity, and housing under the Pradhan Mantri Awaas Yojana – Gramin (PMAY-G).

This work requires assimilating thousands of data points from multiple sources and in different formats from across the country, which are checked for data accuracy (triggering alerts if the data feed appears to be inconsistent), and integrated for analysis and reporting. Critical to the success of the programme is for this data to be fed back in real time to administrators, district collectors, government officers and the general public via a visual and intuitive dashboards.

Oracle Autonomous Database is at the heart of data management in this programme. Being self-driving, self-securing and self-repairing, it requires a minimal IT team and resources to support it. The system and the data it manage are protected from system failure and attack by machine learning (ML) and automation.

DATA GOVERNANCE AND ITS CHALLENGES

Data governance is the systematic orchestration of entities, processes, and technologies to allow an entity to utilize its data as an asset to the enterprise. It also assigns data owners and stewards to data assets to improve transparency and allow data resource access, promoting data use. And by providing reporting, quality measurements, and scores, it can directly impact the quality of the results. Machine learning algorithms can track and enhance the quality of data in an organization with data governance enabled by analytics, self-learning as issues have been resolved. Governance programs essentially confirm compliance with regulatory requirements, without analytics. Organizations can proactively recognize areas where they may have breaches with analytics and active monitoring — a capability that will only become more relevant as legislation such as the General Data Protection Regulation (GDPR) falls into effect.

Necessity of self service

There should be a mutually beneficial relationship of data governance and analytics, but it must be said that if an organization delegates all responsibility for analytics to IT or an elite group of data scientists, it is largely a moot point. Today, data governance and data management will concentrate on empowering business users with a self-service approach to accessing and managing data properties, because they are the power users looking to use the data. IT resources are limited, and it is possible to invest their time rather than handling a continuous stream of overlapping demands from other business users.

Problems in data governance

Unfortunately, only about one-third of data-driven initiatives are successful in bigger companies. The reasons for the failure often stem from inadequate data governance. When a company indicates that the data quality is inadequate to be used for decision-making, this will signal poor data management. When the data cannot be processed, this would be a symptom of inefficient data management.

Ineffectiveness

Issues involving data and analyses, ranging from small errors in calculations to instability in data sets, all build to create a culture of data mistrust and, by extension, mistrust in any data-based decision. The root cause of these issues is often ineffective in data governance. Organizations may be unaware of critical governance tools such as data dictionaries, data catalogues, and heuristic layers, or the organization may suffer from a lack of communication between the data owners and the end-users.

Analytics as the helping hand

Ultimately, with so much data entering the data lake, companies need to streamline the most basic tasks of data governance that affect analytical efforts. Data quality controls can be automated today so that organizations can take full advantage of data governance enabled by analytics. AI and machine learning allow teaching data governance systems to help track the data's health so that more team members can trust the data, depending on it and create more (and better) analytics projects with it.

Conclusion

According to McKinsey, “Data governance is critical to capturing value through analytics, digital, and other transformative opportunities”. Although, many organizations are struggling to get it right, each business will thrive by changing its attitude from thinking about data governance as structures and policies to strategically embedding it into the way the organization operates every day.

QUIZ

1. Which of the following is not a “source” of data protection requirements?
    a. Legal
    b. Regulatory
    c. Geographical location
    d. Customer list
    e. Business partner contracts

2. On the basis of an enterprise data protection framework which of the following standards could be leveraged upon?
    a. AS/NZ 4360
    b. COBIT 4.1
    c. ISO 17799:2005
    d. None of the above
    e. All of the above

3. What are the key supporters of Data Governance 3.0?
    a. Dev Ops
    b. Domain Lineage and Discovery
    c. Automated workflow collaboration
    d. Intelligent glossary association
    e. b, c & d.

4. In the journey of Data governance and compliance which is the FINES use case?
    a. Avoid risking through data privacy and compliance
    b. Understanding business value through analytics
    c. Fueling business through enterprise data governance

5. Which of the following is not the requirement for a data governance 3.0?
    a. Scale across systems and business to ensure the efficient operation and accurate analytics
    b. Defining the policy for quality access, protection, retention and deletion
    c. A data management system which is comprehensive enough to support all cases
    d. None of the above
    e. All of the above

6. Among the following which is not a part of the data life cycle?
    a. Storage
    b. Anonymization
    c. Sharing
    d. Destruction
    e. Acquisition

7. The pre-requisite work that is recommended to be addressed before embarking on an enterprise data protection framework includes:
    a. Determining the source/existence of data
    b. The design role of enterprise
    c. Determining the scope of element
    d. a & c
    e. All of the above

8. One of the things that you could not do with the enterprise data protection framework is:
    a. To develop a roadmap for the enterprise data protection program
    b. Segregation of duties issues in an ERP system are addressed
    c. To remediate data protection gaps
    d. Perform a gap analysis of existing data protection controls

9. Which of the following is not data governance tool?

   a. OvalEdge
   b. Collibra
   c. Talend
   d. Hadoop
   e. None of the above

10. Put the correct order for sales life cycle for data governance opportunity

    1- Discovery, 2- Qualification, 3- Initial presentation, 4- Prospecting.

   a. 1-3-4-2
   b. 4-1-3-2
   c. 2-3-4-1
   d. 1-4-2-3

11. Which of the following is a data governance organization?

   a) DAMA
   b) DGPA
   c) DGMO

12. Which of the following is not a data governance role?

   a) Data Pilot
   b) Data Steward
   c) Data Manager

13. Which of the following is not a typical data governance task?

   a) Data Cataloging
   b) ETL Development
   c) Data Quality Monitoring

ANSWERS:

1) d
2) e
3) e
4) a
5) b
6) e
7) d
8) b
9) d
10) a
11) a
12) a
13) a

6. DATA GEEK CREW

SAILJA BEHERA

RENUKA DUTTA

FENI MARTINA FERNANDO