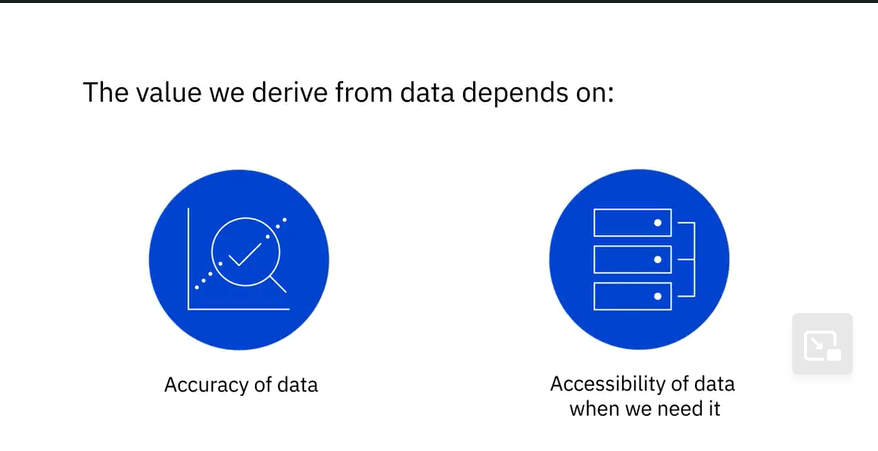
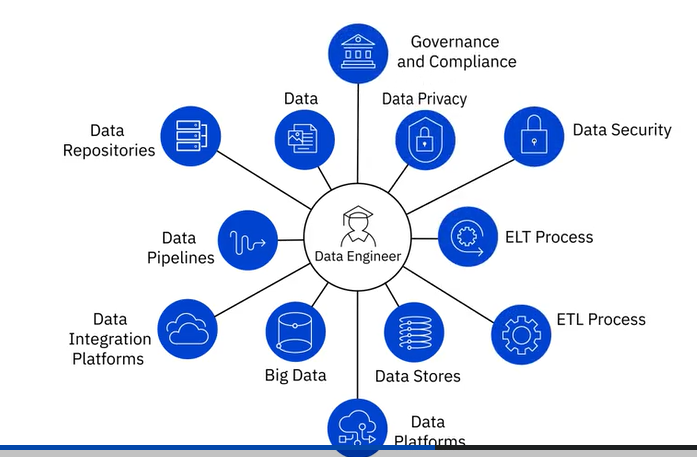
Welcome to Into to Data Engg.

Data is being generated every day

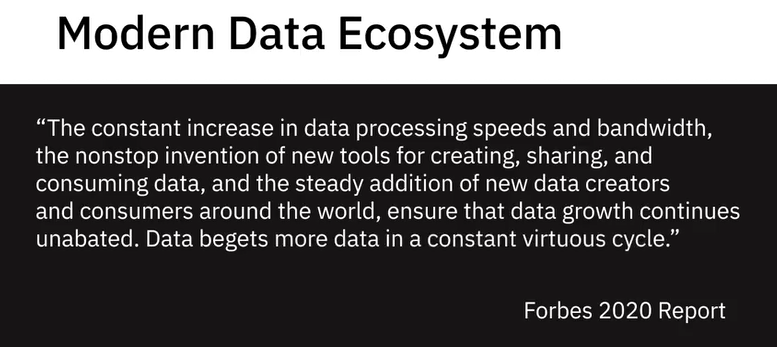


Growth of data engg jobs is 50% year over year.



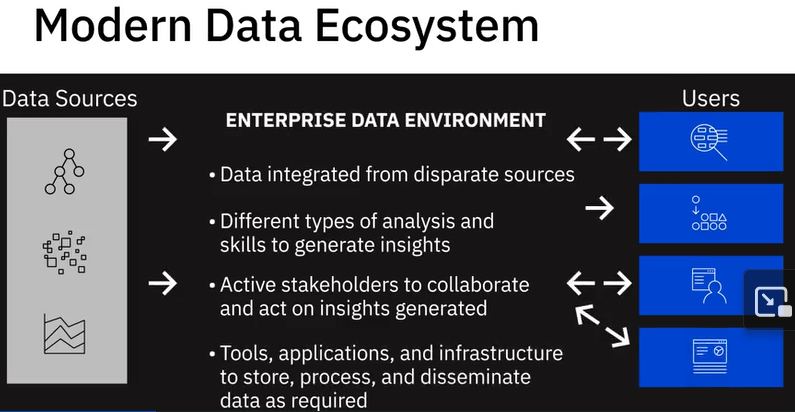
We will learn all these.

Modern Data EcoSystem

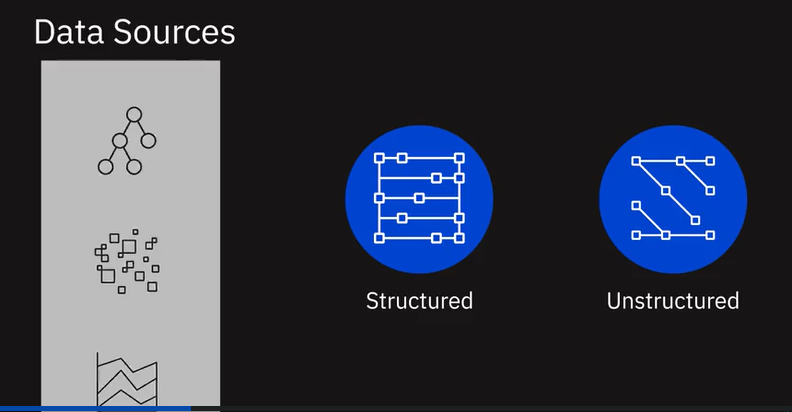


Modern Data Ecosystems contain Interconnected, Independent, Continually evolving entities .

Moden Data Ecosystem contains the following



Data Sources contains structured and unstructured data from different sources.



Data is available in a variety of structured and unstructured datasets, residing in text, images, videos, clickstreams, user conversations, social media platforms, the Internet of Things (or IoT) devices, real-time events that stream data, legacy databases, and data sourced from professional data providers and agencies.

We need to pull the data from these different sources to a data repository by aquiring data that we need, working with data formats, sources, and interfaces through which this data can be pulled.

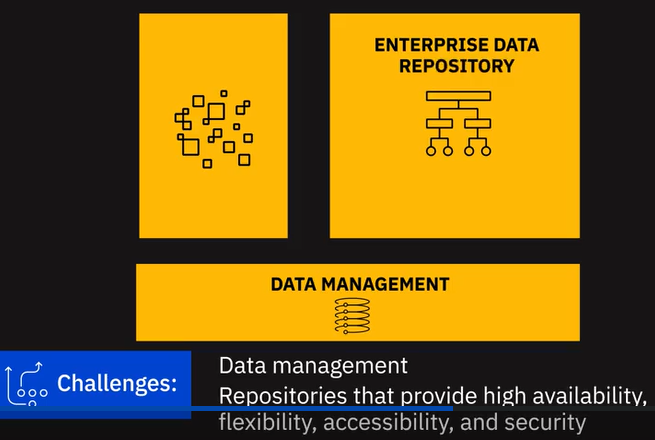
Challenges: reliability, security and integrity of the data.



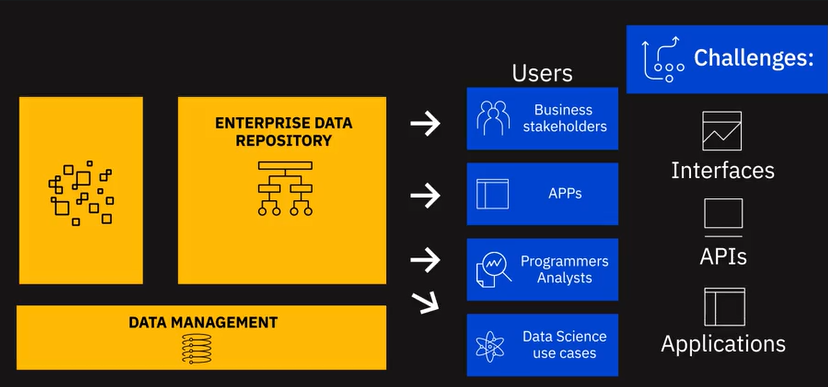
Data can be only used when the data is clean and organized.

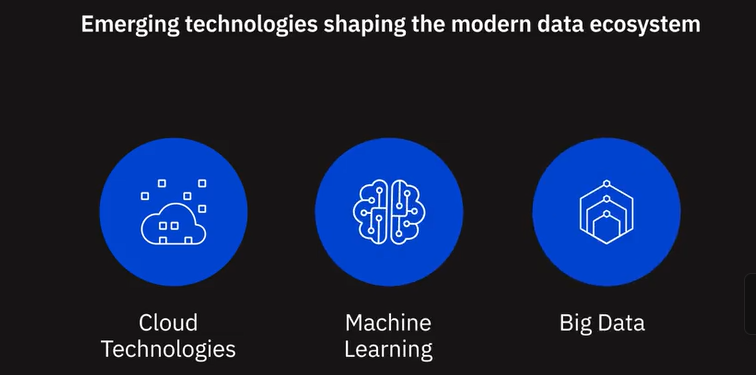
We need to conform to regulations and guidelines when the data is sensitive like healthcare, biometric and any other kind of personal data.

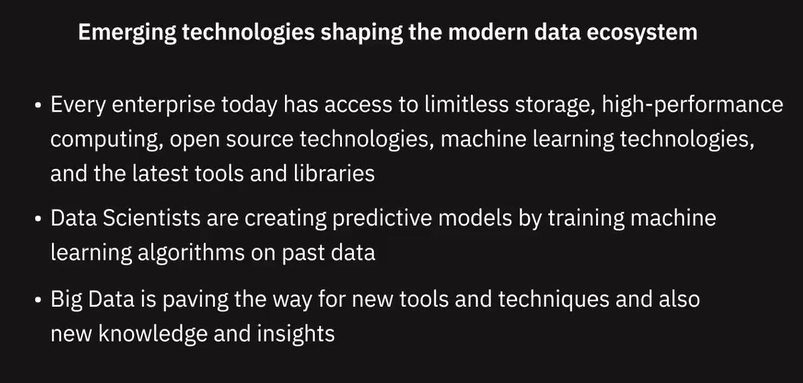
Adhering to master data tables within the organization, to ensure standardization of master data across all applications and systems of an organization, is another example. The key challenges at this stage could involve data management and working with data repositories that provide high availability, flexibility, accessibility, and security.



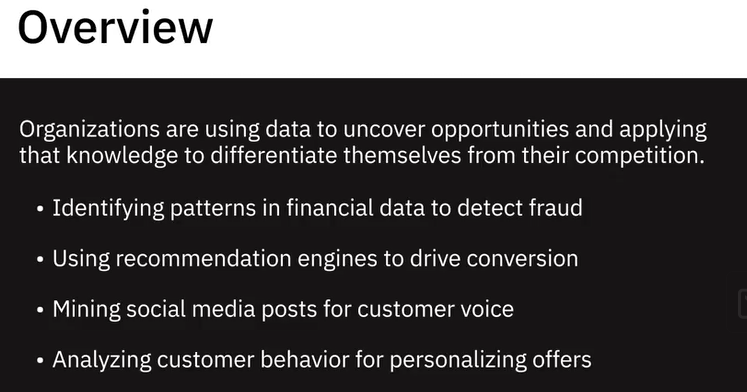
Finally, we have our business stakeholders, applications, programmers, analysts, and data

science use cases all pulling this data from the enterprise data repository.The key challenges at this stage could include the interfaces, APIs, and applications that can get this data to the end-users in line with their specific needs.For example, Data Analysts may need the raw data to work with, business stakeholders may need reports and dashboards, applications may need custom APIs to pull this data. 





Key Players in the Data Ecosystem



​

Today, organizations that are using data to uncover opportunities and are applying that

knowledge to differentiate themselves are the ones leading into the future.

Whether looking for patterns in financial transactions to detect fraud, using recommendation

engines to drive conversion, mining social media posts for customer voice, or brands

personalizing their offers based on customer behavior analysis, business leaders realize

that data holds the key to competitive advantage.

To get value from data, you need a vast number of skillsets and people playing different

roles.

In this video we’re going to look at the role Data Engineers, Data Analysts, Data Scientists,

Business Analysts, and Business Intelligence (or BI) Analysts play in helping organizations

tap into vast amounts of data and turn them into actionable insights.

It all starts with a Data Engineer.

Data Engineers are people who develop and maintain data architectures and make data

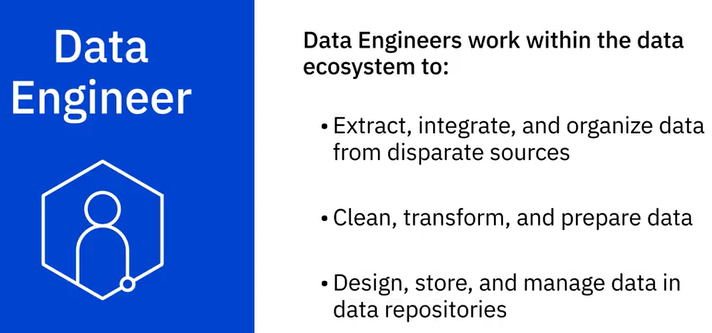
available for business operations and analysis.



Data Engineers work within the data ecosystem to extract, integrate, and organize data from

disparate sources; clean, transform, and prepare data; design, store, and manage data in data

repositories.



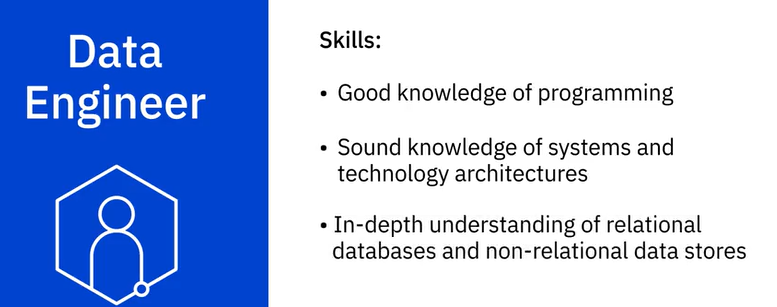
They enable data to be accessible in formats and systems that the various business applications,

as well as stakeholders like Data Analysts and Data Scientists, can utilize.

A Data Engineer must have good knowledge of programming, sound knowledge of systems and

technology architectures, and in-depth understanding of relational databases and non-relational

datastores.

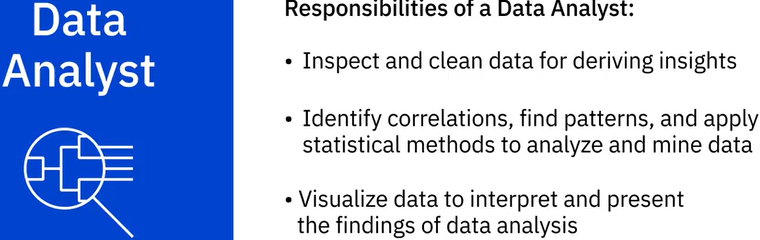


Now let’s look at the role of a Data Analyst.

In short, a Data Analyst translates data and numbers into plain language, so organizations

can make decisions.

Data Analysts inspect, and clean data for deriving insights;identify correlations, find patterns, and apply statistical methods to analyze and mine data; and visualize data to interpret and present the findings of data analysis.



Analysts are the people who answer questions such as “Are the users’ search experiences

generally good or bad with the search functionality on our site” or “What is the popular perception

of people regarding our re-branding initiatives” or “Is there a correlation between sales

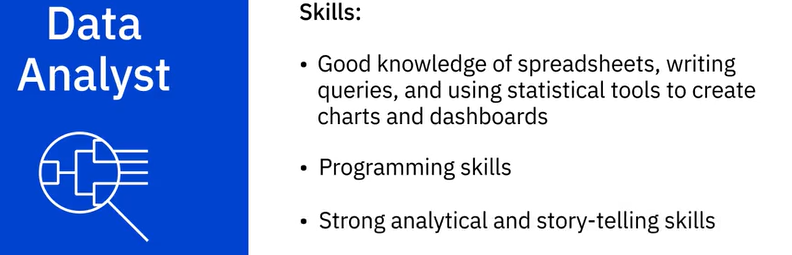
of one product and another."

Data Analysts require good knowledge of spreadsheets, writing queries, and using statistical tools

to create charts and dashboards.

Modern data analysts also need to have some programming skills.

They need strong analytical and story-telling skills.



And now let’s look at the role Data Scientists play in this ecosystem.

Data Scientists analyze data for actionable insights and build Machine Learning or Deep

Learning models that train on past data to create predictive models.



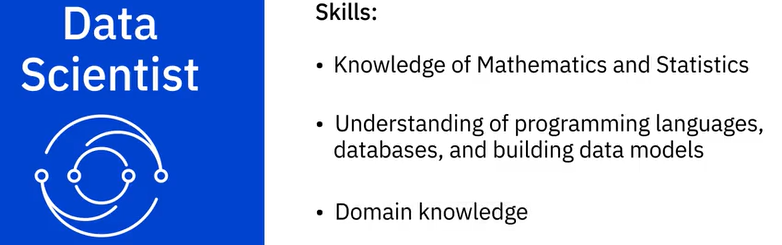
Data Scientists are people who answer questions such as “How many new social media followers

am I likely to get next month?” or “What percentage of my customers am I likely to

lose to competition in the next quarter” or “Is this financial transaction unusual

for this customer?”.

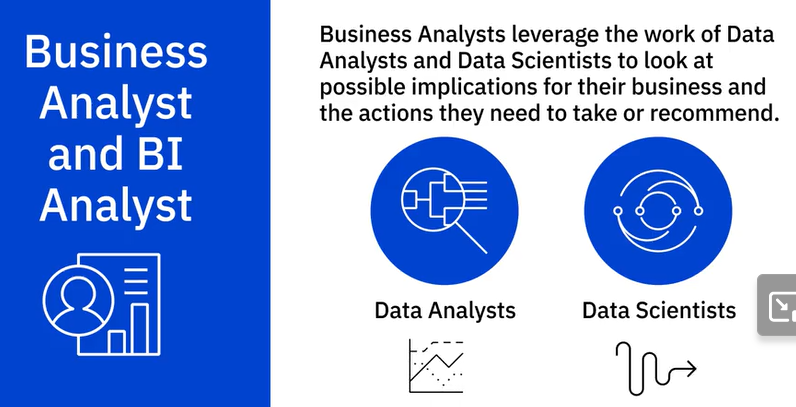
Data Scientists require knowledge of Mathematics, Statistics, and a fair understanding of programming languages, databases, and building data models. They also need to have domain knowledge.



Then we also have business Analysts and BI Analysts.

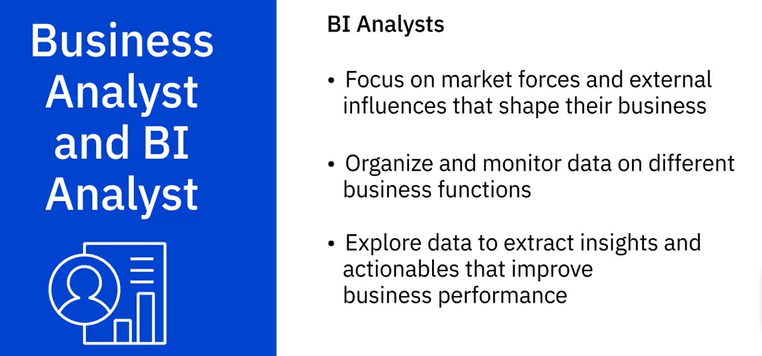
Business Analysts leverage the work of Data Analysts and Data Scientists to look at possible

implications for their business and the actions they need to take or recommend.

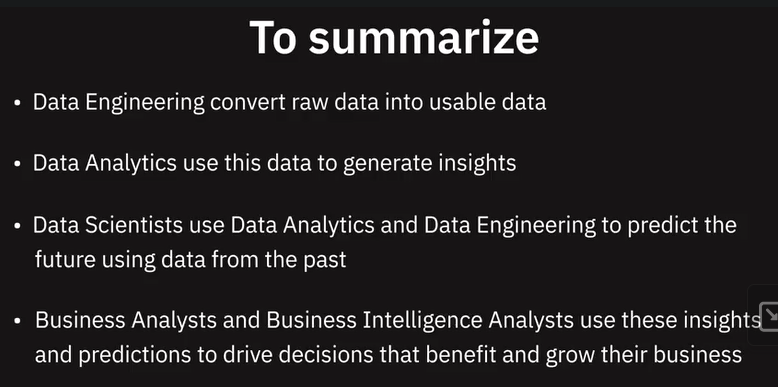


BI Analysts do the same, except their focus is on the market forces and external influences

that shape their business. They provide business intelligence solutions by organizing and monitoring data on different business functions and exploring that data to extract insights and actionables that improve business performance.



To summarize in simple terms,



Interestingly, it’s not uncommon for data professionals to start their career in one

of the data roles and transition to another role within the data ecosystem by supplementing

their skills.