**The Value of Data.**

**Introduction.**

Data is a term that's used a lot in today's business world, and there's a good reason for that.

Capturing, managing, and leveraging data is central to redefining customer experience and creating new value in almost every industry.

In this module, I'll start with the basics.

What is data, and what's its role in the digital transformation of your business?

Then I'll discuss how you can leverage data in your organization.

Next, I'll break down the types of data, and finally, I'll go through some important data considerations for every business using data in the cloud.

Let's get started.





**The role of data in digital transformation.**

Let's start by asking a simple question: what is data? Data is any information that is useful to an organization.

Imagine numbers on a spreadsheet or text in an email.

These are both examples of data.

Other examples include audio or video recordings, images, and even just ideas in employees' heads.

Businesses now have access to data like never before.

This includes internal information, data from inside your organization, and external information, customer and industry data.

For example, as organizations have digitized their operations, all kinds of business data has become available, such as financial information, logistics data, production output, and quality reports.

Businesses also have access to new kinds of data about their customers.

Consider digital interactions such as the length of time a user spends on a web page or reaction to a social media post.

These are totally new and very rich sources of information about customer behavior.

The Internet has also increased access to external data, such as industry benchmarking reports.

Capturing and leveraging this data to unlock business value is central to digital transformation.

Large enterprises with traditional IT infrastructures face several limitations in leveraging the value of data.

These limitations include processing volumes and varieties of new data, either at regular time intervals known as batch or in real time.

Finding cost-effective solutions for setting up and maintaining data centers.

Scaling resource capacity up or down, and regulating their capacity globally, especially during peak demand times throughout the year.

Accessing historical data that is often stored in different formats and on different platforms.

Deriving insights from historical and new data in time and cost-effective ways.

Public cloud services like Google Cloud offer organizations economies of scale, rapid elasticity, and automation where there was manual overhead.

They allow organizations to bring together data points and platforms fragmented across their whole ecosystem.

In particular, the cloud provides data solutions that were once almost impossible.

Businesses can now consume, store, and process terabytes of data in real time and run queries at its request to retrieve and use data instantly.

Resources are now distributed across a global network.

This means that multiple data centers can create resilience against data loss or service disruption, but without any extra overhead for businesses.

And data can be combined, analyzed, and served to business teams quickly and cost-effectively.

For the first time in many businesses, this means data insight is highly accurate and accessible across the business, and now an enabler of transformation.

Let's look at some examples of organizations that have transformed their business by unlocking the value of data.

Budget airlines don't provide food as part of their service.

Instead, they charge customers for meals if they want that.

This may seem like a cost-effective solution, but it's often difficult to estimate the number of meals required onboard.

If the airline overestimates the number of meals required, they risk wasting food and losing revenue.

But if they underestimate the number of meals needed, they risk selling out of food, providing poor customer service, and losing potential revenue.

One budget airline in Asia embraced digital transformation and reimagined how they could solve this problem using data.

First, they identified factors to help estimate stock such as the size of the plane and the number of passengers, but they soon discovered that estimates based on these factors were not highly accurate.

This meant that they had to think about their data differently.

So they analyzed additional information such as destination, time of day, and flight connections before and after their journey.

Using this information, they uncovered actionable insights.

For example, they learned that of the total vegetarian and non-vegetarian meals required on each flight, flights to and from India required 73% more vegetarian meals.

With these new insights, the airline was able to predict the amount of meals required more accurately.

As a result, they provided a more positive customer experience and improved the profitability of their food service.

Let's take a look at another example.

Traditional retailers have access to a range of data about their stores, including stock levels, items purchase, and average spend per customer.

However, they've never been able to capture information about more nuanced in-store customer behavior.

One video security company noticed this problem in the retail sector and reconsidered how they could use their existing technology and data to overcome it.

Traditionally, security monitoring systems were used for one main purpose: to detect criminal behavior in stores.

But what if they could reimagine the purpose of this technology?

By using cloud computing to mine data from video cameras and devices, this company was able to generate insights on customer retail footpath, sentiment, and dwell time.

This means that businesses can now correlate data on shopper behaviors in the store to improve safety, operational efficiency, and top-line growth.

Manufacturing is another great example of an industry that is using data and embracing digital transformation.

Companies and high-speed manufacturing industries such as pharmaceuticals, food and beverage, and consumer packaged goods require continuous production.

They can't afford any downtime because that can significantly impact revenue, customer experience, and product quality.

In these always on-manufacturing environments, maintaining production health is key.

One technology company helps businesses perform vital monitoring of their manufacturing and production lines.

They do this by combining IoT, Internet of things devices, with manufacturing and analytics.

With cloud computing, they analyze historical data and live information generated by sensors to assess machine health, predict maintenance, and ensure that production lines are always running.

We'll talk more about IoT in the next module.

These are just a few examples of how cloud technology can unlock new value by reimagining data.

No matter where you are in your company, you too can leverage data to solve challenges.

In the next video, I'll discuss how you can start to think about and map the data in your organization to uncover new business value.

**Leveraging data in your organization.**

So far you've learned that leveraging data is a critical part of digital transformation.

And no matter what your role is, you can unlock new business value by thinking about data in new ways.

A helpful starting point is to identify and map your data.

Let me explain.

A data map is a chart of all the data used in end-to-end business processes.

For example, imagine that you own a chain of apparel retail stores.

What might you include in your data map?

A customer purchases an item in one of your stores.

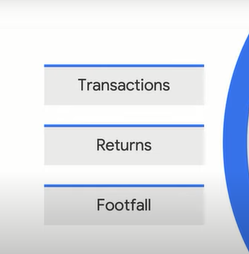
That's a data point.

If you aggregate that with other purchases across stores in each region, you have a type of data: transactions.

We call this a dataset.

Another dataset might be item returns.

Another is footfall in your stores.



You may have noticed that all of these datasets are about your users.

**User data is therefore your first data bucket**. This category contains all data from customers who use or purchase your services and products.

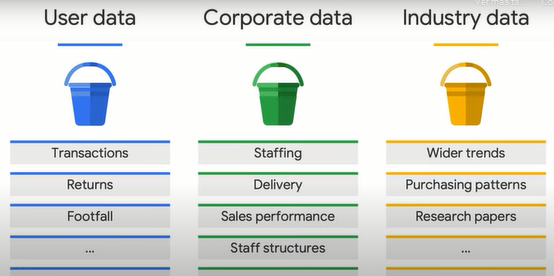
Now let's think about data that is more operational. For example, data about staffing levels in each store, stock delivery dates, overall sales performance of each store, and store staffing structure. That's, how many people are in fitting rooms versus at the cash register.

**These fall into the second bucket of data: corporate data.** Corporate data includes data about the company such as sales patterns and operations. Can you think of more datasets that you'd have as the retail owner?

A third category serves as the umbrella for user and corporate data.

We'll call it **industry data.** Industry data is data found outside an individual organization that everyone in the sector needs to view or access to gain knowledge about a specific domain. This could include wider trends, purchasing patterns, and publicly available research papers.

**These three buckets make up your data map**.



As you add more datasets to each bucket, you'll eventually gain richer insights.

But how can I sort these datasets further?

Well, I can use a check mark to identify the datasets I currently have and a question mark to label the datasets I think I can get.

Now your data is mapped out with a list of the different datasets you have or think you can get.

How can you make your data actionable?

Start playing with the intersections between your datasets.

Take two or more datasets and ask yourself, "What insight could I gain if these datasets were combined?"

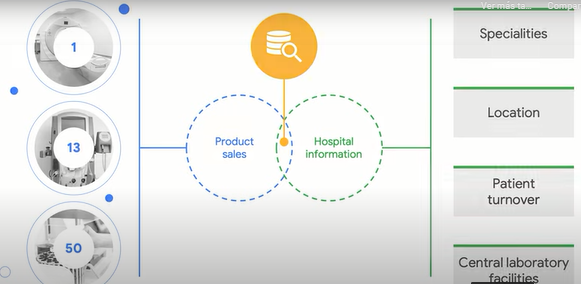
For example, suppose you're a sales manager at a biomedical diagnostics company.

You provide a range of diagnostic tools to hospitals and laboratories across the region.

One of your datasets is sales of each product.

You also have some datasets about the hospitals themselves, including specialties, location, patient turnover, and central laboratory facilities.

In this case, by integrating how many products are sold at each hospital with all the other datasets you have for the hospital, you have a clear insight into what makes your ideal customer.



Here's another example.

If you took introduction to digital transformation with Google Cloud, you might remember Jane.

She's a personal banker.

What does her data ecosystem look like?

Well, user datasets might include user demographics, user financial history, and previous user interactions.

Corporate datasets might include sales by financial product, sales conversation call logs, and performance metrics of financial portfolios.

Banking is a heavily regulated industry, so a lot of industry data is available, including industry benchmarks.

Another set of industry data is stock performance and other investment trends.

Jane, her colleagues, and her competitors at other banks all have access to this data.

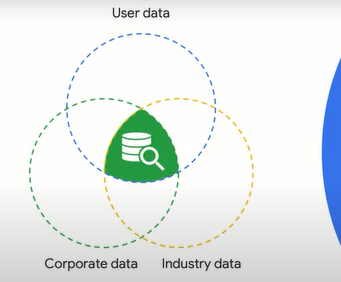
By integrating her user demographics data with product purchases, she can begin uncovering demographic indicators that are significant for predicting product purchases.

Now it's your turn.

Take a piece of paper and write user data, corporate data, and industry data.

Think about what dataset you have in each.

Then consider how different datasets can be combined to create valuable insights.



In the next section, we'll explore the different data modalities and why they matter in digital transformation.

**Understanding data types.**

We've talked a lot about different data formats throughout the course, such as images, audio files, and social media interactions, as well as tabular data like sales figures.

Most importantly, we've uncovered how you can combine and leverage data in new ways to revolutionize business models and create new value.

To better understand how to combine this data throughout your digital transformation, let's explore the two types of data and what they mean for businesses.

We can categorize data in two main types: s**tructured and unstructured.**

Structured data is highly organized.

Examples include customer records consisting of names, addresses, credit card numbers, and other quantitative data.

Structured data can be easily stored and managed in databases.

By contrast, unstructured data has no organization and tends to be qualitative.

Examples of unstructured data can include word processing documents, audio files, images, and videos.

This data can be stored as objects.

An object consists of the data itself, a variable amount of metadata, and a globally unique identifier.

Some unstructured data can be stored in a format called a BLOB.

This stands for Binary Large Object.

Images, audio, and multimedia files can all be stored as BLOBs.

Organizations rely on both structured and unstructured data to gain insight and make intelligent decisions.

However, unstructured data has historically been very difficult to analyze.

Cloud technology changes this.

With the right cloud tools, businesses can extract value from unstructured data by using an application programming interface, or API, to create structure.

An API is a set of functions that integrate different platforms with different types of data so that new insights can be uncovered.

Let's look at an example of a used car dealership that was able to leverage both structured and unstructured data.

Whenever customers brought cars into the dealership, the agents had to manually upload and label photos of each car and then set a price depending on the model and condition.

This process took them, on average, 20 minutes per car.

In this case, how could the car dealership use automation to make this process more efficient?

Well, they decided to develop a custom train machine learning model, a topic I'll cover more in detail in an upcoming module.

To combine unstructured data, the photographs, with structured data, they used car prices.

They could then use this combined data to predict the price value of a vehicle.

With this new approach, the overall process to photograph and evaluate a car has now dropped from 20 minutes to 2 to 3 minutes per car.

That's a 95% reduction in time and a massive improvement in overall service.

Another example is Bloomberg.

Bloomberg aggregates data-driven news, global insights, and expert analysis from over 2,700 journalists in more than 120 countries.

They also make their content, their structured and unstructured data, available in multiple languages worldwide.

This is inevitably a big challenge.

Their customers expect up-to-date and accurate news that's accessible in their specific languages.

So how do you translate the data you're receiving and then localize it into your global audiences, all in real time?

Well, Bloomberg uses Google Translate API, which enables them to make financial news, global insights available to their customers in as many as 92 languages.

Using an API to translate content into multiple languages is a common way that businesses apply the power of the cloud to unstructured data, creating new value while saving time and costs.

More importantly, they can reach more customers and provide a better, more personalized service.

Understanding structured and unstructured data can help you define what's possible with the data solutions you have.

We'll explore this in the next module when we cover data consolidation and analytics.

But first, any conversation about data needs to include a reference to security, privacy, compliance, and ethics.

We'll cover that in the next section..

**Important data considerations.**

Capturing, storing, and analyzing vast amount of data is key to adopting Cloud technology.

But handling this volume and diversity of data comes with its own ethical considerations and requires alternative ways of thinking about security.

Google believes that capturing and managing data demands responsibility and accountability.

Not all information that can be captured should be captured.

In other words, businesses are accountable for making responsible decisions about which data they collect, store, and analyze.

This also extends the data that businesses already own.

In this case, it's essential to examine who has access to the data and how they'll be using it.

First, consider the source of the data, how it's being collected, and where it's stored.

If it's personal or sensitive data about a customer or an employee, it needs to be securely collected, encrypted when stored in the Cloud, and protected from external threats.

Additionally, only a subset of users should be granted permission to view or access the private data.

Data security and privacy becomes more complex in a global economy.

Regional or industry-specific regulations often guide data policies.

Google Cloud offers a range of solutions and best practice resources that companies can leverage.

Another consideration is whether all the data is relevant and appropriate.

Let me explain where this can be particularly important.

Suppose, for instance, you want to use thousands of lung X-ray images to train an ML model to automatically identify tumor markings in new patient X-rays.

What you need are the X-ray images.

This is the relevant data.

What's not relevant is patient's personal data.

You need to ensure that any source data about individuals such as names or addresses is omitted or redacted.

There's also some information that is not personally identifiable and should still not be included in the modeling for ethical reasons.

A good example of this is whether or not individuals have health insurance.

This is not relevant for educating the model to identify tumors.

And if the data is included, the solution could be discriminatory.

These ethical and privacy considerations are particularly complex when you're working with unstructured data.

For example, a customer support team that resolves hundreds of customer's issues a day via mail might want to use an automated tool to find patterns in the email passages and develop targeted solutions.

It's true that emails contain valuable data that can be mined to solve this challenge, but it's essential to be conscious of protecting customer privacy at the same time.

Ethical and fair considerations are particularly important and applicable when you work with artificial intelligence, AI, and machine learning.

We'll cover ML and important factors to consider in a later module.

For now, I want you to remember that human bias (*prejuicios, parcialidad*) can influence the way datasets are collected, combined, and used.

Because of this, it's always important to include strategies to remove unconscious biases as you start to leverage data to build new business value.

In this module, you learned the importance of data in digital transformation.

Unlocking the value of data enables a business to both rethink how they serve their customers and reimagine how they operate.

Ultimately, using data effectively enables any business, large or small, to better achieve its mission.

Move on to the next module to learn about data consolidation and analytics.

How you store and manage your data affects what you can do with it.

So in the next module, we'll examine the challenges, solutions, and use cases for different data consolidation and storage systems on-premises or in the Cloud.

**Quiz.**

1. What are the key benefits of using cloud technology to unlock value from data, especially for traditional Enterprises? Select the two correct answers.

Customers can collaborate with corporations to create industry trends.

Customers can now gain access to their own data instantly.

**Businesses can query their data and retrieve results instantly.**

Businesses can access open source data like never before.

**Businesses can process terabytes of data in real-time.**

2. Lucinda is creating a data map for her online learning company. Her datasets include learner demographics, their purchases, and browsing history. What data 'bucket' would these datasets fall into? Select the correct answer.

Industry data

Corporate data

Cloud data

**User data**

3. Eduardo is using a machine learning model to improve recruitment efficiency for his company. What candidate data is appropriate and relevant for training the model? Select the two correct answers.

Gender

**Years of experience**

Ethnicity

Address

**Education**

4. Mark owns a large pharmaceutical company that manufactures essential medical supplies. The production lines are required to operate efficiently at all times. How can Mark use cloud technology to ensure his production lines are meeting optimal performance requirements? Select the correct answer.

Evaluate historic data to inform new product development

Evaluate real-time data to monitor competitor landscape

Evaluate consumer feedback to identify customer sentiment

**Evaluate real-time data to predict maintenance requirements**

5.Images and videos are examples of what type of data? Select the correct answer.

Semi-structured

Structured

Organized

**Unstructured**