

# Innovating with Data and Google Cloud

## The value of Data

Data → any info useful to an organization.

### The role of data in DX

Capturing and leveraging data to unlock business value is central to DX

Google Cloud offers

Economies of scales | Rapid elasticity  
Automation | Data access

### Leveraging data in your organization

#### DATA MAP

Chart of all the data used in E2E business processes

i. e. User Data, Corporate Data, Industry Data

#### Making Dataset actionable?

Take 2 or more datasets and ask "what insight(s) could I gain if these datasets were combined"

### Types of data

#### Structured

i. e. customer record containing name, address, and other quantitative data

#### Unstructured

i. e. audio files, images, videos

some unstructured data can be stored in format called BLOB

BOLB = Binary Large Object

Cloud Technology made analysis of unstructured data easier.

With right cloud tools, business can extract value from unstructured data by using an API

### Important considerations for using data in the cloud

Capturing & managing data demands responsibility and accountability

#### 1. Consider source of data

How is it being collected?  
Where is it stored?

If it is personal or sensitive data, It needs to be securely collected, encrypted when stored in Cloud

#### 2. Ethical & fair considerations

Important for AI/ML

#### 3. Human Bias

It can influence the way datasets are collected

## DATA SECURITY & PRIVACY

# Innovating with Data and Google Cloud

## Data Consolidation and Analytics

### Migrating data to the cloud

When data is stored on-premises:

Things to consider: IT infrastructure, maintenance, capacity control...

Risks: Downtime, Dissatisfied user

Cloud = space rent from public cloud provider like Google Cloud

Data storage and compute power is elastic

Speed at which user ingest & use data increased  
(business can now ingest data in real-time)

### Databases, data warehouses and data lakes

#### Database

Organized collection of data, generally stored in tables and accessed electronically from computer system

i. e. Cloud SQL, Cloud Spanner

#### Data integrity & scale

Data integrity, or transactional integrity, refers to the accuracy and consistency of data stored in a database. Data integrity is achieved by implementing a set of rules when a database is first designed and through ongoing error checking and validation routines as data is collected

With Cloud, businesses can build & deploy faster, deliver transformative applications, maintain portability and control on their data

#### Data warehouses

Assembles data from multiple sources, including databases  
Built to enable rapid analysis and multidimensional datasets  
Converts Unstructured data into semi-structured data  
Links storage & computer together

Cloud data warehouse allows businesses to consolidate data that is structured and semi-structured.

When combined with connector tools, data warehouses can transform unstructured data into semi-structured data that can be used for analysis.

I e. Big Query : serverless

This transformed data can then be sent directly from Dataflow to BigQuery, where it is made immediately available for analysis.

Pub/Sub and DataFlow can bring unstructured data into the cloud and transform it into semi-structured data.

#### Data lakes

Store structured, semi-structured, Unstructured data

Repositories for raw data

Hold for backup data  
(Businesses protected against data loss)

#### Cloud storage benefit

(1) any amount of data

(2) low latency

(3) accessible from anywhere

### Business Intelligence solutions

**Looker** Google Cloud business intelligence solution

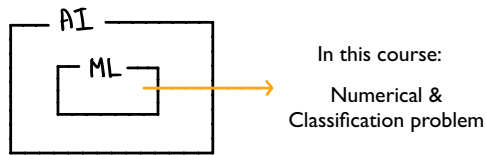
it's a data platform that sits on top of an analytics database and makes it simple to describe user data and define business metrics

# Innovating with Data and Google Cloud

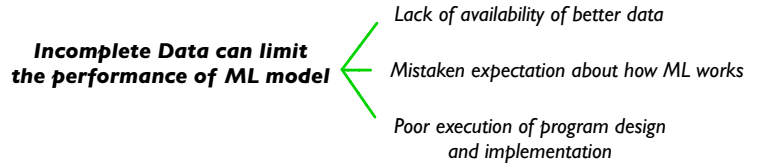
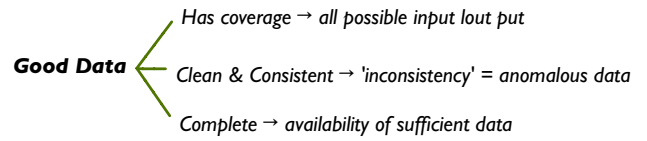
## Innovation with Machine Learning

### The definition of ML and AI

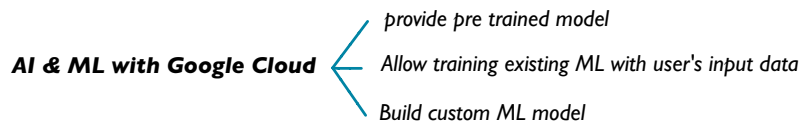
ML is a way to use standard algorithms or standard models to analyze data in order to derive predictive insights and make repeated decisions at scale.



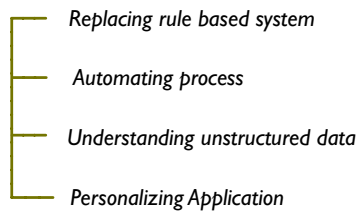
### Data quality considerations



## Real world ML use cases



### Real-world Use cases



OPEN SOURCE  
**Tensor Flow**

Google Cloud made TPU available in the Cloud with Compute Engine Virtual Machine

**API** simple methods & tools to connect various applications