Data on GCP

INTRODUCTION TO GCP

Nabeel Imam
Content Developer at DataCamp

datacamp

Data-driven culture

- Data is an asset
  - Drives operations and strategies
- From intuition to evidence-based decisions

Data-driven culture

The role of the cloud

- Vast storage in the cloud
- Real-time processing and analysis
- Responsive and dynamic business practices

Decoding data types

Structured

	Column Title	Column Title	Column Title	Column Title
First Row	250.05	45	230	95.65
Second Row	320	784.65	25	370
Third Row	1560	1570	2875	876.15
Fourth Row	895.25	375	485	685
Fifth Row	85	125	375	65
Sixth Row	275.7	570.35	85.35	185
Total	**3386**	**3470**	**4075.35**	**2276.8**

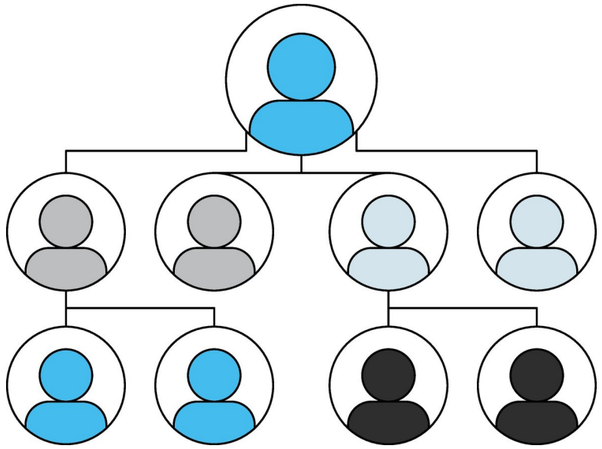
- Row-column table structure
- Follows rigid schema
  - Hard to scale
- **Google Cloud SQL**

Unstructured

- Multimedia files and data objects
- No common structure ("non-relational")
  - Flexible and easy to scale
- **Google Cloud Storage**

Semi-structured data

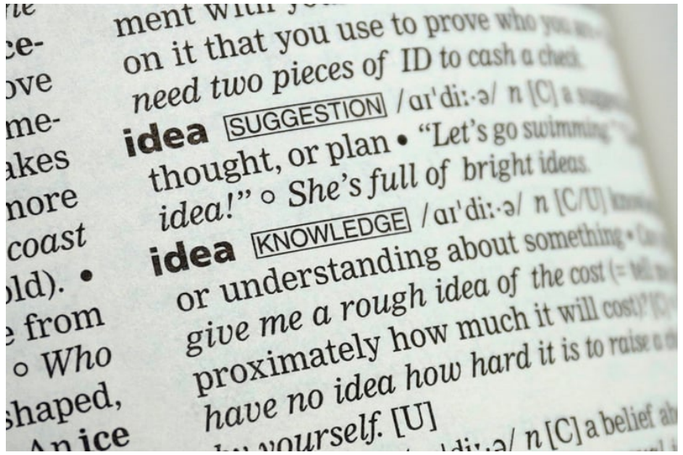
Trees



- Non-relational but semi-structured

Key-value pairs

- One key links to one or many values
  - E.g. the English dictionary



- One key could have one or multiple values

Semi-structured data

Flexible, scalable, and fast

- Used to store
  - Emails and social media posts
  - Insurance claims
  - Health records

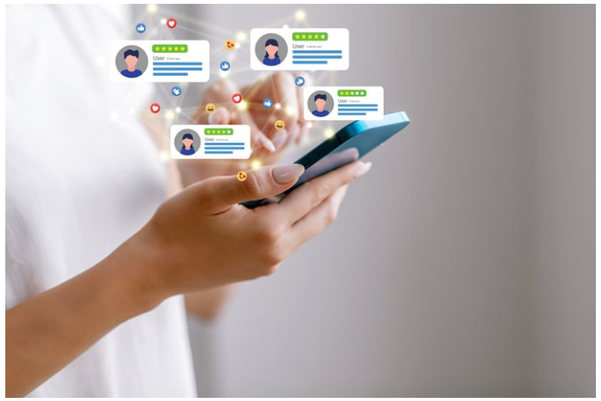
GCP Bigtable

GCP in retail

GCP in retail

Coud SQL: inventory and sales

Bigtable: social media posts and profiles



Cloud Storage: multimedia files

Let's practice!

INTRODUCTION TO GCP

Smart analytics

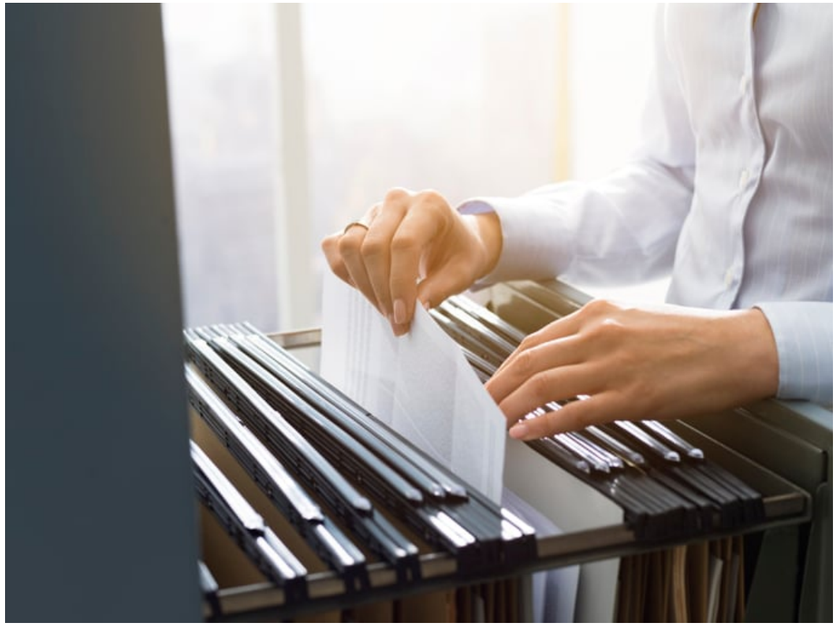
INTRODUCTION TO GCP

Nabeel Imam
Content Developer at DataCamp

datacamp

Databases

- Store structured data in tables
- Provide easy access and editing



GCP's Cloud SQL

The global bank

The infinite database

Cloud Spanner

- Unlimited scale
- Supports multiple read write operations at the same time

Analytic needs

How do we run analytics on big databases?

- Data warehouses collect, sort, and collate data
  - Get it ready for analysis

Analytics in the warehouse

- Centralized updates for fast, consistent analytics
- Optimized for reading, complex queries; less suited for writing or data entry

Analytics with GCP

BigQuery

- Fully managed serverless warehouse
- Optimized for complex analytics
- Scales up automatically

Looker

- Business Intelligence platform
- Interactive visualizations and real-time analytics
- Dashboards with BigQuery

Data lakes

- Centralized store for structured, semi-structured, and unstructured data

Data lakes

- Offer high flexibility
  - Central repository for diverse data types
- Handle petabytes from sources like social media
- Enable big data analytics and machine learning

- **GCP Data Lake service:** scalable, cost-effective solution for data from various sources
- **GCP BigLake:** integrates BigQuery to data lakes

Let's practice!

INTRODUCTION TO GCP

GCP's compute services

INTRODUCTION TO GCP

Nabeel Imam
Content Developer at DataCamp

Engines and functions

Server-based

Compute Engine

- Customizable virtual machines

Serverless

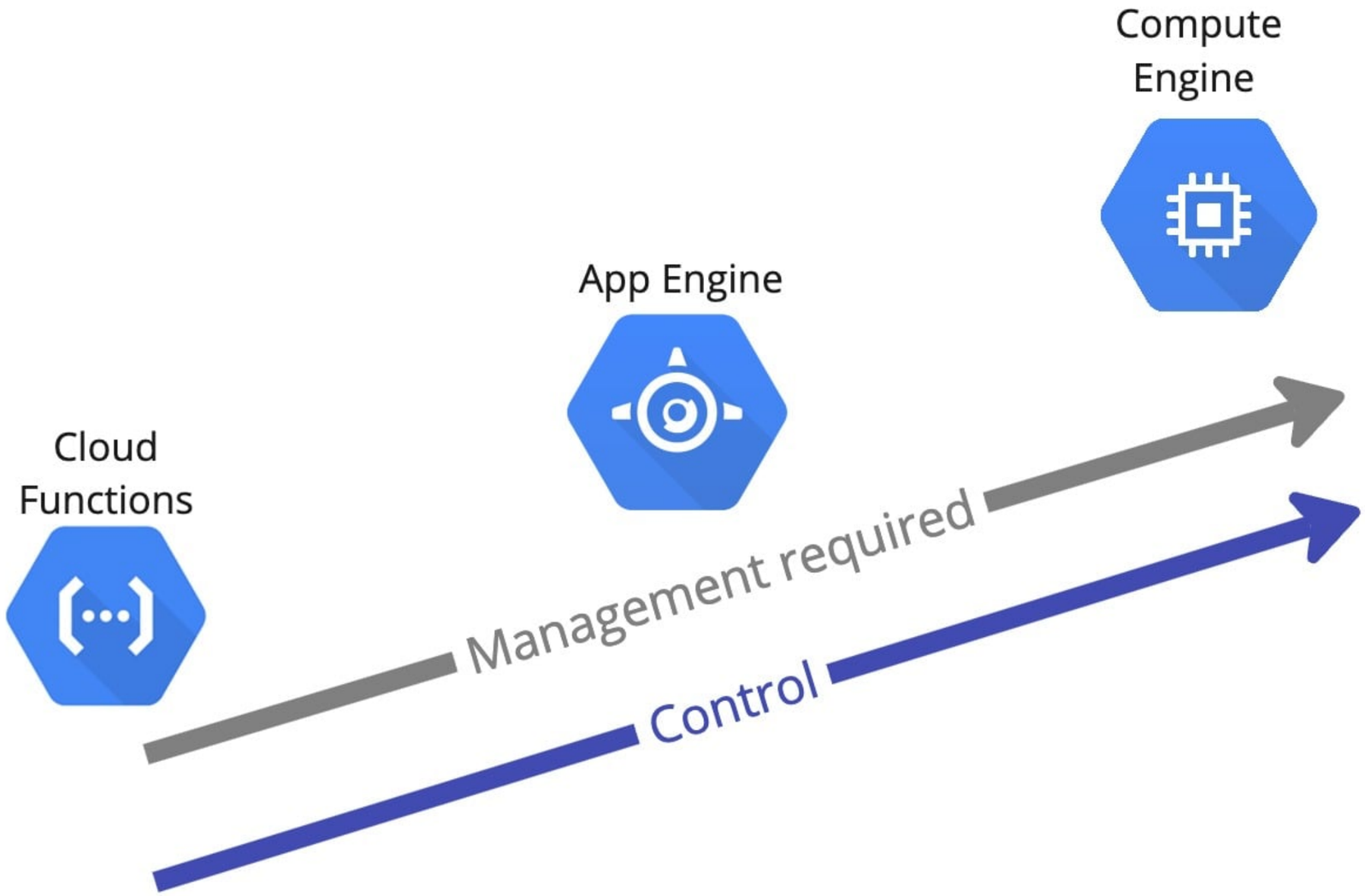
App Engine

- Scalable and efficient app-development

Cloud Functions

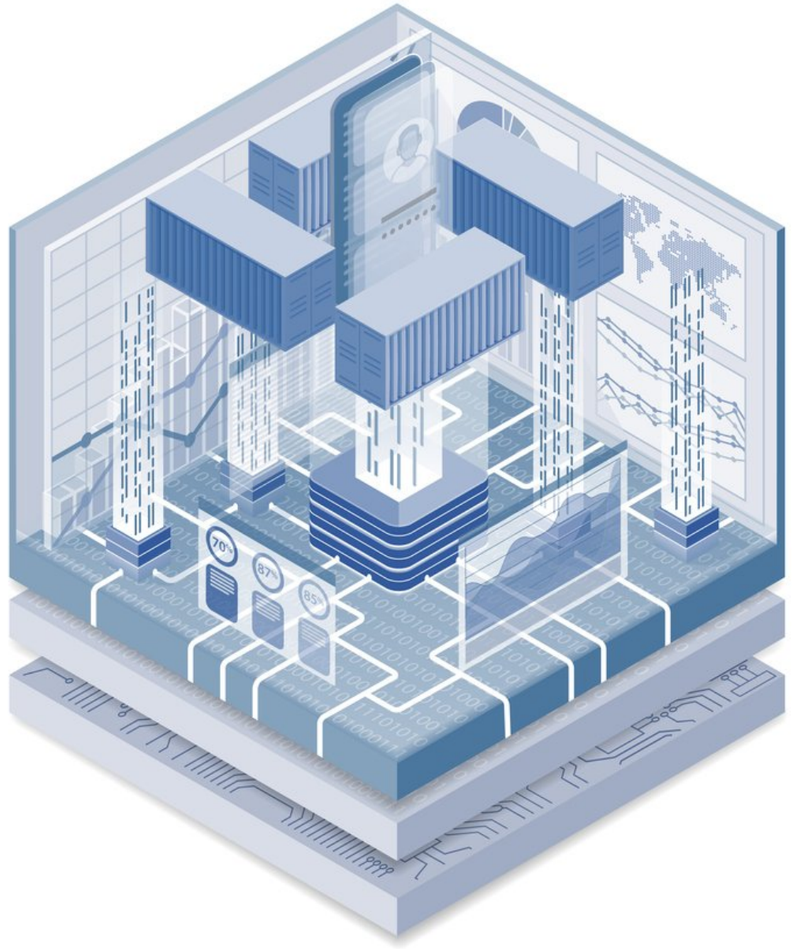
- Event-driven computation

Serverless spectrum



Microservices in containers

- Containers: packages containing all essentials (libraries, code, data)
- Each containerized app provides a microservice
  - Together microservices make up the service



The photo-sharing app

An upload surge

- Increase number of containers for upload service
- Trigger an update of recommendations service

Orchestration is crucial...

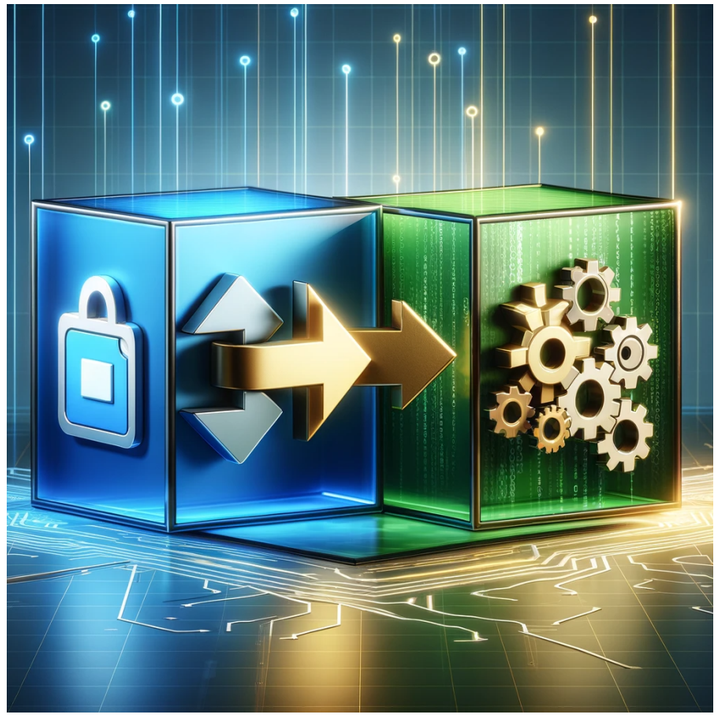
Kubernetes

- Open-source software for container orchestration
- Invented by Google

Google Kubernetes Engine (GKE)

- Fully managed Kubernetes service
- Automated scaling and updates
- Integrated with other GCP services

Handling the upload surge



- GKE scales up the containers for uploads
  - Scales them down once surge goes away
- Uploads microservice sends a message to recommendations

Complicated architectures

- Data on premises for regulatory reasons
- Compute services on the cloud

Bridging platforms

- Deploy apps across various environments
  - GCP, on-premises, other cloud providers
- Simplified operations, increased compliance
- Apps must be container-based
  - Facilitates move to container architecture

Anthos

Let's practice!

INTRODUCTION TO GCP

AI and machine learning

INTRODUCTION TO GCP

Nabeel Imam
Content Developer at DataCamp

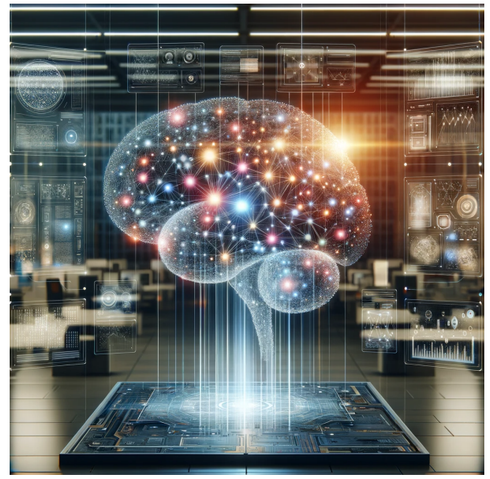
datacamp

The world of AI and ML

Artificial intelligence (AI)

- Designed to mimic human thought processes
  - Learning, problem-solving, and decision-making

Machine Learning (ML)

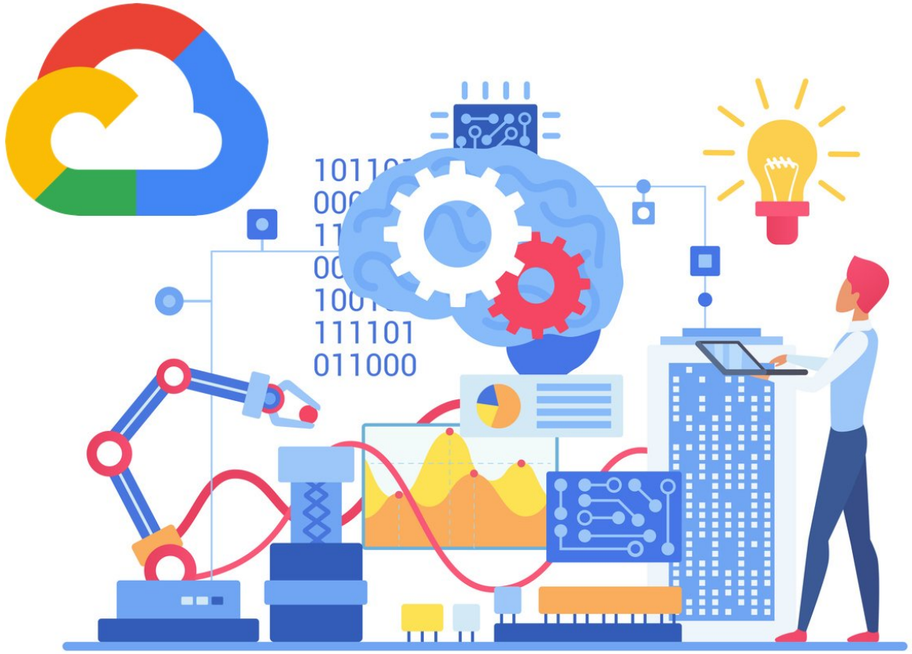


- A key branch of AI
- Machines learn from data to make decisions or predictions
- Machines do not need to be explicitly programmed

Data quality in ML

- ML algorithms ingest and train on data
  - "Garbage in, garbage out"
- High quality data:
  - Clean
  - Comprehensive
  - Relevant
  - Up-to-date

GCP's AI and ML Solutions



- Low latency, high throughput network for deployment
- A suite of AI and ML services

Pre-trained models on GCP

Vision AI

- Labels images
- Detect objects
- Read hand-written documents

Video AI

- Recognize places
- Extract video metadata

Pre-trained models on GCP

Natural Language AI

Enable businesses to implement AI without development and maintenance

- Extract insights from texts
  - Find the topic of a long post
  - Sentiment of a review
  - And much more...

ML with BigQuery

- BigQuery ML training with SQL
- Seamless transition from analytics to AI/ML insights

Vertex AI

- Integrated AI/ML platform
- Build, train, and deploy on one platform

Vertex AI

AutoML

- No-code ML solutions
- Trained on user's data
- Available for tabular, image, text data, among others

Custom training

- User control and flexibility
- Model fine-tuning and customization

AI in business

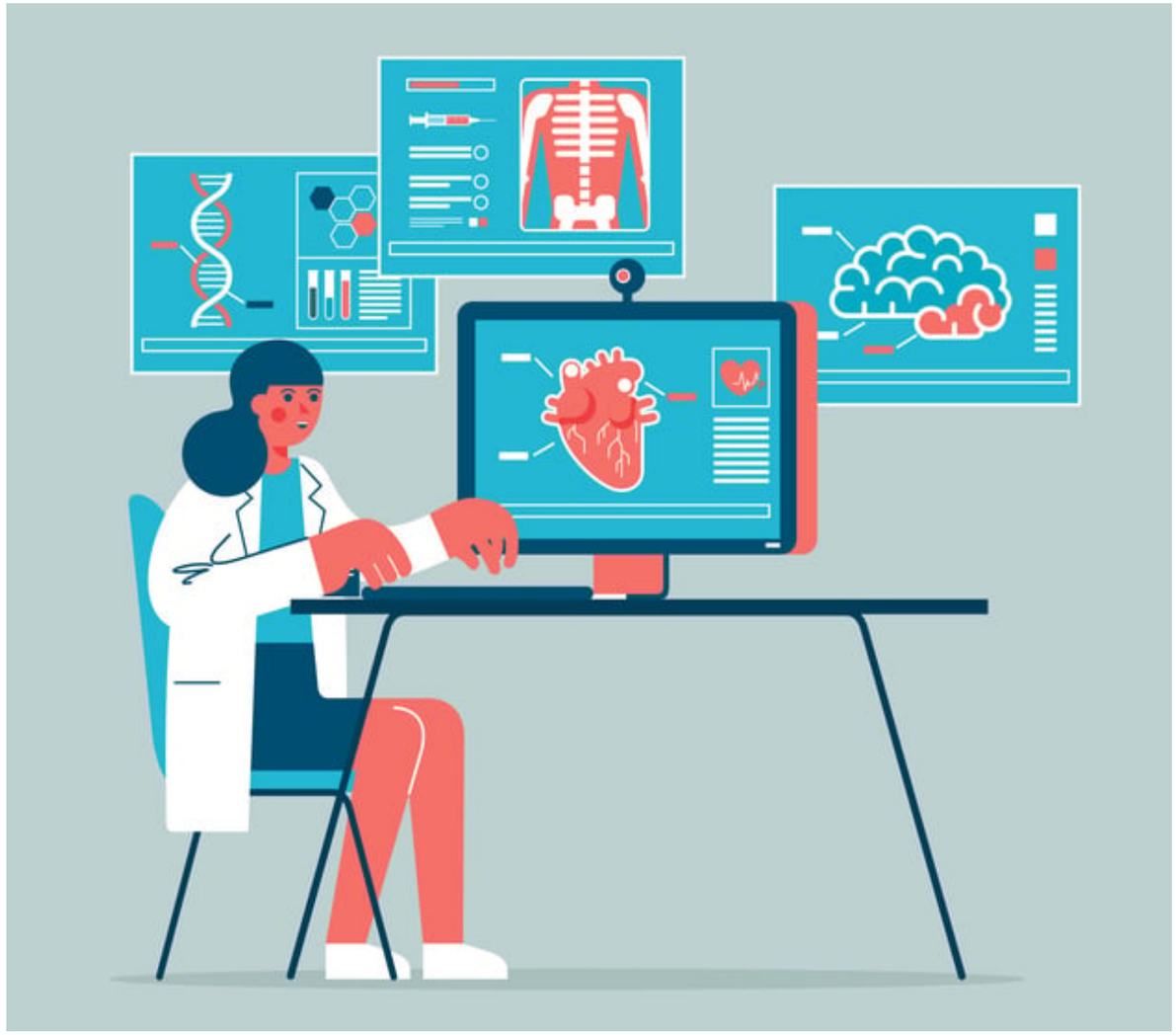
Customer churn model in retail

- Built on BigQuery ML
- Deployed on Vertex AI

AI in business

Provisional diagnoses with AutoML

- Trained on images through AutoML
- Label images as "Disease" or "Healthy"



AI in business

Stock recommendations

- Recommendation model through custom training on Vertex AI
- Train model on diverse sources of data
- Fine-tuned for each customer's needs

Let's practice!

INTRODUCTION TO GCP