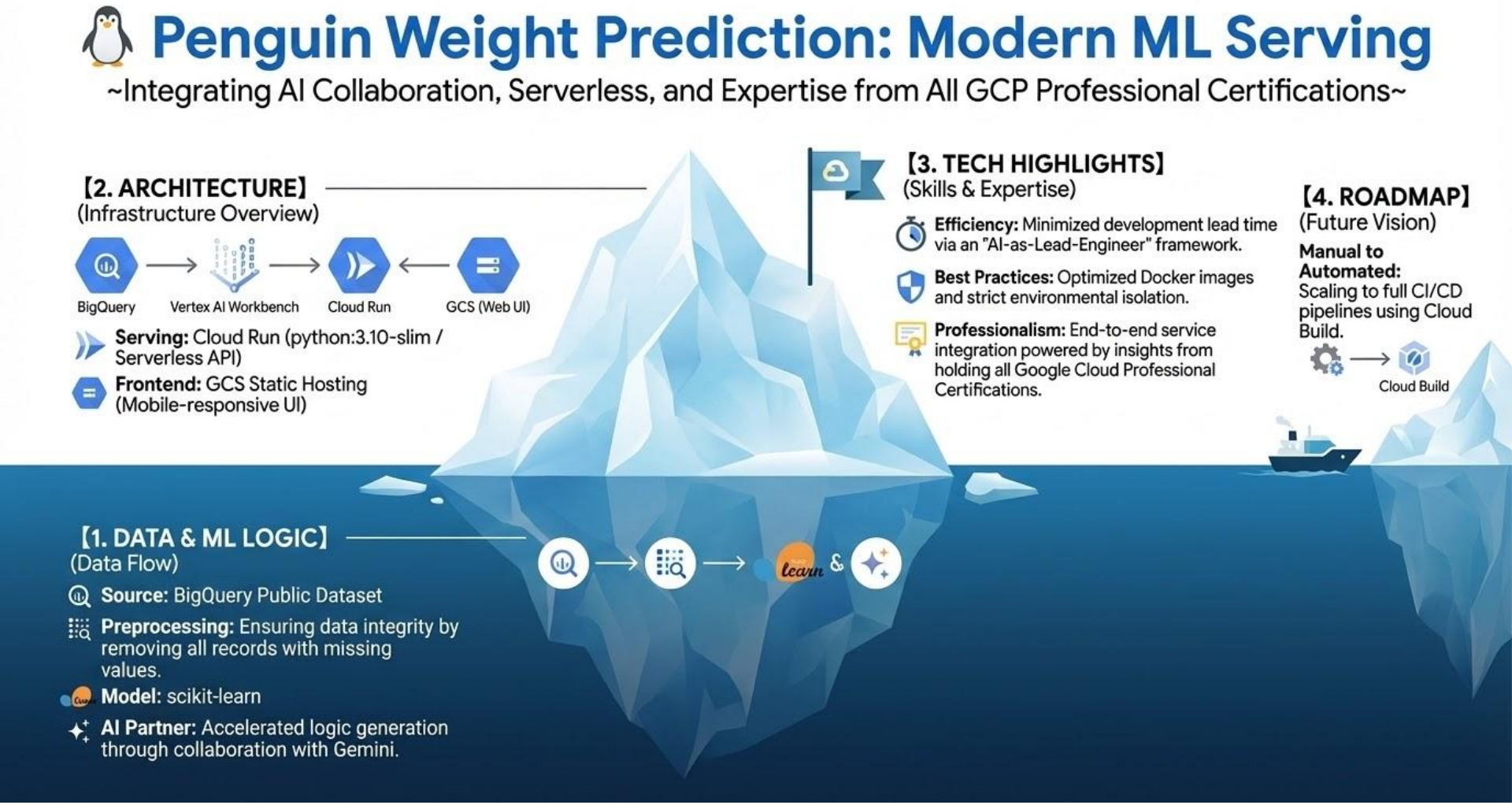
https://storage.googleapis.com/ml-test001-publicfiles/index.html



Project Overview & Architecture Design

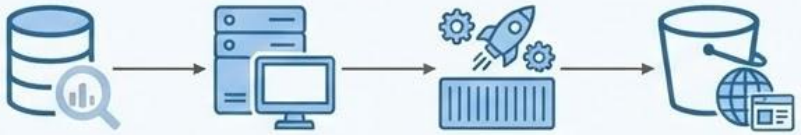
— Building Serverless ML Serving via AI Collaboration —

Project Objectives

Web Deployment of Practical ML Demo: Publicly hosting a machine learning demonstration using the BigQuery public dataset.

Proof of Concept for AI-Augmented Development: Utilizing AI (Gemini) as a co-development partner to validate methods for minimizing development lead time from conception to deployment

System Architecture



Data: BigQuery (ml_datasets.penguins)

Development Environment: Vertex AI Workbench (JupyterLab)

Inference API: Cloud Run (Python / Docker / scikit-learn)

User Interface: Google Cloud Storage (Static Website Hosting)

Data Strategy & AI Collaboration Process

— Ensuring Data Integrity & Model Validation —

【Data Engineering】

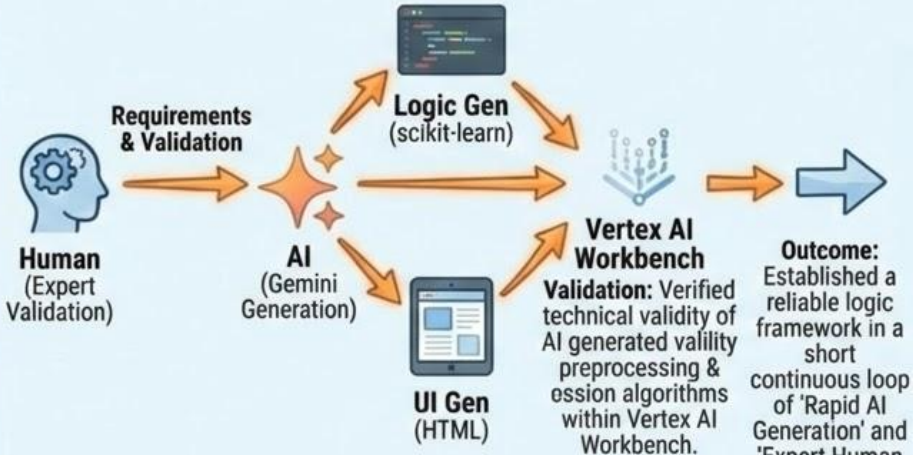
Dataset Features: Species, Island, Physical Characteristics (Bill length/depth, Flipper length), Sex, and Body Mass.

Target Variable: Body Mass (g)

Features: All variables excluding Body Mass.

Handling Missing Values: Prioritized stability of prediction accuracy by removing all records containing missing values to ensure a clean dataset.

【AI-Human Co-Development Workflow】



Deployment, Integration & Roadmap

【Container-Based Deployment】

Cloud Run Deployment: Adopted a lightweight container strategy using python:3.10-slim as the base image.

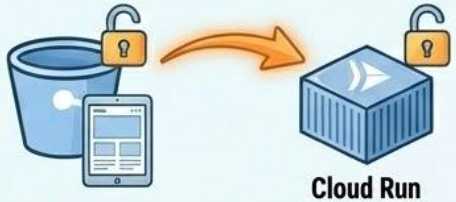
Build Optimization: Implemented Docker best practices, such as using –no-cache-dir, to optimize image size and build efficiency.

AI-Assisted Workflow: Validated the Dockerfile configuration proposed by Gemini through successful end-to-end deployment testing.

【Service Integration & Accessibility】

Access Control: Prioritized accessibility for this demonstration by configuring Cloud Run and GCS buckets for public access (allUsers).

Service Integration: Successfully established a seamless communication flow between the GCS-hosted UI (HTML) and the Cloud Run inference API.



GCS Bucket

Cloud Run
Container instance

【Roadmap: Evolving from Workbench to MLOps】

Current State: Successful implementation of manual deployment and model development within Vertex AI Workbench.

Next Action: Transition from manual operations to an automated CI/CD pipeline leveraging Cloud Build for continuous deployment.

Current vs. Future

Vertex AI Workbench

Automated CI/CD

Summary: Key Outcomes & Core Competencies

— A Message to Recruiters —

Skills Demonstrated Through This Project

Deep Understanding of the GCP Ecosystem: Proven ability to select and integrate **optimal services**—from data extraction to web deployment—to build a cohesive solution.

Modern Development Methodology: Established an 'AI-Augmented' workflow, leveraging Gemini as a lead engineer to maximize development velocity and efficiency.

Practical Engineering Foundation: Applied comprehensive knowledge gained from holding all Google Cloud Professional Certifications to implement ML development (BigQuery/Workbench), containerization (Docker), and access management.

Demonstrated end-to-end ownership by designing a **mobile-friendly UI** (HTML), drawing on previous front-end development experience.