**PROJECT: AI-Driven Personalized Health Plan Recommender**

**Background / Problem Statement**

* The Healthcare Insurance and Wellness Provider (Blue Cross Blue Shield) faced challenges in delivering **personalized healthcare plan recommendations** to existing and prospective members and internal employees.
* Multiple machine learning models were developed but remained siloed and **difficult to operationalize** due to:
  + Lack of **standardized MLOps pipelines**.
  + No centralized control over **data versioning, experiment tracking, and deployment automation**.
  + Manual workflows led to **deployment delays**, lack of traceability, and high operational overhead.
  + The absence of **automated CI/CD pipelines for training, inference, and monitoring** resulted in delayed insights, inconsistent recommendations, and low scalability across the provider network.

**UST Solution (Built Using AWS SageMaker MLOps AI Workbench)**

* UST designed, developed, and deployed a **Hybrid Recommendation System** combining:
  + **Content-Based Filtering** (Plan attributes and patient preferences)
  + **Collaborative Filtering** (User behavior patterns)
  + **Ranking Model** (Optimizing recommendation relevance)
* The solution was built using **AWS SageMaker MLOps Template-based Architecture**, incorporating:
  + **CI/CD-driven Training Pipelines**: Automated model retraining triggered by data changes using SageMaker Pipelines, GitHub, EventBridge.
  + **DataOps & Data Versioning**: Leveraging S3, SageMaker Ground Truth, and AWS Glue for clean, versioned, and governed data pipelines.
  + **CI/CD Inference Pipelines**: Auto-deploy multi-algorithm inference via SageMaker endpoints inside secure VPCs behind Load Balancers.
  + **CI/CD Model Monitoring Pipelines**: Using SageMaker Model Monitor and Clarify for drift detection, explainability, fairness, and performance insights.

**Business Impact**

* Achieved approximately **21% improvement** in healthcare plan recommendation acceptance rate through a hybrid ML model combining content-based, collaborative filtering, and ranking techniques.
* **Accelerated time-to-market**, reducing model deployment duration from several weeks to **under 2 hours** using fully automated CI/CD pipelines built on SageMaker Pipelines and GitHub Actions.
* **Enhanced member and employee experience** by delivering more accurate, timely, and personalized healthcare plan suggestions.
* Ensured **regulatory compliance and ethical AI** by integrating SageMaker Clarify for **bias detection**, **model explainability**, and **transparency**.
* Delivered **scalable real-time inference** capabilities supporting millions of users simultaneously using **multi-model SageMaker endpoints** hosted securely within VPCs.

**Cost Impact**

* Realized **~30% reduction in infrastructure costs** by leveraging AWS **Spot Instances**, SageMaker **model lifecycle management**, and **on-demand autoscaling** of endpoints.
* Reduced **ML Ops overhead** by automating training, inference, and monitoring pipelines using AWS **EventBridge triggers** and SageMaker workflows.
* Lowered **data processing and storage expenses** by adopting **versioned S3 data lakes**, **serverless data pipelines**, and **cost-effective data labeling** using SageMaker Ground Truth.
* Eliminated repetitive **manual intervention costs** through full CI/CD automation across the MLOps lifecycle—training, deployment, and model monitoring.

**AWS SageMaker MLOps Benefits**

* Standardized CI/CD templates ensured **repeatable and governed deployment**.
* Version-controlled data/model lineage helped with regulatory compliance (HIPAA).
* Event-driven retraining and pipeline orchestration led to **resilient production ML workflows**.

**Summary of AWS SageMaker MLOps Template-based Components Used**

* **Data Pipeline**: S3, Ground Truth, EventBridge, AWS Glue – *for DataOps, Governance, and Triggering Pipelines*
* **Training Pipeline (CI/CD)**: SageMaker Pipelines + GitHub + CodeBuild
* **Inference Pipeline (CI/CD)**: ECR-based container deployment to SageMaker Endpoints in private VPCs
* **Monitoring Pipeline (CI/CD)**: SageMaker Model Monitor + Clarify → metrics to CloudWatch + alerts via EventBridge