**VPBank Technology Hackathon 2025 – Senior Track**

General Brief

Please fill up this table and use this document as a template to write your proposal.

| **Challenge Statement** | Write your Challenge Statement Title here |
| --- | --- |
| **Team Name** |  |

**Team Members**

Full Name

Role

Email Address

School Name

(if applicable)

Faculty / Area of Study

LinkedIn Profile URL

Member 1

Member 2

Member 3

Member 4

Member 5

**Content Outline**

Page No.

Solutions Introduction

Impact of Solution

Deep Dive into Solution

Architecture of Solution

**DRAFTING (XÓA TRƯỚC KHI NỘP)**

* **Bài toán & Mục tiêu**
  + **Hệ thống khuyến mãi truyền thống chạy batch chậm, thiếu linh hoạt, hiệu quả thấp.**
  + **Mục tiêu: Nền tảng dữ liệu khuyến mãi hiện đại, cho phép business tự thiết kế/cấu hình campaign (no-code), hỗ trợ batch + near real-time + real-time, cá nhân hóa đúng lúc.**
* **Giá trị kinh doanh**
  + **Tối đa hóa Net Profit Uplift, giảm lãng phí nhắm nhầm Sure Things/Lost Causes.**
  + **Rút ngắn time-to-market (no-code rule changes), tăng hiệu quả chiến dịch.**
  + **Explainability minh bạch cho từng quyết định (lý do đủ điều kiện).**
* **Năng lực cốt lõi**
  + **Business User Configurability: Rule Builder UI (kéo-thả/form), versioning, audit, publish/rollback tức thì.**
  + **Rule Engine + Guardrails + Optimizer: Lọc eligibility → xếp hạng theo uplift → tối ưu ngân sách/ràng buộc.**
  + **Explainability (core): Trả về reason codes/SHAP top-K, threshold\_met theo rule.**
  + **Xử lý dữ liệu end-to-end: Ingest, cleaning, enrichment, transformation; structured + unstructured.**
* **Kiến trúc tổng thể (đầy đủ)**
  + **Experience & Config**
    - **Campaign Configuration UI (AppSync GraphQL + Cognito)**
    - **Rule DSL (JSON/YAML), workflow phê duyệt, versioning**
  + **Rule Store & Cache**
    - **DynamoDB: RuleSets (versioned, compiled predicates)**
    - **ElastiCache Redis: Cache rule/feature, leaderboard, session (TTL)**
  + **Decisioning & ML**
    - **API Gateway → Lambda Decision Service (Provisioned Concurrency)**
    - **Rule Engine runtime (Lambda/EKS container)**
    - **Uplift Model Endpoint (SageMaker Endpoint; TreeSHAP inline)**
    - **Guardrails (caps/frequency/DNC) + Knapsack Optimizer (ngân sách/ràng buộc)**
  + **Feature Platform**
    - **SageMaker Feature Store: Online (low-latency) + Offline (S3/Parquet)**
    - **S3 Data Lake (raw, curated), Glue Catalog**
  + **Streaming & Events**
    - **Amazon MSK (Kafka) hoặc Kinesis (streams + Firehose)**
    - **Stream processing (Lambda/Glue Streaming) cập nhật aggregates**
  + **Batch & Analytics**
    - **Glue/EMR Serverless (Spark) cho cleaning/enrichment/transforms**
    - **Amazon Redshift (DWH) cho lịch sử campaign, phân tích, dashboard**
  + **MLOps & Orchestration**
    - **Step Functions: train/evaluate/deploy, Model Cards, approvals**
    - **SageMaker Model Monitor: data/concept drift → retrain**
  + **Monitoring & Alerting**
    - **CloudWatch Dashboards + Alarms (API P95/P99, Lambda errors, Endpoint latency)**
    - **Redshift (QueryDuration, WLMQueueLength), Redis (CacheHitRate)**
    - **SNS thông báo sự cố**
  + **Bảo mật & Tuân thủ**
    - **IAM least privilege, KMS, log bất biến, audit trail rules/models**
* **Luồng dữ liệu chính**
  + **Batch: S3/Redshift → Glue/EMR (clean/enrich/aggregate) → Offline FS → Scoring → Rule evaluate → Output Redshift/S3.**
  + **Real-time: MSK/Kinesis event → Aggregates/Redis → API GW → Decision Lambda (Rule → Uplift → Guardrails/Optimizer) → Explanation JSON → Notification/Leaderboard → Redshift.**
* **Data Handling (structured + unstructured)**
  + **Structured: profiles, balances, transactions; time windows 7/30/90d; velocity/RFM/preferences.**
  + **Unstructured: call transcripts/email/social/clickstream JSON → embeddings/sentiment/topics → Feature Store; raw lưu S3.**
* **Explainability (MVP)**
  + **Response gồm: decision, uplift\_score, explanation.primary\_factors (SHAP/reason codes), threshold\_met. P95 < 100ms (TreeSHAP cho tree-based; fallback reason codes + SHAP async nếu cần).**
* **Performance & Benchmarks**
  + **Batch: 1M giao dịch ≤ 6h; E2E ≤ 8h; scale đến 10M/ngày (EMR Serverless).**
  + **Real-time: P95 < 100ms, P99 < 150ms; ≥ 10k RPS; ≥ 100k concurrent sessions.**
  + **Ingest: MSK/Kinesis ≥ 50k events/s; Feature Store writes ≥ 10k rec/s.**
* **Adaptability (đa dạng khuyến mãi, no-code)**
  + **Cashback %, fixed; Tiered rewards; Sales contest (leaderboard); Challenge; Flash sales; Merchant/category campaigns.**
  + **Rule change qua UI, lưu JSON/YAML ở DynamoDB, hiệu lực ngay (cache invalidation), versioning/audit.**
* **Demo Scenarios (để trình bày)**
  + **Scenario 1: Batch-driven Cashback (Dining ≥ 5M/tháng)**
    - **Dữ liệu: Redshift → Glue/EMR → FS → Scoring → Rules → Output + Explanation**
    - **Mục tiêu: 1M txn ~ 6h; báo cáo eligibility + reason**
  + **Scenario 2: Real-time Sales Contest (flash 2h, top 100)**
    - **Dữ liệu: Kafka/MSK → Decision (<100ms) → Leaderboard (Redis) → Notification**
    - **Mục tiêu: P95 < 100ms; refresh leaderboard ≤ 1s**
* **Tech stack**
  + **Languages: Python, Node.js/TypeScript**
  + **AWS: API Gateway, Lambda, SageMaker (Feature Store/Endpoint/Monitor), Step Functions, S3, Glue/EMR, Redshift, DynamoDB, ElastiCache Redis, MSK/Kinesis, CloudWatch, SNS, Cognito, AppSync**
  + **Containers: Docker; EKS cho rule/inference nâng cao khi cần**
* **Agile delivery (rút gọn)**
  + **Sprint 0–3 (pilot): baseline E2E, Rule UI, realtime + batch demo, monitoring.**
  + **Ceremonies 2 tuần; DoR/DoD ML/Data/Infra; KPI: Profit@K, ROI, latency, error rate.**

**Solutions Introduction**

Uplift Engine is a modern, configurable promotion data platform that turns promotions into profit by targeting persuadable customers—those who convert because of the treatment. It combines a no-code Rule Engine (for business users) with Causal AI uplift modeling, low-latency decisioning, and transparent explanations.

**Tiêu chí đánh giá thành công dự án (mở rộng)**

1. Ba Ràng Buộc Cốt Lõi (Triple Constraint) – “Cái Tam Giác”

* Phạm vi (Scope): Sản phẩm/dịch vụ/kết quả đúng yêu cầu đã định nghĩa.
* Thời gian (Time): Hoàn thành đúng hạn; yếu tố ít linh hoạt nhất.
* Chi phí (Cost): Hoàn thành trong ngân sách được phê duyệt.
* Lưu ý: Nhà tài trợ thường xếp hạng ưu tiên giữa Phạm vi–Thời gian–Chi phí để đội dự án cân bằng phù hợp.

1. Các Tiêu Chí Thành Công Mở Rộng

* Sự hài lòng của Khách hàng/Nhà tài trợ: Chỉ số cảm nhận, mức độ tin tưởng và chấp nhận.
* Đạt được Mục tiêu Chính: Tạo/tiết kiệm tiền, ROI, hoặc outcome đã cam kết.

1. Yếu Tố Quan Trọng cho Thành công Dự án CNTT

* Quản lý các Bên liên quan (Stakeholder Management): Xác định–hiểu–đáp ứng kỳ vọng; bỏ qua stakeholder dễ dẫn đến thất bại.
* Sponsorship cấp cao (Executive Sponsorship): Yếu tố thành công hàng đầu; bảo trợ/ra quyết định giải toả bế tắc.
* Sự tham gia của Người dùng (User Involvement): Đồng thiết kế, phản hồi sớm; đảm bảo phù hợp nghiệp vụ.
* Nguồn lực & Chuyên môn (Resources & Expertise): Vốn, con người đúng kỹ năng là nền tảng cho tiến độ/chất lượng.
* Kỹ năng mềm & Giao tiếp (Soft skills & Communication): Truyền thông rõ ràng, thường xuyên; giảm hiểu nhầm/rủi ro.
* Phương pháp Quản lý Dự án chính thức: Tổ chức vận hành tốt đạt tỷ lệ thành công cao vượt trội.

Tóm lại, thành công dự án CNTT không chỉ là “đủ tam giác” (Scope–Time–Cost) mà còn bao gồm sự hài lòng của khách hàng/bên liên quan và giá trị kinh doanh thực nhận (ROI). Nghiên cứu thực tiễn cho thấy nếu chỉ bám ba tiêu chí truyền thống, tỷ lệ thành công bền vững rất thấp; do đó cần bổ sung tiêu chí mở rộng và các yếu tố then chốt ở trên.

**Key features:**

* Business-user configurability: Campaign Configuration UI (React/Amplify) + AppSync GraphQL + DynamoDB for versioned rules; publish/rollback without deployments.
* Explainability-by-design: Every decision includes human-readable reason codes (e.g., “Customer spent 5M VND in Dining this week, reached Gold tier”). TreeSHAP for tree models; async SHAP fallback for complex learners.
* Full-spectrum processing: Batch (Glue/EMR), near real-time **message-queue** (Kinesis/MSK + Lambda), and real-time (<100ms) via API Gateway + Lambda + SageMaker **Feature Store** + Endpoint + Redis cache.
* AWS-native, **serverless-first**: Lambda (Provisioned Concurrency), Step Functions, SageMaker (Feature Store, Training, Endpoints), DynamoDB, Redshift, ElastiCache Redis, Kinesis/MSK.
* Optimizer & guardrails: **Knapsack optimizer** for budget allocation; do-no-harm guardrails (DNC, confidence lower bound, frequency capping).

**Impact of Solution**

**Business impact:**

* Focus squad/product: Initial pilot with Credit Cards squad (or CASA) to maximize measurable **Net Profit Uplift** and speed up adoption across squads.
* **Higher ROI, lower waste:** Focus budget on Persuadables; avoid Sure Things/Lost Causes; reduce wasted spend (up to ~70%) and improve ROI (simulated +308%), while preserving brand trust via guardrails.
* **Faster time-to-market:** Business users change rules without code; publish instantly with audit/versioning.
* Personalization at scale: Real-time eligibility + uplift scoring with P95 < 100ms; near real-time streaming updates for contests/challenges.

**Why this solution is better:**

* Causal AI advantage: Optimizes net profit uplift, not propensity—aligns with true business outcomes.
* Consistency by design: SageMaker Feature Store removes training-serving skew; same features across offline/online.
* Operational excellence: Redis caching, observability with CloudWatch, and performance targets/benchmarks included.

**Differentiators (USP):**

* Explainability as a core MVP feature in inference (reason codes/SHAP).
* No-code Rule Engine tightly integrated with uplift/optimizer.
* Tiered architecture (Glue/EMR) for cost-performance balance, SERVERLESS, SCALABLE

**Deep Dive into Solution**

**End-to-end flows: INPUT DATASOURCE – UI INTERACTION**

* Data ingestion & features: Structured (profiles, transactions) and unstructured (text/clickstream) data handled via Glue/EMR pipelines; point-in-time correctness enforced; features stored in SageMaker Feature Store (Offline/Online).
* MLOps: Step Functions orchestrates feature jobs, parallel training (CatBoostUplift, DR-Learner, CausalForest), best model selection by Profit@K, registration/deployment to SageMaker Endpoint.
* Real-time decisioning: API Gateway → Lambda (Provisioned Concurrency) → Online Feature Store (GetRecord) → Uplift Endpoint → Guardrails → Optimizer → Response with explanation; exposures logged to Kinesis/MSK and fed to Redshift for analytics.
* Rule Engine: Business UI → AppSync → DynamoDB (draft/publish) → Lambda CompileRule → Redis cache of compiled rules; Decision Lambda evaluates eligibility in ~1–5ms via Redis.
* Streaming & near real-time: Kinesis/MSK updates aggregates and triggers eligibility/contest logic (leaderboards in Redis/Redshift).

Contract examples:

* Request (client → API): { customerId, context }
* Response: decision, offer, uplift\_score, uplift\_std\_error, explanation.primary\_factors, threshold\_met

**Demo scenarios (prototype-ready):**

* Batch cashback: Weekly/monthly cap + rule threshold; batch evaluation via Glue/EMR; outputs to Redshift with explanations.
* Real-time sales contest: First N customers meeting spend within a time window; streaming updates + real-time rule evaluation and leaderboard refresh.

**Architecture of Solution**

**AWS services and how they’re used:**

* API Gateway + Lambda (Provisioned Concurrency): Low-latency decision API; cold-start eliminated.
* SageMaker Feature Store: Offline (S3/Parquet) for training/analytics; Online (DynamoDB) for <10ms feature reads.
* SageMaker Endpoints: Real-time uplift inference; TreeSHAP inline where applicable; autoscaling by RPS/CPU.
* AppSync + Cognito + S3/CloudFront (UI): Business Rule Builder UI with auth; GraphQL mutations to author/publish rules.
* DynamoDB + Redis: Versioned rule storage and 1–5ms compiled rule cache; invalidation on publish/rollback.
* Kinesis/MSK + Firehose: Exposure/outcome streams; sinks to S3/Redshift for analytics/monitoring.
* Glue/EMR Serverless: Tiered batch/feature engineering pipeline with cost/performance optimization.
* Redshift: Analytics warehouse for campaign KPIs, eligibility batches, explainability logs, and historical attribution.

**Performance & Benchmarks:**

* Real-time decisioning: P95 < 100ms, P99 < 150ms (API Gateway + Lambda + Feature Store + Endpoint + Rule eval + Guardrails).
* Batch processing: ≥ 1,000,000 transactions in ≤ 6h on Glue/EMR Serverless; scalable to 10,000,000/day with autoscaling and partitioning.
* Streaming ingest: Kinesis/MSK ≥ 50,000 events/sec; Feature Store writes ≥ 10,000 records/sec with parallel writers.
* LATENCY
* VALIDATE: BO SUNG

Processing modes covered as required: Batch, Near real-time, and Real-time.

**Architecture diagram:**