HCLTech
Supercharging Progress™

Analyst Data CoPilot

Gen AI Use Case Draft Proposal

Contents

1. Analyst Data Copilot (MVP) 1

1 Overview of the Use Case 1

2 Scope Description 1

2.1 MVP (target for UAT) 1

2.2 High-level release cycle (post-MVP) 2

3 19-Week MVP Timeline 2

4 Dependencies 3

5 Key Assumptions 3

6 Timelines 4

7 Success Criteria 4

8 Technology Stack (GCP) 5

9 Squad Composition 6

9.1 HCLTech squad (core team) 6

9.2 Required ING stakeholders 6

1. Analyst Data Copilot (MVP)

1 Overview of the Use Case

This proposal outlines a pragmatic, deliverable-focused **MVP** we'll complete as a **co-innovation project** between HCLTech and ING. The solution is an **Analyst Data Copilot** designed to accelerate the initial investigation of **Financial Crime and Fraud Prevention (FCFP)** alerts.

- **Problem:** FCFP L1 analysts spend significant manual effort gathering and reviewing data from multiple systems, primarily FCRM, just to perform an initial triage. This repetitive, time-consuming process slows down case resolution.
- **Solution (MVP):** We'll build a **GenAI-powered solution** (an "AI Assistant") that provides two core functions:
    1. **Proactive Summarization:** When an analyst provides a Case ID, the Copilot will automatically gather the relevant data and present a concise, natural-language summary.
    2. **Interactive Q&A:** The analyst can then have a natural-language conversation with the Copilot to "interrogate" the case data (e.g., "What was the total value of outbound transactions?").
- **Value:** This focused MVP provides immediate value by automating "info gathering" *and* "info investigation." The solution will **not** perform any triage, create reports, or make proposals, ensuring it *aids* the analyst without interfering with other systems (like SATM). The goal is to deliver a solution ready for formal **UAT**.

---

2 Scope Description

To ensure deliverability within the agreed co-innovation budget, the scope is tightly defined.

2.1 MVP (target for UAT)

- **Functionality**:
    - A standalone web application where an analyst can input a Case ID
    - Data Source**:**
        - The system will be built and tested using a **high-fidelity mock dataset** (based on the FCRM/CDT data model)
    - The UI will provide:
        - A **proactive, natural-language summary** of the key facts from the mock data
        - An **"AI Assistant" chat interface** to ask follow-up Q&A about the loaded case data

- **Out of Scope for MVP**:
    - Any decision-making, triage, recommendations, or proposals (e.g., "Close" or "Escalate")
    - Automated report generation for case closure
    - Direct "write-back" integration to FCRM (e.g., no auto-closing cases)
    - Automated aggregation or creation of new alerts
    - Live data integration from FCRM API, DAP, Quantexa, or TMA (this will be planned for Release 2)

2.2 High-level release cycle (post-MVP)

- **Release 2 (Live data integration):** Connect the UAT-validated Copilot to a live ING data source (e.g., FCRM API or DAP), as decided by ING
- **Release 3 (Embedded integration):** Embed the Copilot as a native widget within the FCRM front-end UI
- **Release 4 (Multi-system data integration):** Expand the Copilot's "tools" to query additional systems like Quantexa and the TMA Tool
- **Release 5 (Agent-in-the-loop actions):** *Pending business and AI governance approval,* explore adding features for report generation or "write-back" capabilities

---

3 19-Week MVP Timeline

The timeline isn't driven by the complexity of the UI, but by the "non-visible" engineering, data, and governance work required to build a reliable and secure AI agent for a high-stakes compliance environment. Unlike traditional software, an AI MVP's effort is front-loaded in data and model validation, not feature development.

The key drivers consuming the co-innovation budget are:

1. **Data Modeling & Mocking (The "Data" Challenge):**
    a. **High-Fidelity Mocking:** Creating a mock dataset that accurately reflects the complexity and relationships of the **FCRM and CDT data model** is a significant engineering task. This is critical for building a tool that provides real value.
    b. **Data Transformation:** We will build the data pipeline logic to transform this complex data model into a format that a GenAI model can reliably understand for both summarization and Q&A.
2. **Copilot Development & Tuning (The "AI" Challenge):**
    a. **Dual-Function AI:** We are now building two distinct AI functions: a "one-shot" summarizer *and* a conversational Q&A (RAG) engine.
    b. **Reliability & Factual Accuracy:** The biggest risk is "hallucinations." A large portion of the timeline is dedicated to "validation cycles", rigorously testing the Copilot's summaries and Q&A answers against the mock data to ensure they are factually correct, accurate, and free of errors.
3. **Compliance, Governance & Explainability:**

a. **Audit & Explainability:** For UAT sign-off, the Copilot can't be a "black box." We must build robust logging and explainability layers so that every summary and answer is "transparent, auditable, and consistent."
b. **Risk Mitigation:** The timeline accounts for the effort to build deterministic "guardrails" that prevent the Copilot from inferring or implying decisions, ensuring it strictly adheres to its "information only" mandate.

4 Dependencies

Successful delivery requires a collaborative partnership with clear responsibilities.

4.1.1 Dependencies on ING

- **Product Owner:**
  - **Arthur Van Bronswijk** to act as the dedicated **Product Owner (PO)**, responsible for prioritizing the backlog, providing domain decisions and validating outcomes
- **Data Model (Critical Path):**
  - Provision of a comprehensive **data model and sample data** for the **FCRM Case Disposition Template (CDT)** via **Vikas (HCL) and Upendra**. This is the primary blocker to starting development.
- **SME Access:**
  - Access to **one to two** L1 FCRM analysts for **three to four** hours per week for requirements gathering (Sprints 1-3) and formal **UAT** execution (Sprints 6-8)
- **Project Setup & Onboarding:**
  - Onboarding of **1-2 HCLTech resources (e.g., Kush, Vikas)** to the ING network
  - Assistance from an onboarded resource (helped by Arthur) to **fill in the necessary forms** for project approval, GCP setup, and AI governance
  - Provision of an **Azure DevOps (ADO)** project board for tracking
- **Platform Access:**
  - A decision on whether the target environment will be ING's GCP, and provisioning of that **GCP UAT environment** when ready

4.1.2 Dependencies on HCLTech

- Provide the dedicated squad composition (see section 8) for the project's duration
- Manage the project budget and provide transparent burndown charts
- Create the high-fidelity mock dataset based on the data model provided by ING
- Lead the GenAI design, technical development, and deployment

5 Key Assumptions

- The project will commence development on HCLTech's side using a mock dataset. The connection to a live ING data source (FCRM API or DAP) will be planned for a subsequent release.
- The **FCRM/CDT data model** provided by ING (via Vikas (HCL)/Upendra) will be accurate and comprehensive enough to build a realistic and valuable mock dataset.

- ING has the necessary **Google Cloud Platform (GCP)** subscriptions (e.g., **Gemini Enterprise**) to enable the Vertex AI Gemini APIs for the project.
- The agreed service hours for this co-innovation project are dedicated to our (HCLTech's) squad effort. All ING-side efforts (PO, SMEs, platform engineering, compliance review) are separate.
- The scope of the MVP is strictly limited to the features described in Section 1.
- ING compliance and risk teams will be engaged (via the required forms) to align on requirements, enabling a timely review process that doesn't block the start of UAT.

6 Timelines

The proposed squad (Section 8) consumes ~160 hours/week. The total co-innovation budget of 3,000 service hours thus supports an ~18–19 week project.

- **Weeks 1–2 (Sprint 0):** Kick-off, **receive FCRM/CDT data model**, finalize MVP scope, set up ADO board, and **begin ING project/GCP setup forms**.
- **Weeks 3–8 (Sprints 1–3): (Data & Mocking)** Develop and validate the high-fidelity mock dataset. Build and test the data pipeline logic.
- **Weeks 9–14 (Sprints 4–6): (Copilot & Q&A)** Build the standalone Copilot web application (UI). Develop the core summarization and Q&A (RAG) functions. Begin iterative testing and tuning of AI outputs.
- **Weeks 15–17 (Sprints 7–8): (Validation & UAT) Formal UAT phase** (on the mock dataset). Support L1 analysts in test execution. Gather feedback, fix bugs, and finalize accuracy and explainability logs.
- **Week 18 (Sprint 9):** Final UAT report, demo of the validated Copilot, and handover of the final UAT-ready application, ready for deployment to ING GCP and connection to live data.

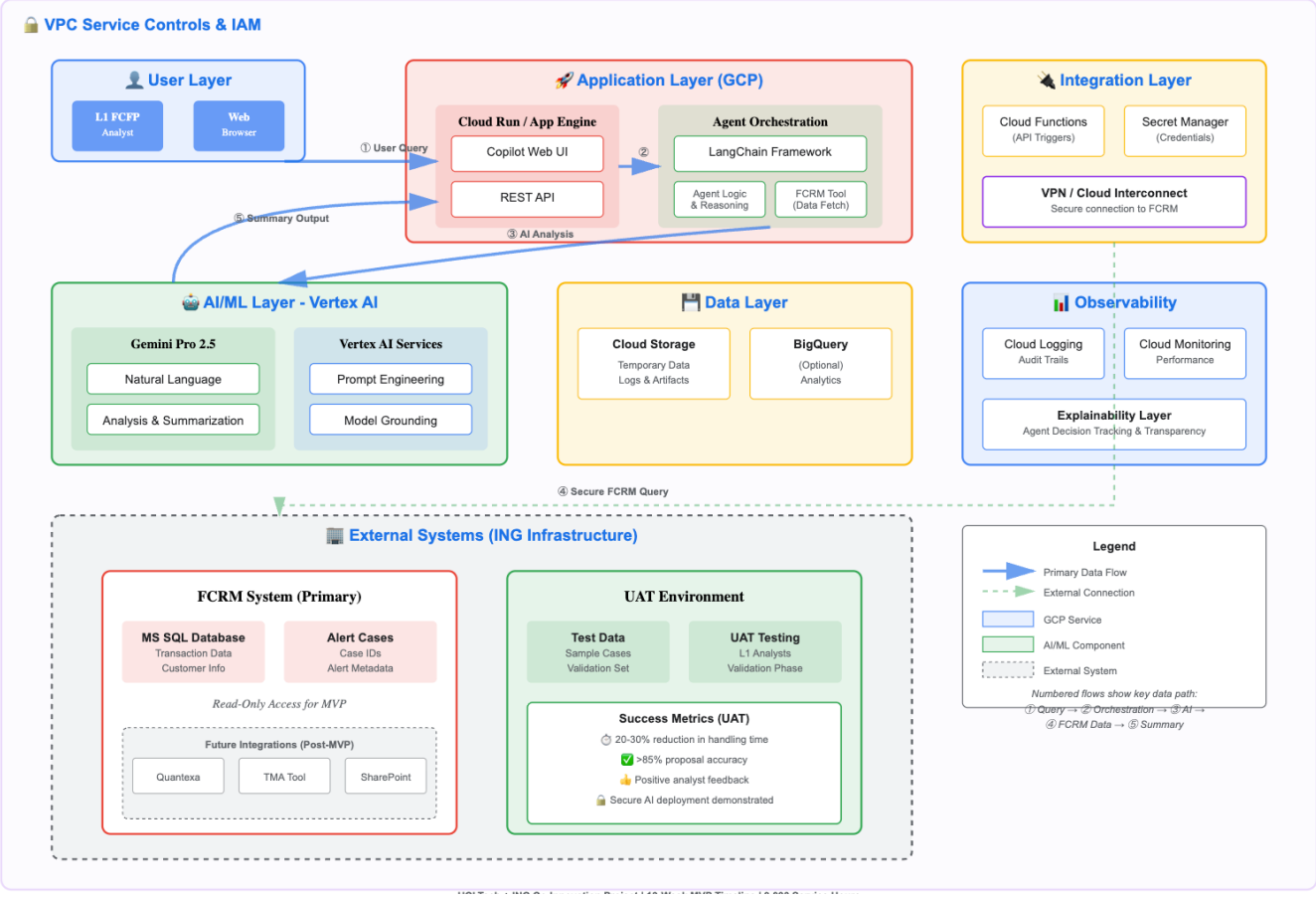
7 Success Criteria

- **Project success**
  - Deliver the defined MVP scope (on mock data) to a UAT-ready state within the agreed co-innovation budget
  - Have a clear path to deployment on ING GCP, with approval forms submitted
- **Quantitative success (to be validated during UAT)**
  - **Time Savings:** Achieve a **30%–50% reduction** in average *information gathering & investigation time*
  - **Factual Accuracy:** The generated summary and Q&A answers are **>95% factually correct** and free of hallucinations
  - **Query Relevance:** >90% of answers from the AI Assistant are deemed relevant and helpful by SMEs during UAT
- **Qualitative success (to be validated during UAT)**
  - Positive qualitative feedback from L1 analysts (e.g., "This is useful and saves me time")

8 Technology Stack (GCP)

Here is a high-level overview of the components. The architecture consists of a secure web front-end that communicates with a backend service, which in turn securely calls the data source and the Gemini large language model.

- **Analyst Interface (UI):** A simple, ING-branded web application hosted on **Google Cloud Run**.
- **Copilot Backend Service (Cloud Run):** The "brain" of the Copilot (formerly "Agent Orchestrator"). This service, built with a **GenAI Framework (like LangChain)**, will receive requests from the UI and orchestrate the calls to the data source and the AI model.
- **AI Model (LLM): Google Vertex AI**, accessing the latest **Gemini Pro** models.
- **Data Source (MVP):** A **mock database** (e.g., Cloud SQL or BigQuery) populated with the high-fidelity mock data.
- **Data Connector ("Tool"):** A secure API connector, built as a **Cloud Function**, that allows the Backend Service to query the data source (mock DB for MVP, FCRM/DAP for Release 2).
- **Security & Networking: VPC Service Controls**, **IAM**, and **Secret Manager** (for credentials).



9 Squad Composition

This squad is designed for a ~19-week (4.5-month) engagement, aligned to the total co-innovation budget.

9.1 HCLTech squad (core team)

- **One** Senior AI Lead / Architect (Full-time): Leads technical design, co-manages backlog with the ING PO and performs code reviews
- **Two** AI / Machine Learning (ML) Engineers (Full-time): Build the data pipeline, AI logic and application UI/API
- **One** Scrum Master (Part-time, 50%): Facilitates ceremonies, tracks burndown of the project hours and manages dependencies

(Estimated Total Burn: ~2,660 hours over 19 weeks, leaving a ~340-hour buffer for contingency within the 3,000-hour budget.)

9.2 Required ING stakeholders

- **One** Product Owner (Arthur Van Bronswijk, Part-time): Dedicated time for backlog grooming, sprint reviews and unblocking the team – 4 hours per week. There is no need to attend daily standups.
- **Two** L1 Analyst Subject Matter Experts (SMEs) (Part-time): For domain expertise (Sprints 1–3) and formal UAT execution (Sprints 7–8)
- **One** ING Platform Engineer (As-needed): To provision the GCP/UAT environments and configure the secure data access.

10 Key decisions

10.1 Key Decisions & Pending Approvals

Decisions Made: (14th of November)

- **Scope (No Triage):** The solution will **only aid the analyst**. It will *not* include any decisioning, triaging, or automated report generation for now.
- **Scope (Q&A Included):** The MVP scope **will include an AI Assistant** (Q&A) feature, allowing analysts to "interrogate" the case data.
- **"Agent" (Not Allowed):** The team is **not allowed to use an "Agentic AI"** solution in the autonomous sense, as ING's AI board has a hold on new agentic projects.
- **"GenAI" (Allowed):** Using **GenAI for summarization and Q&A is allowed**.
- **Data (Mock):** The MVP will be built using a **mock dataset**.
- **Platform (GCP):** If built inside ING, the solution will be on **GCP**.
- **Tech Stack:** The solution will use **LangChain** and deploy to **Cloud Run** and **Cloud Functions** within a **VPC**.

Decisions Pending (Blockers):

- **Environment (Inside vs. Outside):** Arthur must confirm whether the POC can be built **inside the ING estate**. This is the preference for all.
- **Data Source (Live):** The team must decide *if* and *how* to connect to a live data source (e.g., FCRM database, DAP) *after* the mock-data MVP is built.

HCLTech | Supercharging Progress™

hcltech.com