Proof of Concept Project Report

October 14, 2025

**Customer:** Enviable

**Development team:** Gintare Saali (lead), Natalia Kamysheva (developer)

**Disclaimer**

This proof of concept (PoC) is provided as a demonstration of potential functionality and is intended for evaluation purposes only.

It is the responsibility of the customer to thoroughly review and verify all relevant licenses, copyrights, and terms of use associated with any third-party models, libraries, datasets, APIs, or other tools included in or used to build this PoC before implementing, distributing, or commercializing it in any environment.

Contents

[1. Objective 4](#_Toc211348021)

[Primary Goal 4](#_Toc211348022)

[Problem Statement 4](#_Toc211348023)

[Solution 4](#_Toc211348024)

[Success Criteria Met 5](#_Toc211348025)

[2. Architecture Overview 5](#_Toc211348026)

[High-Level System Design 5](#_Toc211348027)

[Technology Stack 6](#_Toc211348028)

[3. Workflow Description 6](#_Toc211348029)

[End-to-End Process 6](#_Toc211348030)

[Typical Timeline 7](#_Toc211348031)

[4. Implementation Details 8](#_Toc211348032)

[Data Ingestion 8](#_Toc211348033)

[Discovery Engine 8](#_Toc211348034)

[Background Job Orchestration 10](#_Toc211348035)

[Real-Time Data Streaming 10](#_Toc211348036)

[Database Design 10](#_Toc211348037)

[Deployment Configuration 11](#_Toc211348038)

[5. Results 11](#_Toc211348039)

[6. Challenges and Limitations 14](#_Toc211348040)

[7. Recommendations 15](#_Toc211348041)

# 1. Objective

## Primary Goal

The Intelligent Building Analytics Platform demonstrates an AI-powered solution for proactive HVAC system monitoring in commercial buildings, automatically detecting anomalies and intelligently communicating findings to the right stakeholders in the right format—eliminating both manual analysis burden and notification fatigue.

## Problem Statement

Traditional building management systems collect extensive sensor data but face two critical challenges:

Technical Challenge: Lack of intelligent analysis. Facility managers must manually monitor multiple sensors, identify patterns, and correlate data across systems. This reactive approach leads to delayed equipment malfunction detection, energy inefficiency, and increased maintenance costs.

Communication Challenge: Generic alert fatigue. When issues are detected, systems send identical notifications to all stakeholders, from engineers to executives, resulting in information overload for managers and insufficient technical detail for engineers. This one-size-fits-all approach leads to most notifications being ignored.

## Solution

This PoC delivers an autonomous monitoring platform with intelligent stakeholder engagement:

* Ingests real-time data from Fan Coil 201 Unit via MQTT protocol
* Analyzes patterns using AI with 19 specialized analytical tools
* Automatically discovers anomalies (e.g., setpoint mismatches, efficiency issues)
* Intelligently routes discoveries to relevant stakeholders based on their role and interests
* Generates customized notifications adapting technical depth and presentation format per user
* Creates visual dashboards for discovered patterns
* Provides natural language chat interface for building queries

## Success Criteria Met

* Real-time data ingestion from physical HVAC equipment
* Autonomous pattern discovery without manual configuration
* Persona-based notification system with contextual insights
* 48-hour operational data retention with automatic cleanup

# 2. Architecture Overview

## High-Level System Design

The platform consists of five main layers working together:

**Physical Layer**

* 49 Fan Coil Units (FCUs) with LonWorks ((Local Operating Network)) protocol sensors
* Measurements: temperature, setpoints, heating/cooling output, fan speed. In PoC only one FCU (201) processed.

**Data Ingestion Layer**

* HiveMQ Cloud MQTT broker for secure data transmission
* Automated background worker collecting data every 5 minutes

**Storage Layer**

* Vercel Postgres database with 48-hour data retention
* Automatic daily cleanup preventing database bloat
* Stores telemetry data, discoveries, notifications, and activity logs

**AI Analysis Layer**

* Anthropic Claude AI with 19 (69 generated) specialized analysis tools
* Automated scheduler running analyses at configurable intervals
* Discovers patterns, correlations, anomalies, and efficiency issues
* Intelligent persona matching: Routes discoveries to relevant stakeholders based on role
* Adaptive notification formatting: Customizes technical depth, language, and presentation per recipient

**Presentation Layer**

* Web-based command center with real-time dashboards
* Live MQTT monitor with streaming charts
* AI chat interface for natural language queries
* Published dashboard sharing for stakeholder communication

## Technology Stack

**Core Technologies**

* Next.js 15 (Frontend framework)
* TypeScript 5.6 (Programming language)
* Anthropic Claude Sonnet 4.5 (AI/LLM)
* Vercel Platform (Hosting & database)
* HiveMQ Cloud (MQTT broker)
* Inngest (Background job orchestration)

**Key Features**

* React Server Components for optimal performance
* Server-Sent Events (SSE) for real-time data streaming
* Drizzle ORM for type-safe database operations
* Prompt caching achieving 90% token cost reduction

# 3. Workflow Description

## End-to-End Process

**Phase 1: Data Collection (Every 5 Minutes)**

1. FCU hardware publishes sensor readings to MQTT broker
2. Background worker connects and retrieves data from all 49 units
3. System extracts and normalizes FCU-201 data (focus unit for deep analysis)
4. Data stored in database as individual telemetry streams
5. Result: 20-30 data points collected per cycle

**Phase 2: AI Analysis (Configurable Interval)**

1. Auto-scheduler checks system configuration and timing
2. If analysis needed, AI engine loads historical context and persona profiles
3. System queries recent telemetry data from monitored streams
4. AI applies up to 19 analysis tools to discover patterns
5. Generates discoveries with evidence, reasoning, and recommendations
6. **Intelligent persona matching**: Evaluates which stakeholders should be notified based on:
   * Discovery category (energy, comfort, maintenance, safety)
   * Severity level (low, medium, high, critical)
   * Persona interests (configured preferences for types of issues)
   * Role responsibilities (engineer vs. manager vs. executive)

**Phase 3: Notification Generation (Background Processing)**

1. For each matched persona, AI generates a **customized notification** with:
   * **Adaptive language**: Technical detail for engineers, business impact for managers
   * **Role-specific focus**: Root cause for engineers, budget implications for executives
   * **Personalized tone**: Formal for executives, collaborative for team members
   * **Action-oriented recommendations**: Tailored to persona's decision-making authority
2. Creates optional visual dashboards illustrating the pattern (if beneficial for persona) and send notification by email
3. Saves notifications to database for UI display
4. Updates discovery status as "notified"
5. Process includes automatic retries for reliability

**Phase 4: User Interaction**

1. Stakeholders access web interface to view discoveries
2. Real-time MQTT monitor displays live system status
3. Chat interface allows natural language queries
4. Published dashboards can be shared with team members

## Typical Timeline

* **00:00** - MQTT data arrives, worker processes within 5 seconds
* **01:00** - Scheduler initiates analysis if criteria met
* **01:45** - AI generates discovery (e.g., "Setpoint Paradox detected")
* **01:52** - Notifications created for relevant stakeholders and send them by email
* **User access** - Discoveries visible immediately in web interface

# 4. Implementation Details

## Data Ingestion

Approach:

Serverless background worker on 5-minute schedule

Process:

* Connects to MQTT broker over secure TLS connection
* Subscribes to HVAC topic receiving all 49 FCU updates
* Parses LonWorks protocol messages extracting numeric values
* Normalizes field names for consistency
* Stores each metric as separate database stream
* Automatically disconnects after 50-second timeout

Configuration:

* Host: HiveMQ Cloud (EU region)
* Port: 8883 (MQTT over TLS)
* Topic: Building HVAC FCU Level 2
* Target: FCU-201 (expandable to all units)

## Discovery Engine

Approach:

Anthropic Claude AI with specialized analysis toolset and persona-aware notification system

19 (could be up to 69) Analysis Tools Include:

* Anomaly detection using statistical z-score
* Correlation analysis between data streams
* Pattern recognition for recurring behaviors
* Statistical calculations (mean, variance, trends)
* Time window comparisons
* Peak and trough identification
* Seasonality analysis
* Predictive forecasting
* Change point detection
* Volatility measurement
* And other additional analytical tools

Persona-Aware Notification System:

The platform goes beyond generic alerts by implementing intelligent stakeholder routing:

**Persona Profiling**: Each stakeholder has a configured profile with:

* Role (Building Engineer, Facility Manager, Operations Manager, etc.)
* Interest categories (energy efficiency, comfort, maintenance, safety)
* Communication preferences (technical detail level, update frequency)
* Decision-making authority (actionable recommendations scope)

**Discovery-to-Persona Matching**: AI evaluates each discovery and determines:

* Which personas should be notified (based on interest alignment)
* Priority level per persona (same discovery may be critical for engineer, informational for manager)
* Whether to include dashboard visualization (valuable for managers, optional for engineers)

**Adaptive Notification Generation**: For each matched persona, AI creates a customized message:

* Building Engineer receives: Technical root cause analysis, system parameters, troubleshooting steps
* Facility Manager receives: Business impact summary, user comfort implications, decision options
* Operations Manager receives: Performance metrics, cost implications, strategic recommendations

**Key Features**:

* Temperature setting: 0.8 (encourages creative discovery)
* Prompt caching: 90% reduction in API costs
* Context includes: persona profiles, recent discoveries, monitored streams, notification history

## Background Job Orchestration

**Approach**: Inngest event-driven architecture

**Benefits**:

* No serverless timeout limitations (analyses can run indefinitely)
* Automatic retries with exponential backoff
* Step-by-step execution tracking
* Individual step retry without restarting entire job
* Dashboard observability for monitoring

**Workers Implemented**:

* FCU data ingestion (every 5 minutes)
* Analysis auto-scheduler (every 5 minutes)
* Notification processor (event-driven)
* Telemetry cleanup (daily at 3 AM UTC)

## Real-Time Data Streaming

**Approach**: Server-Sent Events (SSE) for live dashboard updates

**Implementation**:

* Continuous connection from browser to server
* MQTT messages forwarded to connected clients in real-time
* Charts update automatically with new data points
* Keep-alive mechanism prevents connection drops
* Displays last 100 data points for performance

## Database Design

Key Tables:

* TelemetryTick: Sensor readings with timestamp and value
* TelemetryAnomaly: Detected anomaly records (deprecated)
* JeevesState: System configuration and monitored streams
* JeevesDiscovery: AI-generated findings with evidence
* JeevesNotification: Persona-specific customized alerts
* Persona: Stakeholder profiles with roles and communication preferences
* PersonaMemory: Historical interaction patterns for learning
* JeevesActivityLog: Execution history for debugging

**Optimization**:

* Indexed queries for fast retrieval (<100ms response time)
* 48-hour automatic cleanup preventing unlimited growth
* Type-safe schema with Drizzle ORM

## Deployment Configuration

Platform: Vercel (serverless architecture)

Key Settings:

* Production deployment with automatic scaling
* Extended timeout (300s) for analysis endpoint
* PostgreSQL database with connection pooling
* Environment variables for secure credential management

External Services:

* HiveMQ Cloud: MQTT message broker
* Anthropic API: Claude AI inference
* Inngest Cloud: Background job orchestration
* Optional: Redis for deduplication, Resend for email delivery

# 5. Results

**Performance Metrics**

**Data Ingestion:**

* MQTT connection time: ~2 seconds
* Message processing: ~500ms per message
* Database insertion: ~50ms per stream
* Total cycle time: ~5 seconds for 20-30 streams
* Success rate: 99.8%

**AI Analysis:**

* Discovery generation: 30-60 seconds
* Token usage (with caching): 800-1,000 tokens
* Token usage (without caching): 8,000-10,000 tokens
* Token cost savings: 90%
* Tool calls per analysis: 3-8 (average 5)
* Success rate: 100%

**System Availability:**

* Uptime: 99.9% (Vercel SLA)
* Cold start time: <3 seconds
* Page load time: <1 second
* Database query time: <100ms

**Storage Efficiency:**

* Data points per FCU: ~16,000 rows (48h window)
* Database size: ~1-2 MB per FCU
* Cleanup duration: ~2 seconds daily
* Free tier capacity: 256 MB (200+ MB safety margin)

**Example Discovery**

**Title:** "Half-Degree Setpoint Paradox" **Discovered:** October 13, 2025 at 18:45 UTC **Severity:** High **Confidence:** 95%

**Finding:** User interface shows setpoint of 22.5°C, but FCU consistently targets 22.0°C, a persistent 0.5°C offset observed across 193 data points over 6 hours.

**Evidence:**

* User setpoint: 22.5°C (constant)
* Effective setpoint: 22.0°C (constant)
* Current temperature: 23.2°C (varying)
* Pattern consistency: 100%
* Correlation: 0.00 (setpoints don't influence each other)

**Hypothesis:**

* Energy savings mode with programmed offset
* Deadband zone configuration
* Control algorithm requiring calibration
* Special operating mode active (eco/night)

**Persona-Specific Notifications Generated:**

**Building Engineer (Alex Chen) - Technical Focus:**

"**Half-Degree Setpoint Offset Detected in FCU-201**

I've discovered a persistent 0.5°C offset between user setpoint (22.5°C) and effective setpoint (22.0°C) over 6 hours with 100% consistency.

**Technical Analysis:**

* Correlation: 0.00 (setpoints independent)
* Room temp: 23.2°C (1.2°C above target)
* No active heating/cooling response

**Recommended Investigation:**

1. Check FCU-201 control parameters for programmed offset
2. Verify deadband configuration (may explain lack of cooling at 23.2°C)
3. Review operating mode (eco/night mode may enforce offset)
4. Examine control logic for calibration requirements

This appears to be either intentional energy-saving configuration or a control loop issue requiring calibration."

**Facility Manager (Maria Rodriguez) - Impact Focus:**

"**Temperature Control Issue in FCU-201 Affecting User Comfort**

FCU-201 is not reaching the user's requested temperature setpoint of 22.5°C, instead consistently targeting 22.0°C. The room is currently 23.2°C, above both setpoints.

**Impact Summary:**

* User comfort: Room 0.7°C warmer than requested
* Duration: 6+ hours without correction
* Energy efficiency: System inactive despite temperature deviation

**Decision Required:** Should we maintain current settings (possible energy-saving mode) or restore full user control? Engineering team can investigate root cause and present options."

**Result:**

* Issue identified within 1 hour of data collection, before user complaints
* Each stakeholder receives notification tailored to their role and decision-making needs
* Engineer gets technical details for troubleshooting
* Manager gets business context for decision-making

# 6. Challenges and Limitations

Technical Challenges Resolved

**API Rate Limiting**

* Issue: Exceeded Anthropic token limits during heavy usage
* Solution: Implemented automatic retry with exponential backoff plus prompt caching
* Result: Zero rate limit failures after implementation
* Additional FCU Integration will need request loop per FCU

**Serverless Timeouts**

* Issue: Combined discovery and notification processing exceeded 300s limit
* Solution: Split into two phases using background event-driven worker
* Result: 100% completion rate

**Current Limitations**

**Single FCU Focus**

* Current: Only FCU-201 data ingested for deep analysis
* Reason: Validate approach before scaling to all 49 units

**48-Hour Data Retention**

* Current: Automatic cleanup deletes data older than 48 hours
* Reason: Prevent database bloat within free tier limits
* Future: Implement data aggregation (hourly averages) for longer retention

**Manual Configuration**

* Current: Monitored streams must be configured manually
* Reason: No auto-discovery of interesting streams yet
* Future: AI-powered stream selection based on variability

**No broad Email Delivery**

* Current: Notifications send only for one email
* Reason: Domain verification need in the Resend service
* Future: Integrate email service for proactive delivery

**AI Considerations**

**Non-Deterministic Discovery**

* Observation: Same data may produce different discoveries
* Reason: Intentional creativity for pattern exploration
* Mitigation: User feedback loop to learn preferences

**Tool Call Limitations**

* Reason: Balance between thoroughness and token usage
* Future: Iterative analysis for complex patterns

**Prompt Dependency**

* Observation: Discovery quality sensitive to prompt wording
* Approach: Iterative refinement based on user feedback, prompt engineering via UI
* Result: Continuous improvement over time

# 7. Recommendations

**1. Security Hardening**

* Move MQTT credentials from code to environment variables
* Implement credential rotation policy
* Impact: Critical security improvement

**2. Email Integration**

* Verify domain and Integrate email service
* Create HTML email templates per persona
* Add unsubscribe functionality
* Impact: Proactive stakeholder engagement

**3. Multi FCU and Multi-Building Support**

* Add building identifier to database schema
* Implement building selector in UI navigation
* Support multiple MQTT broker connections
* Impact: Portfolio scalability

**4. Long-Term Data Retention**

* Create aggregates table (hourly/daily averages)
* Aggregate before cleanup (retain summaries forever)
* Enable historical trend analysis
* Impact: Strategic planning capabilities

**5. Comprehensive Error Handling**

* MQTT auto-reconnect with backoff
* Database transaction rollback and retry
* LLM graceful degradation
* Impact: System reliability improvement

**6. Expand FCU Coverage**

* Add 5-10 additional units (mix normal + faulty)
* Enable building-wide pattern detection
* Impact: Richer comparative insights

**7. Persona Feedback Loop**

* Add processing of "Helpful/Not Helpful" buttons on notifications
* Track which notification styles resonate with each persona
* Adjust AI communication style based on feedback
* Learn persona preferences: technical depth, dashboard preference, action item style
* Impact: Continuous personalization improvement, higher notification engagement

**8. Custom Alert Thresholds**

* UI for per-persona alert configuration
* Check thresholds before notification
* Impact: Reduced notification fatigue

**9. Custom Anomaly Model**

* Train HVAC-specific machine learning model
* Reduce LLM reliance for basic anomalies
* Reserve AI for complex interpretation
* Impact: Cost reduction, faster detection

**10. Predictive Maintenance**

* Analyze historical failure patterns
* Predict equipment failures 24-48 hours ahead
* Integrate with work order systems
* Impact: Proactive maintenance scheduling

**11. Energy Efficiency Scoring**

* Calculate building energy efficiency score
* Compare against industry benchmarks
* Recommend optimization actions
* Impact: Cost savings identification

**12. Decouple Discovery and Notification Generation**

* Separates discovery analysis from per-persona notification creation
* Enables on-demand notification generation for existing discoveries
* Allows re-generating notifications without re-analyzing data
* Reduces API costs and enables flexible iteration

**13. Dashboard Quality Validation Agent**

* AI agent reviews generated dashboards before publishing
* Validates data visualization best practices
* Checks clarity, chart type appropriateness, and accessibility
* Suggests improvements for unclear visualizations