MarTech Systems Architecture Review

Scalability, Cost Efficiency, and Data Foundation Strategy

Context

This memo evaluates the current marketing technology stack through the lens of long-term scalability, cost discipline, and data reliability.

The objective was not simply to audit tools, but to clarify system roles and identify structural gaps that would constrain reporting, attribution, and operational flexibility as volume increases.

The central question:

> How should CRM, CDP, automation, and data infrastructure be structured to support scale without introducing technical debt or vendor lock-in?

Executive Summary

The current stack is functional for early-stage execution, but lacks key structural components required for durable growth:

- No Customer Data Platform (CDP) to unify product and marketing events

- No centralized data warehouse for historical storage and cross-system modeling

- Functional overlap across outbound and automation tooling

- Measurement exposed to evolving compliance and deliverability constraints

Without separating execution from storage and activation, performance metrics degrade as complexity increases.

The solution is architectural clarity — assigning each system a defined responsibility and preventing the CRM from becoming a bottleneck.

Architectural Principle

Each system should serve a distinct role:

- **Execution (CRM and Campaign Tools)**

- **Storage and Modeling (Data Warehouse)**

- **Collection and Activation (CDP)**

Attempting to centralize all business logic inside a CRM introduces scale ceilings over time, particularly around property limits, API thresholds, and long-term historical storage.

A composable approach preserves flexibility while maintaining consistent, decision-grade reporting.

System Role Definitions

CRM — HubSpot

HubSpot remains appropriate for the current stage.

It serves as the operational system of record for:

- Contacts

- Companies

- Deals

- Campaign execution and workflow automation

However, it is not designed for:

- Unlimited telemetry

- Deep historical retention

- Cross-system modeling

- Machine learning feature engineering

As scale increases, complex logic and reporting should move upstream into a warehouse environment.

CDP — Composable Approach

A composable CDP (e.g., Segment or RudderStack) separates event collection from storage.

Core responsibilities:

- Capture product, web, and server-side events

- Resolve identity across channels

- Sync enriched data to CRM, advertising platforms, and product systems

Where possible, vendor consolidation should be prioritized to avoid unnecessary overlap between event collection and reverse ETL providers.

The CDP extends visibility beyond what the CRM can natively observe.

Data Warehouse — Missing Foundation

A data warehouse (Snowflake, BigQuery, or Redshift) is the structural gap today.

It should own:

- Raw and modeled tables

- Cross-system joins

- Long-term historical storage

- Analytics and forecasting

- Feature generation for scoring and activation

A warehouse is not file storage.
 It is a structured analytical database optimized for modeling and scale.

Without a warehouse-first strategy, reporting becomes tool dependent and vulnerable to vendor limitations.

Stack-Specific Observations

Automation and Orchestration

n8n introduces operational instability unless configured in queue mode with workers, increasing DevOps overhead.

For business critical automations, reliability outweighs marginal cost savings.

Recommendation:
 Standardize core automations in Zapier.
 Reserve n8n only for controlled, developer managed workflows or retire it.

Outreach Stack

Apollo and Instantly overlap in sequencing and outbound capabilities.

Maintaining both likely duplicates spend without proportionate gain.

Additionally, Google and Yahoo sender requirements (SPF, DKIM, DMARC, complaint thresholds) reduce long-term effectiveness of mailbox warmup strategies.

Outbound tooling should be consolidated with compliance resilience in mind.

Analytics and Measurement Gaps

Current gaps include:

- Product analytics

- Consent management for EU and UK compliance

- Offline conversion stitching for LinkedIn Ads

Offline conversion syncing from CRM outcomes back into LinkedIn is critical for accurate ROI measurement in higher-ticket sales cycles.

Without proper attribution hygiene, paid performance metrics deteriorate even when underlying revenue improves.

Risks If Unaddressed

- Tool sprawl and duplicated spend

- HubSpot scale ceilings (properties, objects, API volume)

- Deliverability degradation from tightening compliance rules

- Weak identity resolution causing duplicate records

- Reporting fragmentation across platforms

- Increased future re-platform risk

These risks compound over time.

Phased Recommendations

Phase 1 — Immediate Stabilization

- Move business critical automations to Zapier

- Consolidate outbound tooling (choose Apollo or Instantly)

- Implement a Consent Management Platform

- Configure LinkedIn Offline Conversions

Phase 2 — Structural Layer

- Deploy Segment or RudderStack for event collection

- Add product analytics (Mixpanel or PostHog)

- Introduce reverse ETL if activation needs exceed CDP native capabilities

---

Phase 3 — Foundational Infrastructure

- Stand up a data warehouse (Snowflake, BigQuery, or Redshift)

- Shift complex logic and reporting out of HubSpot

- Formalize data governance and AI data handling policies

- Evaluate Fivetran vs Airbyte for sync efficiency

Conclusion

A scalable revenue architecture requires:

- HubSpot for operational execution

- A composable CDP for event collection and activation

- Zapier for reliable automation

- A centralized data warehouse as the source of truth

- Consolidated outbound tooling

- Early investment in consent, attribution, and identity hygiene

This structure reduces overlapping spend, preserves architectural flexibility, and protects metric integrity as scale increases.

Most importantly, it prevents expensive re-platforming later by separating execution from storage and modeling early.