AI-Enabled Data Pipeline + Analytics Architect (Contract)

Objective: Build a scalable, production-ready data infrastructure that allows our team to conversationally prompt AI tools (Perplexity Enterprise Max, GPT-5 Pro) to query our unified warehouse and generate real-time visual reports.

About Us

Manifest Commerce is a tech-enabled 3PL (third-party logistics) specializing in high-volume fulfillment, ecommerce integrations, and analytics-driven operations. We already leverage AI tools heavily, but we're now ready to build the underlying data infrastructure needed to automate workflows end-to-end.

We're looking for an experienced contractor who understands **data engineering + AI connectors + enterprise automation** to help architect and implement this system.

Scope of Work

Phase 1 — Data Engineering & Warehousing

Goal: Get all operational data flowing cleanly and consistently into our BigQuery warehouse.

Responsibilities

- Review our existing data ecosystem (ShipHero → Make.com → Parabola → BigQuery).
- Design a scalable data model aligned to our 3PL operations (orders, shipments, SLAs, POs, etc.).
- Build or improve pipelines to ensure **clean, deduped, normalized, and query-ready** data.
- Implement monitoring / observability for pipeline failures.
- Set up incremental refresh logic and change-data-capture where appropriate.

Primary data sources include:

- **ShipHero WMS** (API + webhooks + internal events)
- **Google Sheets** (including form-driven Sheets)
- **HubSpot** CRM
- **Zendesk** (tickets, SLA events, tags)
- **Lateshipment.com** (carrier performance, claims info, reimbursement values)
- **Deputy** (workforce management/labor data)
- **QuickBooks Online** (and eventually ERP such as Microsoft Dynamics)
- Other structured/external sources as needed

Phase 2 — AI Connector Layer (Perplexity & GPT-5 Pro)

Goal: Enable conversational querying of our warehouse and automated insights generation.

Responsibilities

- Integrate BigQuery with **Perplexity Enterprise Max** and/or **GPT-5 Pro**.
- Architect secure access patterns so LLMs can query the warehouse through:
    - Direct API integrations
    - A middleware layer (Make.com, n8n, or custom)
    - Embeddings-powered semantic retrieval where needed
- Implement guardrails, role-based access, and cost-controlled query strategies.
- Produce example conversational workflows such as:
    - “Show me week-over-week shipment volume by client.”
    - “Graph time-in-transit trends for DHL over the past 90 days.”
    - “Show me week-over-week on-time fulfillment SLA achievement rates broken out by D2C vs. B2B orders”

Phase 3 — Reporting, Dashboards, & Visual Output

Goal: Create executive-ready visualizations AI can produce on demand.

Responsibilities

- Set up AI-driven generation of charts, tables, and summaries (e.g., Looker Studio, custom visualization endpoints, or direct image rendering).
- Build reusable templates for:
    - WoW KPI graphs
    - Client-level shipment analytics
    - SLA breach reporting
    - Labor + throughput forecasting
- Ensure charts/graphs generated by AI can be exported, shared, or embedded.

Required Skills

You should have strong experience in:

Data Engineering & Warehousing

- BigQuery (must)
- Data modeling for operational/transactional systems
- ETL/ELT pipeline building (Make.com, n8n, Parabola, or custom)

- API integrations + webhook listener architectures

AI Integration

- Connecting LLMs to databases securely
- Perplexity Enterprise or OpenAI GPT-4/5 API experience
- Middleware automation tools (Make, n8n, Zapier)
- Vector databases + embeddings (nice to have)

Tech Stack Familiarity (Strongly Preferred)

- ShipHero WMS API (or similar WMS/OMS APIs)
- HubSpot API
- Zendesk API
- Google Sheets API / Apps Script
- QuickBooks Online API
- Visualization tools (Looker Studio, Retool, Superset, Mode, etc.)

What Success Looks Like

Within 6–10 weeks, we want:

1. **A clean, reliable, production-ready BigQuery warehouse** for all operational data.

2. **AI-driven conversational analytics** (Perplexity + GPT) working against real data.

3. **Executive-friendly visual outputs** generated automatically from AI prompts.

4. Reusable pipelines and documentation for ongoing expansion (ERP integration in 2026).
