



---

# The PrimaForja Synthetic Distributor Dataset

A production-scale, fully configurable, deeply realistic data  
foundation for analytics, BI, and machine-learning teams

**A PRIMAFORJA WHITEPAPER**

---

PrimaForja · [info@primaforja.com](mailto:info@primaforja.com)

## Abstract

---

Analytics and data teams need data that looks and behaves like a real operating business — at scale, across every department, and internally consistent from the general ledger down to the individual shipment. Real production data cannot be used for this: it carries privacy obligations, security exposure, contractual restrictions, and access friction, and it is almost never available to the vendors, prospects, students, and new hires who need it most.

The **PrimaForja Synthetic Distributor Dataset** solves this. It is a complete, self-consistent, entirely synthetic dataset modeling a multi-billion-dollar North American wholesale distributor across two full calendar years — **62 interrelated tables and 809 million rows** spanning orders, fulfillment, inventory, logistics, procurement, receivables and payables, pricing, promotions, workforce, fleet, safety, and a full executive-grade financial layer including a line-level margin waterfall.

Two things set it apart from every other sample dataset on the market. First, **nothing is fixed**. The entire business is generated from more than **3,300 individually adjustable values** — every count, rate, range, cost, and behavioral rule can be dialed up or down to model the business *you* want. Second, the realism is **deep and interlocking**: a distribution center's size drives its headcount, which drives its payroll and benefits; a facility's region drives its wage bands, utility seasonality, and insurance; whether it is owned or leased changes its entire cost structure — and all of it reconciles back to revenue. Every record is fabricated; no real customer, employee, transaction, or personally identifiable information appears anywhere in it.

---

## 1. Why synthetic data — and why this dataset

---

Four recurring needs drive demand for realistic, shareable data:

- **Demonstration and evaluation.** BI vendors, consultancies, and platform teams need a rich, believable dataset to show what a dashboard, model, or pipeline can do — without exposing a customer's confidential numbers.
- **Development and testing at scale.** ETL/ELT pipelines, warehouse schemas, and query performance behave very differently against a hundred rows than against hundreds of millions.
- **Machine-learning experimentation.** Forecasting, churn, pricing, and anomaly-detection models need labeled, patterned, temporally coherent data.
- **Training and enablement.** New analysts, students, and course authors need a safe, shareable sandbox that mirrors real enterprise complexity.

Most public sample datasets fail these needs the same ways: too small, single-department, no cross-table integrity, obviously synthetic patterns, no executive/financial depth, fixed and unconfigurable, and murky provenance. The PrimaForja dataset was built specifically to close every one of those gaps.

---

## 2. Everything is configurable — nothing is set in stone

**Read this before the numbers that follow: every figure in this paper is a default, not a constraint.** The dataset is produced by a configuration-driven engine with more than **3,300 individually tunable values**, organized into **440+ named parameter groups** across roughly **30 business domains**. Each value is documented with its units and valid range, and the engine validates the whole configuration for internal consistency before it generates a single row.

That means the dataset flexes to your scenario:

- **Company size.** The default models a national distributor doing \$4.38 billion a year — \$12.0 million every day across ~50,000 orders. Turn the dials and it becomes a \$400M regional player or a \$40B national giant — revenue targets, order volume, and entity counts all move together.
- **Footprint and geography.** Change the number of distribution centers, customers, vendors, manufacturers, carriers, and sales reps. Reshape the geography across the five U.S. regions, or point it at a different geography entirely.
- **Time window.** Generate two years, one quarter, or five years; seasonality, aging, and year-over-year logic adapt automatically.
- **Business mix and behavior.** Adjust customer segment shares, order sizes and frequency, product mix, discount and promotion policy, fill rates, on-time-delivery targets, return rates, payment behavior, forecast accuracy, and much more.
- **Cost structure.** Wage bands, fuel prices, freight rates, facility economics, benefits burden, and commission plans are all parameters.

The point: this is not one frozen dataset — it is a *dataset generator* delivered as data. Two teams with different needs get two different, equally coherent businesses from the same engine.

For readability, the rest of this paper describes the **default** configuration. Assume every number can change.

## 3. The modeled business (default configuration)

By default, the dataset portrays a single fictional enterprise — a national industrial and office-supply wholesale distributor operating across the United States:

| Dimension            | Default profile (all adjustable)                        |
|----------------------|---------------------------------------------------------|
| Annual revenue       | \$4.38B (\$12.0M per day), seasonally distributed       |
| Distribution network | 27 distribution centers across five U.S. regions        |
| Customers            | 6,763 business accounts across five commercial segments |
| Product catalog      | ~37,000 SKUs across a full category hierarchy           |
| Supply base          | 391 vendors and 593 manufacturers                       |

| Dimension          | Default profile (all adjustable)                                        |
|--------------------|-------------------------------------------------------------------------|
| Sales organization | 113 field reps with quotas and territories                              |
| Carriers           | 23 carriers across parcel, LTL, courier, and company fleet              |
| Time window        | Two complete calendar years (prior-year comparison built in)            |
| Order volume       | 39.0 million orders (49,990 non-cancelled per day, \$240 average order) |

Crucially, the business is **coherent across every department**. Revenue in the finance tables reconciles to orders in the sales tables. Labor hours in the warehouse tables tie to throughput in the fulfillment tables. Vendor performance in procurement flows through to inventory availability and, ultimately, to customer fill rates. A question asked in one domain can be answered — and cross-checked — in another.

## 4. The realism engine

Volume and configurability mean little without believability. What makes the PrimaForja dataset stand apart is how much genuine business logic is wired into it, and how the pieces interlock the way they do in a real company. A few of the systems built in:

### Facility economics — owned vs. leased, and driven by size and region

Each distribution center has a size, a capacity scaled to the inventory it holds, and a utilization profile. Its **occupancy cost structure depends on whether it is owned or leased**: leased sites carry rent and CAM; owned sites carry depreciation, property tax, and mortgage interest. On top of that sit utilities that follow **region-specific seasonal curves** (heating-driven in the north, cooling-driven in the south), plus insurance, security, and maintenance that vary by region and facility. The result supports fully-loaded cost-per-order analysis by facility — and it all traces to a capital-expenditure schedule that feeds the depreciation.

### Workforce and payroll — headcount derived from the work, not assigned

Headcount is not a number someone typed in. **A distribution center's size and actual throughput drive its staffing** through labor standards: more volume means more receiving, picking, packing, and shipping FTEs. Warehouse roles are modeled distinctly from office and corporate roles, each with **region-specific wage bands** (defaulting to roughly \$16–\$34/hour depending on region and role), overtime that rises in peak months, and a **benefits/burden load** on top of base pay. Warehouse turnover runs realistically higher than corporate; seasonal flex and temp labor appear when demand spikes. A separate corporate roster (executives, finance, IT, HR, marketing, procurement, customer service, sales ops) ties the SG&A salary line to actual people, so revenue-per-employee and headcount-vs-workload analyses hold up.

### Sales and compensation — pay tied to real performance

Sales reps carry quotas and territories, and their compensation is modeled end to end: a **regional base salary**, attainment against quota, **tiered commissions paid on pocket margin** (not just revenue), and an accelerator above 100% attainment. Because commissions pay on pocket margin, the compensation tables reconcile to the same profitability metric the finance tables use.

## Transportation, fuel, and fleet

The engine **tracks fuel prices by state and month**, applies fuel surcharges to freight, and prices carriers by tier and mode (parcel, LTL, courier, company fleet). Cartonization, multi-parcel and partial shipments, and accessorial fees (liftgate, inside delivery, residential, hazmat) are all modeled. For company-fleet DCs, a truck register carries vehicle age that drives a **maintenance cost curve** — older trucks cost more per mile.

## Pricing, promotions, and rebates

List, catalog, and contract pricing coexist, with **segment- and category-specific discounts**, volume breaks, price floors and ceilings, and price elasticity that varies by item class. A promotion engine produces percent-off, fixed-off, BOGO, and clearance events with marketing-campaign attribution, rebate accruals, and measurable promotional lift.

## Working capital — both sides of the balance sheet

Accounts receivable models segment-specific **payment personalities** (early, on-time, late, delinquent), aging buckets, credit holds, and write-offs. Accounts payable mirrors it on the vendor side, with terms by vendor tier, early-pay discount capture, and deliberate payables stretching in peak-cash months — enabling days-payable-outstanding and a full cash-conversion-cycle analysis alongside inventory.

## Demand, inventory, and forecasting

Demand follows realistic monthly seasonality, weekday/weekend patterns, holiday spikes and lulls, and year-over-year growth, personalized by segment and region. Items carry industry-standard **ABC (value) and XYZ (variability) classifications derived from their own realized demand**, and pass through lifecycle stages from introduction to end-of-life. Forecast error scales the way it does in practice — tightest for high-volume, stable items and widest for slow-moving, erratic long-tail SKUs. Inventory carries health profiles (healthy, out-of-stock, below-reorder, overstock), aging, and velocity-class behavior per DC.

## Safety and internal operations

Recordable safety incidents are calibrated to standard industry rates per 200,000 labor hours by department — computable precisely *because* real labor hours exist in the data. Warehouse equipment (forklifts, conveyors, dock doors) generates planned and unplanned maintenance with downtime. And the company consumes its own catalog: office and breakroom supplies drawn at cost, driven by headcount.

## It all interlocks

These systems are not independently generated and stitched together. Vendor reliability affects inventory availability, which affects fill rates and backorders, which affect customer satisfaction and returns, which affect margin. DC size drives headcount drives payroll drives overhead drives EBITDA. Region drives wages *and* utilities *and* insurance *and* delivery performance. That causal wiring is what makes the data hold up under an expert's scrutiny — and what makes it genuinely useful for analysis rather than merely large.

## 5. The executive financial layer — the C-level differentiator

Most sample datasets stop at operational transactions. The PrimaForja dataset extends into the financial and executive analytics a CFO, COO, CRO, and CEO actually review — the layer that turns a transactional dataset into a boardroom-grade one, and the one that makes it uniquely suited to **C-level dashboards and visualizations**.

At its heart is a **line-level pocket-margin waterfall** (~121 million rows) that walks each sale from list price, through on-invoice, promotional, and contract discounts, to invoice price; then through early-pay discounts, rebate accruals, freight recovery, and returns credits to a *pocket price*; then through cost of goods to *pocket margin*; and finally through allocated cost-to-serve — warehouse labor, freight, receivables carrying cost, returns processing — to true *contribution*. This makes it possible to analyze profitability the way leading distributors actually do: to surface customers, reps, segments, and categories that show healthy gross margin but eroded pocket margin once the real cost of serving them is counted.

Around it sits a complete executive layer: facility and occupancy economics, an accounts-payable ledger and working-capital cycle, monthly corporate overhead and a budget-vs-actual plan that supports a full revenue → gross → pocket → contribution → **EBITDA bridge**, workforce headcount and corporate payroll, sales-rep compensation tied to attained margin, a delivery-fleet register with age-driven maintenance, internal consumption, equipment maintenance, safety incidents, and a capital-expenditure schedule feeding depreciation.

This is what lets the dataset drive genuine executive dashboards — CEO, CFO, COO, and CRO views — not just operational reports. Very few datasets on the market can populate a C-suite dashboard at all; this one is built for it.

## 6. How it compares to other datasets

|                             | Typical sample datasets          | PrimaForja Synthetic Distributor Dataset                     |
|-----------------------------|----------------------------------|--------------------------------------------------------------|
| Scale                       | Thousands to a few million rows  | ~800 million rows, 62 tables                                 |
| Breadth                     | One or two departments           | End-to-end: sales, ops, logistics, finance, workforce, fleet |
| Cross-table integrity       | Loose or absent                  | Enforced everywhere; every domain reconciles                 |
| Executive / financial depth | Rare; usually none               | Full C-level layer + line-level margin waterfall             |
| Realism                     | Random values, visible artifacts | Interlocking business logic (size→staff→payroll→EBITDA)      |
| Configurability             | Fixed                            | 3,300+ tunable values; model any size or mix                 |

|                               | Typical sample datasets | PrimaForja Synthetic Distributor Dataset                                                                          |
|-------------------------------|-------------------------|-------------------------------------------------------------------------------------------------------------------|
| <b>Reproducibility</b>        | Varies                  | Deterministic and versioned — identical for every recipient                                                       |
| <b>Provenance / licensing</b> | Often unclear           | Fully synthetic, no PII; licensed for your use — showcase demos & dashboards publicly, no raw-data redistribution |

The short version: other datasets give you *rows*. This one gives you a *business* — coherent, configurable, and complete all the way up to the boardroom.

## 7. Scale and delivery

Delivered at production scale by default (~800 million rows across 62 tables, two years, tens of gigabytes of Parquet), and available in the formats analytics teams already use:

- **Parquet** — columnar, compressed, typed; ideal for data lakes and distributed compute (Spark, DuckDB, Snowflake, BigQuery, Databricks).
- **CSV** — universally importable.
- **PostgreSQL schema** — a complete, documented relational schema with keys, indexes, constraints, and table-level documentation.
- **Materialized analytics views** — a companion library of pre-built views that pre-aggregate the most common executive and operational metrics, so dashboards run fast out of the box.

Reduced-scale editions preserve the same structure and relationships for quick iteration. Every table is documented with a generated data dictionary, and the schema drives BI tools such as Tableau, Power BI, Looker, and Metabase directly.

## 8. Data quality, reproducibility, and privacy

**Quality.** Every generated edition is validated before release against an automated suite of integrity and realism checks — referential integrity, arithmetic and financial identities (the margin waterfall reconciles; receivables tie to revenue; payroll and overhead reconcile), distribution properties, and business-rule constraints. An edition is not released unless it passes.

**Reproducibility.** A given edition is fixed and stable, so every recipient works from identical data. Results are comparable across teams, demonstrations are repeatable, test suites are deterministic, and a benchmark run today can be reproduced months from now. Editions are versioned so consumers can pin to a known release.

**Privacy.** Because the dataset is entirely synthetic, it carries none of the risk of production data. Every name, address, email, account, and transaction is fabricated; there is no PII, no real customer, employee, or financial

record, and it is not derived from, sampled from, or anonymized from any real company. With no real personal or financial data, the usual data-protection obligations simply do not apply.

**Licensing.** The dataset is licensed to the recipient for their own use and is **not for public redistribution** — the raw data files and schema you receive are not to be republished, resold, or handed to the public. You are, however, free to share what you *build* with it: demos, dashboards, visualizations, screenshots, analyses, reports, and presentations created from the data can be shown publicly, embedded in courseware, and used in publications. In short — the outputs are yours to showcase; the underlying dataset stays with you.

## 9. Representative use cases

- **BI and executive dashboards** — build and demonstrate CEO/CFO/COO/CRO and operational dashboards against a complete, cross-functional dataset, including C-level views most sample data cannot support.
- **Data-pipeline and warehouse engineering** — develop and stress-test ETL/ELT pipelines, dimensional models, and dbt projects at genuine production volume.
- **Machine learning** — train and validate demand-forecasting, churn, price-elasticity, anomaly-detection, and margin-optimization models on labeled, temporally coherent data.
- **Database and query benchmarking** — evaluate performance, indexing, and partitioning under realistic scale and join complexity.
- **Sales enablement and evaluation** — give prospects a rich, believable environment without touching their confidential data.
- **Training and education** — a safe, enterprise-grade sandbox that mirrors real business complexity.

## Appendix A — the configuration surface

The engine exposes **more than 3,300 individually adjustable values** in **450+ named parameter groups** across **140+ documented sections**. Grouped into buyer-facing domains, the approximate distribution of tunable values is:

| Domain                                   | Tunable values (approx.) |
|------------------------------------------|--------------------------|
| Executive / C-level financial layer      | 440                      |
| Shipping, carriers & on-time delivery    | 410                      |
| Orders, demand & seasonality             | 410                      |
| Pricing, discounts, promotions & rebates | 300                      |
| Cross-domain rules & safety guards       | 215                      |
| Customers & segmentation                 | 215                      |

| Domain                                 | Tunable values (approx.) |
|----------------------------------------|--------------------------|
| Forecasting & demand signals           | 210                      |
| Inventory & warehousing                | 200                      |
| Item catalog, cost & lifecycle         | 195                      |
| Accounts receivable & payment behavior | 140                      |
| Geography, entities & scale            | 130                      |
| Transportation, fuel & labor cost      | 110                      |
| Returns & credits                      | 100                      |
| Vendors & procurement                  | 85                       |
| Revenue targets & distribution         | 40                       |
| Fulfillment & fill rates               | 10                       |

A few illustrative defaults (each adjustable): five customer segments from Enterprise to Wholesale; payment terms from NET-15 to NET-60 plus COD; company on-time delivery in the low-to-mid 90s percent; healthy-inventory share in the 75–90% band; return rates in the 3–7% band; regional wage bands spanning roughly \$16–\$34/hour; and forecast accuracy that scales by item value and variability class.

## Appendix B — the modifiable category catalog

The engine exposes **445 named, individually modifiable categories** (each a group of one or more tunable values). Every category below can be adjusted to reshape the dataset; the specific values, ranges, and rules *inside* each category are part of the generator and are not enumerated here. This catalog is provided so buyers can see the full breadth of what is configurable.

**Geography, entities & scale (15).** Start date, End date, Number of DCs, Number of manufacturers, Number of sales reps, Master account fraction, Region mapping, Region monthly variation, Geonames URL, Geonames cache dir, Geonames US path, Cross-domain rules, Annual cost inflation rate, Default variation range, Open-records final constraints.

**Revenue targets & distribution (12).** Revenue annual range, Revenue calibration, Revenue monthly variation, Revenue DC share tiers, Revenue DC noise, Revenue segment share, Revenue segment noise, Revenue category multipliers, Revenue category noise, Revenue final constraints, Order revenue rules, AR revenue ratio range.

**Customers & segmentation (23).** Number of customers, Customer segment weights, Customer activity tiers, Customer tier line-count multipliers, Customer region distribution, Customer region order multiplier, Customer behavior rules, Customer lifecycle rules, Customer status (active / churned / new), Customer churn monthly probability, Churn risk weights, Churn days (low / high), Customer acquisition monthly rate,

Customer acquisition ramp months, New-customer fraction, Dormant-deep fraction, Customer RFM weights, Preference RFM weights, Segment quantity params, Default segment multiplier, Customer final constraints.

**Orders, demand & seasonality (31).** Number of orders, Order mix rates, Order aging targets, Order line variability, Order status realism, Order type line-count multipliers, Segment order lines, Segment order frequency, Monthly seasonality weights, Weekday weight multiplier, Year-over-year growth rate, Holiday spikes, Holiday dips, Seasonal line-count multipliers, Segment seasonality, Region seasonality multipliers, Category seasonality, Category demand multipliers, SKU demand rules, SKU demand limits, SKU seasonality alignment, Demand shock events, Order-to-ship days range, Order-to-ship variability, Order-to-ship seasonality, and related rules.

**Item catalog, cost & lifecycle (23).** Item classification rules, Item lifecycle rules, Item complexity rules, Item hazmat rules, Item lead-time rules, Item duty rules, Item margin rules, Item inflation rules, Item cost record counts, Item cost staleness / spacing, Item cost version noise, ABC quantity multipliers, Lifecycle quantity multipliers, Category target margins, Margin noise, SKU lifecycle rules, SKU complexity noise, SKU final constraints, Item catalog paths.

**Pricing, discounts, promotions & rebates (61).** Discount ranges, Segment discount rules and noise, Category discount rules and noise, Customer pricing tiers, Region price multipliers, Order-type price adjustments, Volume discount rules, Price floor / ceiling multipliers, Price noise, Annual price inflation, Contract-price discount rules, Price-history versions / channels / jitter, Price elasticity rules and windows, Number of promotions, Promotion duration / penetration, Promotion discount types and magnitudes, Promotion timing / eligibility / assignment rules, Promotion seasonality, Rebate rules and eligibility, Marketing campaign channels / budgets / conversion / impressions, Promotional-lift windows, and related pricing constraints.

**Vendors & procurement (30).** Number of vendors, Vendor tier split, Vendor on-time rate by tier, Vendor defect rate by tier, Vendor fill rate by tier, Vendor lead-time ranges and variability, Vendor monthly variation, Vendor late delivery, Vendor assignment rules, Vendor item rules, Vendor scorecard weights and trend threshold, Max vendors per manufacturer, Purchase-order open / cancelled / partial-receipt rates, PO receipt timing, PO late-days and open-future-days ranges, and open-PO realism.

**Fulfillment, fill rates & backorders (32).** Full-fill rate, Short-ship rate and severity, Segment fill rates, Order-type fill rates, DC fill rates, Fill-rate company range and DC tiers, Fill-rate monthly variation, Substitution rules, Lost-sales rules, Backorder rate and max percent, Backorder reason codes, Backorder fill rate and timing, Backorder resolution methods (filled / lost / cancelled / open), Backorder aging limit, Backorder monthly variation, and open-backorder realism.

**Inventory & warehousing (28).** Healthy / out-of-stock / below-reorder / overstock target ranges, Inventory days of supply, Safety-stock days and factor, Default on-hand and reorder point, Overstock multiplier, Inventory variation, DC inventory health profiles, Inventory health / aging / adjustment rules, Velocity class weights / rules / distribution, Velocity-inventory interaction, Slotting rules and pick-count ranges, Cycle-count accuracy, DC capacity rules / scaling / utilization, DC stock fraction by capacity, and inventory final constraints.

**Shipping, carriers & delivery (50).** Number of carriers, Carrier on-time-delivery tiers, Carrier monthly variation, Carrier SKU-complexity weights, Carrier performance trend threshold, Region and DC on-time-

delivery tiers and variation, SKU-complexity delivery effects, Company on-time-delivery target range and center, Ship-via weights (segment / region / order type), Shipment-type profiles and multipliers, Multi-shipment and partial-shipment rates, Transit rules and seasonality, Tracking-event milestones, Freight rules and upcharges, Shipment cost noise and exception rate, and delivery final constraints.

**Transportation, fuel & labor cost (37).** Transportation base cost, weight and distance multipliers, and variation, Carrier type rates, carrier cost multipliers and variation, Carrier damage proxy rate, Cartonization rules, Accessorial fees and probabilities, Driver hourly / per-mile / per-stop ranges, Fuel gas and diesel base ranges, Fuel state offsets, Fuel change dynamics and quarterly trend, Fuel surcharge rules and by-carrier, Labor cost by region and department, Labor overtime probabilities and multiplier, Labor DC-tier multiplier and monthly multipliers, Labor shipment-volume scaling, State tax rates, and DC operating cost.

**Accounts receivable & payment behavior (24).** Segment payment behavior and personalities, Payment terms by segment, Payment term / timing rules, Credit-limit rules and segment credit limits, Customer credit-limit multiplier, AR aging targets and buckets, AR delinquency rules, AR reconciliation rules, Invoice status / tax / aging rules, Open-invoice percent / age / monthly variation, AR revenue ratio, and AR final constraints.

**Returns & credits (19).** Return rate percent, Return rate by category, Return rate target range, Return window ranges and realism, Return reason codes, Return condition weights and realism, Restock rate and restocking fee by condition, Return seasonal variation, Return timing rules, Credit-memo rules / aging / realism, and returns final constraints.

**Forecasting & demand signals (24).** Forecast accuracy range and by-category, MAPE by ABC class, Forecast ABC / lifecycle / class-seasonality rules, Forecast volatility and elasticity rules, Forecast promo rules, Forecast bias threshold, Forecast noise, Forecast horizon (fixed start / end), Lifecycle-stage weights, Demand seasonality index bounds, Demand promo-uplift baseline, XYZ class rules and CV bounds, and forecast final constraints.

**Executive & C-level financial layer (31).** Facility economics, Corporate overhead rules, Corporate staffing / salary by function, DC staffing model, Workforce rules, Labor standards, Operating schedule, Payroll burden and benefits by region, Salary bands by region, Bonus rules, Sales compensation plans, Pocket-margin rules, Accounts-payable payment rules, KPI budget rules, KPI projection method / horizon / confidence, Fleet rules, Equipment-maintenance rules, Internal-consumption rules, Safety-incident rules, Capital-expenditure rules, and dataset safety limits.

**Cross-domain, calibration & runtime (5).** Cross-domain rules, Global randomization, Cost final constraints, Default noise range, Output settings, and progress/log and chunk-size controls.

---

## Appendix C — table catalog

---

The dataset comprises 62 tables across these domains (each described functionally; row magnitudes are at default scale):

**Reference & master data** — vendors, manufacturers, vendor-item catalogs, distribution centers (with capacity and occupancy economics), customers and contacts, the item catalog, carriers, driver and fuel costs, and sales reps.

**Order-to-cash (tens of millions of rows)** — orders and order lines, invoices and invoice lines, and the accounts-receivable ledger.

**Fulfillment & logistics (tens to ~100M rows)** — shipments and shipment lines, tracking events, and backorders.

**Inventory & warehousing** — inventory and monthly snapshots, adjustments, and warehouse slotting.

**Procurement** — purchase orders, vendor performance, and vendor scorecards.

**Pricing, promotions & demand** — item cost and price history, promotions and promotional orders, marketing campaigns, and returns.

**Analytical time series (monthly)** — demand signals, forecast accuracy, customer lifetime value, price elasticity, backorder resolution, carrier performance, DC throughput, item velocity, replenishment signals, sales-rep performance, and promotional lift.

**Executive financial layer** — the margin waterfall (line-level list-to-contribution), facility costs, accounts-payable ledger, corporate overhead and budget, workforce headcount and corporate payroll, sales-rep compensation, fleet assets and maintenance, internal consumption, equipment maintenance, safety incidents, and the capital-expenditure schedule.

Pre-aggregated KPI tables and a library of materialized analytics views accompany the core tables to accelerate dashboarding and reporting.

---

*The PrimaForja Synthetic Distributor Dataset is a fully fabricated dataset created for analytics, development, demonstration, and educational use. It does not represent any real company, and contains no real personal, customer, employee, or financial data. Every figure herein is a configurable default.*

*To request access, a sample, a reduced-scale edition, or a custom configuration, contact PrimaForja at [info@primaforja.com](mailto:info@primaforja.com).*