How Pure Romance Redesigned Their Data Architecture, Saving 1000s of Hours per Month

Built correctly, a company's data architecture is invisible. The right systems can easily collect and exchange error-free data, and different end-user groups can effortlessly access what they need. However, as Pure Romance found, when the data architecture is optimized for short-term reporting needs and not to follow best practices, this can not only create frustration, but also cause misalignment with medium-to-long-term operational goals, affecting the bottom line.

Pure Romance's business and application logic evolved gradually, and was geared toward financial reporting. Although this helped day to day, it was unsustainable in the long term, as this financial reporting logic was built into the database. The data team became a bottleneck; not only were they relied upon to prepare reports for use by external consultants (with the financial logic extracted from the data), but they were also prevented from performing advanced analytics tasks and providing their real value to the company.

As such, the database was especially restrictive when it came to changing the company's strategic and operational goals. When the financial analysts wanted to make swift changes to their reports (for example, to include discounts in sales reporting), the data team was again tasked with wrestling with the existing logic in the database.

Company

Pure Romance

Pure Romance is a global direct sales company that makes over $350 million a year, with more than 50,000 consultants (all female) worldwide.

Retail

Data Management

Aiming to align the company's data architecture with its day-to-day operations, Pure Romance set about correcting several years' worth of false starts and finding the right data science platform. The company wanted to remove business and application logic from the database, so that internal financial analysts could work independently of the data team, external consultants could produce their own reports, and the data team could be freed to work on more complex tasks.

After a competitive head-to-head proof of concept, KNIME was determined to be the right choice.

The models of the past can be a drag on future business

Before it had a dedicated financial analytics department, Pure Romance built financial reporting right into its database. Since the data model which evolved during this time was primarily geared toward financial reporting, the ETL was heavily propagated with business logic stored on multiple layers. This dramatically slowed the entire process, and locked the data team into supporting exceptionally complex data sources that mixed dimensions, facts, and grain.

Every time internal financial analysts or external consultants wanted to create individualized reports, or whenever data needed to be pulled for advanced analytics (or even checking inventory levels), the company was bogged down by its database.

The company was using SQL Server Reporting Services for commercial reporting and hard-coded, canned reports which weren't editable on the consultant front end, with few reports targeting the business and executive layers. When the company hired a Database Administrator (now Director), this one-person data team had to create and design reports using Tableau, with consistent layouts and dates on the consultant front end.

This was a sub-optimal use of resources (since data experts can solve far more complex and urgent problems), and also restricted the company's overall reporting flexibility, given that business intelligence analysts could create their own reports if given the tools to do so.

It was soon evident that the company's OLAP cube (the structure optimized for quick data analysis) was also inadequately organized. It was predicated upon certain errors in the data not only being expected, but required. **When they tested the system by inputting error-free and corrected data, it caused the cube processing to crash.**

A first attempt to solve this: Tableau

The company purchased two Tableau servers to let both internal (financial analysts) and external (consultants) workers carry out their own reporting. Although Tableau enabled easy reporting with distributed access for different teams, new issues arose. For example, many of the datasets had to be published to both servers, and only SQL queries would enable Tableau's scheduler to access that data. So whether it was internally or externally accessed, both Tableau servers were being hit, resulting in an almost constant churn of data, limiting the window of opportunity to load new data from source systems.

Furthermore, in spite of these changes, the ETL still contained too much business logic, meaning the tables and views created for Tableau were deep and fat. So when data was accessed, either internally or externally, not only was the warehouse accessed twice, but it was done so using complex business and application logic embedded in the ETL — a circuitous and inefficient process. This process is also unconventional, as typically ETL is done before data enters the warehouse.

This prevented the data team from working on advanced analytics tasks, and these problems directly impacted the company's business intelligence structures, and ultimately its bottom line.

Pure Romance is a direct sales company. As Rachel Ambler, Director of Database and Business Intelligence Systems, puts it, they deal with consultants who range from "part-time ladies looking to make some pocket money to 'rock star' team leaders leading million-dollar teams." Any database problems have direct impacts across every business function and operation, with unreliable customer data affecting revenue. For example, poor data can prevent additional sales opportunities, offer incorrect personalized recommendations, and impact whether a product's availability is properly displayed on a sales portal.

Nevertheless, reengineering the warehouse was not a possibility, as it would take up too much redevelopment time and capacity. So the company began looking for a new tool.

Finding a new tool: A head-to-head proof of concept

Pure Romance conducted a proof of concept, comparing KNIME to another data science platform to see which would best solve its problems.

Flexibility and extensibility

Initially the company needed a flexible and extensible tool that could work with Tableau. Later they needed one that could work with both Tableau and PowerBI, both of which its data team now uses. KNIME's flexibility and extensibility was preferred over the competitor; it could only work with Tableau and PowerBI, and could easily integrate with other solutions at no extra cost, while integrations could be costly for the competitor.

What's more, KNIME's reusable components and Python scripting features enabled Pure Romance to remove the need to publish twice on Tableau. The fact that KNIME could run on Mac, Windows, and Linux was also advantageous for a company like Pure Romance, whose staff uses a mix of technologies.

A single source of truth

KNIME was selected over the competitor, as it enabled financial analysts to take the business and application logic from the ETL using their own workflows. They could access the data and reports they needed without making the data team into a bottleneck.

When the company also started using PowerBI, they found that it functions best with Kimballized data for performance, rather than deep and fat datasets. This forced the company to switch from its previous model to a Kimball-data one, a bottom-up approach in which data marts are formed based on the business requirements, as opposed to establishing a data warehouse and trying to make it fit the business needs.

By implementing PowerBI, the data department became responsible for the application logic but not the business logic. In terms of ETL, the department only became responsible for a minimal amount of effort in cleaning up the application layer. The ETL was slowly transforming into more of an EL(T), leaving transformation to the financial analysts.

In this new setup, KNIME is used to impose rules on the business logic of the Kimballized data, which is what domain experts now do. They can simply extract the business logic from the database using a workflow in KNIME, and then access the data they need, while the data team is free to focus on other tasks.

The data models are now better able to accurately represent the source systems data, and thus get Pure Romance ever closer to a single source of truth, and still allow the analysts to add business logic on top of these models to ensure the resulting data matches the needs of the business.

Support for all data sources and types

Having two servers meant Pure Romance needed both to have the same data sources, which meant both were required to run some very heavy scheduled extracts every hour of every day. However, they were restricted by Tableau's scheduler.

As KNIME could work with many data types and sources and enabled smart scheduling, the comparison in the POC was found to be **"a slam dunk for KNIME"**, as it allowed the data team to create a custom component in Python to publish to both Tableau servers simultaneously.

Total cost of ownership

KNIME offered zero startup cost, with KNIME Analytics Platform being open source, so the team could test and play with the tool before scaling to the on-premises version KNIME Server (now KNIME Business Hub).

The commercial offering provided much functionality out of the box, rather than charging extra for APIs, as the competitor tool did. Its ability to integrate with other tools and systems without additional charge was also a bonus for Pure Romance.

Partners, not a customer

Pure Romance found they preferred KNIME over the competitor because "KNIME was concerned with creating a partnership, and not just selling a product" as Ambler put it. She continued: "Everyone we worked with at KNIME, from the business to the technical teams, was extremely helpful, and went above and beyond in supporting our needs. The competitor, on the other hand, was more concerned with making a sale. Plus, I must say, KNIME offers a Smorgasbord of data manipulation goodness in its components and tools."

Outcome

By using KNIME and PowerBI with a new data warehouse, Pure Romance has been able to cut down end-to-end load time of the data from three hours to about one and a half – and are still looking at further time savings as they continue to utilize more and more aspects of KNIME's functionality. Over the course of the year, this time accumulates to many hours saved in data processing.

Furthermore, the data team is slowly being removed as the bottleneck: they are free to work on more advanced sourcing and cleanup tasks, while domain experts can access reports free of business logic without relying on the data team.

"KNIME was concerned with creating a partnership, and not just selling a product"

Rachel Ambler, Director of Database
and Business Intelligence Systems, Pure Romance