BIG DATA FOR BUSINESS

Title:

Current and Proposed data architecture for Novin Charm

PREPARED FOR:

PROFESSOR CATHY LIEM

Written by:

ALI HOSSEINMARDI

Student Number:24109150

THE HAGUE UNIVERSITY OF APPLIED SCIENCES

Abstract

This paper analyses the current data pipeline architecture at Novin Charm and suggests enhancements using insights from the principles of Big Data Analytics. Each step of the existing architecture is evaluated against relevant analytics concepts.

Contents

1. Introduction

Novin Charm is an Iranian company that makes high-quality leather products. They sell natural leather items like clothes, shoes, bags, belts, and accessories. Novin Charm is a trusted brand in the leather goods industry. It caters to the Iranian market and follows international quality standards. In 2003, the company adopted the "Novin Charm" trademark to show it makes high-quality products that meet global standards and suit Iranian tastes. The company has many stores in Iran. They sell leather goods for everyday use and for special occasions. It uses decades of local consumer knowledge and new production techniques to stay ahead in the market.

In my old job as a sales manager at Novin Charm, data collection was a hands-on, collaborative effort. Our methods were simple but time-consuming.

Our company needed a better data system because our business was growing. We needed to modernize to keep up with competition and growing customer bases. Our old system couldn't handle busy times or process lots of data.

To stay competitive, we need to upgrade our systems. By implementing modern tools and technologies, we aim to streamline operations, enhance decision-making, and provide a better experience for our customers. This paper outlines a plan to transition our company to a new data pipeline.

In this paper, I have focused on the CRM system's role in managing Novin Charm's customer feedback, inventory, and sales data from physical stores. Online sales have been excluded from this analysis, as they require separate systems and tools tailored specifically to e-commerce operations, which are beyond the scope of this discussion.

2. Current Data Architecture – Novin Charm

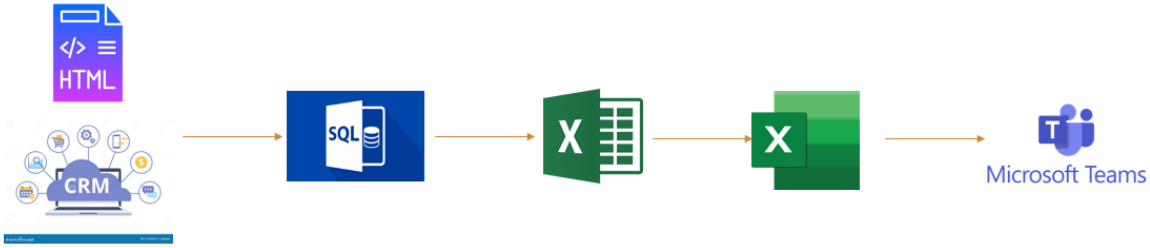

Figure 1. Current Data Pipeline for Novin Charm

2.1 Current Data Collection

In stage one, different data are collected from various branches and online shops. Such as customer feedback, inventory, number of sells. Data enter to **CRM** system. The Novin Charm sales team gathers and manages data. The team makes **HTML forms** to capture data, which is entered into by customers or the sales team. For example:

1. A customer, Sarah, visits Novin Charm's website to look at a new leather product. She fills out a form to leave feedback and subscribe to the newsletter. This data is captured in HTML.
2. The sales team records a new customer purchased five units of a trending leather product. This data is entered by employees manually.

3. Employee inputs inventory updates into the system manually.

As a sales manager at Novin Charm, I often had to enter customer purchase data into the **CRM**. I made sure that all data entries were accurate and complete.

2.2 Current Data Storage

Novin Charm stores data from **HTML** forms and **CRM** systems in an **SQL database**. The database organizes the data into tables, such as customer feedback, sales transactions, and inventory levels. This makes it accessible for further analysis. For instance:

• Sarah's feedback is saved in a "Customer Feedback" table.

• The sales team's data on product purchases is stored in a "Sales Transactions" table.

• Inventory updates are saved in an "Inventory Levels" table.

The IT team ensures the SQL database is secure and accessible.

2.3 Current Data Processing

In the third stage, the data from the **SQL database** is cleaned and prepared for analysis. At Novin Charm, the cleaned data is stored in **Excel**.

The data engineering team extracts data from the SQL database to clean and organize it. This cleaning process is manually now. It takes long time. It needs lots of employees to do it. It is not accurate.

This step ensures the data is reliable and ready for deeper insights.

2.4 Current Data Analysis

In the fourth stage, data is analysed to find patterns and trends. Data in Excel controlled by employees manually. However, Selling online becomes more, it takes long time to analyse all datas. In addition, sometimes mistakes happend by employees.

Novin Charm uses **Excel** for basic analysis and visualizations.

After analysing, they create a report. For example: the trending leather product sold 1,500 units last week, with 30% of sales coming from Sarah's branch. They determine the high sales were driven by a limited-time 20% discount campaign. They share these findings with sales managers.

There are 4 types of Analytics:

- Descriptive Analytics: Summarizes data to answer ‘What happened’’? Which products sold the most last month? (BDB EN week 2.pdf)
- Diagnostic Analytics: It looks for the reason why. The decline in sales is due to limited inventory. (BDB EN week 2.pdf)
- Predictive analytics: “What will happen’’?
- Perspective analytics: ’How can we make it happen’?

2.5 Visualise

At stage 5, Novin Charm uses **Microsoft Teams** for communication and data sharing. Teams lets managers work together, share files, and give updates in real time. However, it can only show static Excel charts or reports, which may not show complex trends or allow users to interact with them.

For example, a dashboard showing key metrics like sales, stock levels, and customer feedback could help executives extend discount campaigns more effectively. It could also alert managers to low stock of trending products, allowing them to approve restock requests and prevent lost sales.

3. Why Transition to Real-Time Analytics?

3.1 what is the current problems

Novin Charm has many shops across Iran and offers online shopping. Data is very important for our company. We need to know the number of products in each shop or warehouse. Categories, sizes, and colours of products are important too. We also care about sales volume, customer feedback, and inventory data. It's essential for us to have access to data from all branches and warehouses.

However, we face problems in different parts of our current data pipeline.

1. **System Performance**: During busy times, like sales seasons, our systems slow down or crash because they can't handle the large amount of data. This prevents customers from shopping online.
2. **Shop Systems**: Sometimes, the systems in our shops stop working, so employees can't check product data properly. In some cases, the system shows incorrect inventory data, causing issues.

3. **Data Analysis**: Our employees have to analyse data manually, which takes a lot of time and effort. They need to identify key insights about products, sales, and inventory to prepare reports. Sometimes the result is not accurate. We also need extra staff to handle this work.

3.2 What goes well?

1- Data Collection:

- Data Centralization: CRM organizes customer and sales data, reducing manual tracking efforts.
- Easy Customer Interaction: HTML forms simplify data collection from customers through surveys or feedback on the website.
- Basic Integration: CRM provides a centralized platform for storing customer interactions and basic operational data.

2- SQL databases are organized, familiar, and cost-effective. They store data in structured tables, making it easier to query data than raw spreadsheets or scattered files. Many team members or IT personnel understand how to use SQL for basic queries and storage tasks.

3- Excel is familiar to most employees at NovinCharm. It's easy to learn and can be used for simple data cleaning, basic calculations, and low-cost data aggregation. Excel is a tool that most employees at NovinCharm are already comfortable with. It requires little to no training for basic use.

Excel allows for manual data cleaning tasks, such as removing duplicates, standardizing formats, or sorting rows and columns.

Employees can perform simple operations like aggregating sales figures or calculating averages directly within Excel.

4- Most employees know Excel, so teams can analyse data without new training or tools. Excel is good for simple tasks like calculating averages or creating pivot tables. It supports many types of analysis and allows employees to customize their workflows.

5- Real-time collaboration. Microsoft Teams lets managers and employees share files, updates, and decisions instantly. It supports group chats, channels, and meetings.

File Sharing: Teams works well with Excel and other Microsoft tools. Managers can share Excel charts or summary files directly within Teams.

Communication: Teams organizes discussions into channels or topics, making it easier to track conversations related to specific projects or departments.

Accessibility: Teams is accessible on desktops and mobile devices, ensuring team members can stay informed from anywhere.

3.3 What does not go well?

1- 1st stage:

Fragmented Data Flow: HTML forms and CRM may not integrate seamlessly, requiring manual efforts to transfer data.

Limited Scalability: As data grows, the CRM may struggle with performance without database support like SQL.

Manual Effort: The current setup lacks automation, requiring significant manual input and management to ensure consistency.

2- As Novin Charm grows, the SQL database may struggle to handle larger datasets, leading to slower query speeds or the need for expensive server upgrades.

Storing and organizing data in SQL often requires manual input or basic scripts, which can be time-consuming and error prone.

3- Manual effort required. Excel is slow and error prone. Employees might delete important rows or fail to standardize fields.

No automation: Excel doesn't allow automated workflows, so repetitive tasks must be done manually, which is time-consuming.

As datasets grow, Excel becomes slow and unwieldy, often crashing or requiring users to break data into multiple sheets.

4- Time-consuming: Analysts must perform calculations manually, which slows down the process and increases the risk of errors.

Limited advanced analysis: Excel is not designed for advanced analytics like predicting trends or segmenting customers. For example, creating a forecast requires complex formulas or manual effort, which is prone to error.

5- Can't visualize much. Teams can display charts or files, but not advanced data. Teams often uses basic insights like Excel graphs or screenshots instead of dashboards.

Manual Effort: You still must prepare and visualize data in other tools before sharing it in Teams. This adds an extra step.

4. Proposed Data Architecture – Novin Charm

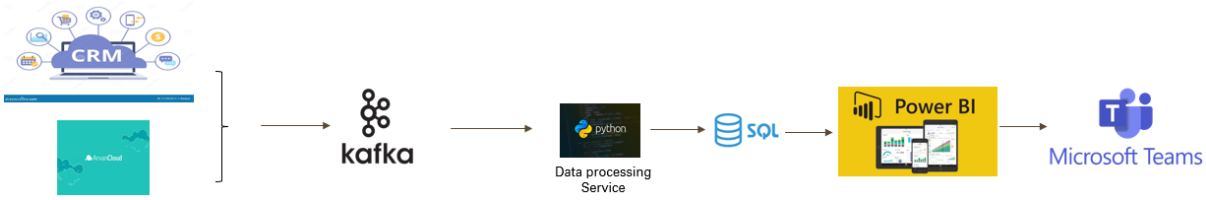

Figure 2. Proposed Data Pipeline for Novin Charm

4.1 suggested data collection

The **CRM system** collects customer interactions, sales data, and feedback from various sources, such as online forms, sales teams, and customer service. This data is then stored in an **SQL database**, which organizes it into structured tables for easier access and querying. For example, customer feedback can be stored in a "Feedback" table, while sales transactions go into a "Sales" table. This setup ensures that all data is centralized, making it ready for further processing in later stages.

4.2 Suggested storage data (ArvanCloud)

The concept of data lakes and distributed systems, in our case ArvanCloud, highlighted the importance of scalable storage and processing in managing and analyzing vast amounts of diverse data efficiently. (BDB EN week 2)

To address the downtime issues during the high seasons and operational costs of maintaining physical servers for the current applications, I suggested to migrate the whole solution to a cloud environment, in this case ArvanCloud. Arvan could is the biggest cloud provider in Iran which provide high quality services for the businesses to scale their applications with minimum operational costs (Like Amazon). In my suggested design, we move the CRM application to cloud to make sure that our online services are going to be responsive during high seasons. We are going to use their cloud database service alongside of their cloud storage. They also provide out of box security services that will help us to protect our customers data without compromising user experience. Also, we can use their Content Distribution Network (CDN) to improve user experience with our website on different devices. This helps our IT department to focus on the improving the website and support the business during the high demand times.

Although there are various good cloud storage, I suggest ArvanCloud due to prohibition exist in my country Iran. ArvanCloud is a cost-effective, scalable storage solution designed for high-demand times. It can store large datasets, ensuring reliable access and performance during peak loads. Its affordability makes it ideal for Novin Charm's budget while providing flexibility for data growth.

4.3 Suggested data processing (kafka)

In the third stage, Novin Charm can use Kafka. One of the issues of the current system is that real time analytics can not be gathered since the current system in busy with processing the incoming requests from the website and the country wide branches. So, any analytic should be done during the time that the system is idle. To address this issue, I suggested an event driven design and the CRM system will generate multiple type of events and send them to another service for processing. This will help us to gather more data about the users' behaviours and use them to improve our services and expand our reach. I suggested Kafka as the data streaming platform and all the users (customers, employees) interactions with the application we generate a new event then it will be sent to Kafka for further processing. With Kafka we can partition the events in different groups and that will give of the flexibility to process each group in their own related services. For example, user can buy, like, comment, tag, and etc a specific product and each of them will be an event that we can use them for analytics and improving the existing services.

Kafka handles real-time data streaming efficiently, making it perfect for processing events like customer transactions or inventory updates. It ensures that data flows smoothly from different sources into ArvanCloud for storage or further processing. Why Kafka:

- Efficient Data Movement: Kafka allows you to process events (e.g., sales transactions, inventory updates) as they happen, ensuring data is immediately available for storage or further processing.
- Scalability: It can handle high data volumes, making it suitable for busy periods when Novin Charm processes many customer interactions.
- Flexibility**:** Kafka integrates seamlessly with tools like Python and Power BI for downstream analysis and visualization.

Departments to Involve:

• IT: Setting up the Kafka environment, including brokers and topics for event streaming. Ensuring integration between Kafka, the SQL database, and downstream systems like.

4.4 Data Processing Service(s) (Python)

A new service(s) suggested to be the downstream of the streaming service with a specific task, processing the data and prepare them for analytic. Data Processing Service will get the events and responsible to process and store them in the related tables which later will be used to generate the business reports and analytics.

Python is suggested for writing this service based on the IT team's experience and useful libraries to work with data. This service, for example, can remove duplicates, fill missing values, and standardize formats, ensuring the data is accurate and ready for analysis.

In Novin Charm we can hire some programmer to code with python for cleaning data.

4.6 Visualization (Power BI)

Prescriptive analytics taught me how optimization models, such as linear programming, are used to minimize costs or maximize profits, bridging the gap between predictive insights and actionable decisions.

The last step is using Power BI. It offers interactive and real-time dashboards for visualizing key metrics like sales, customer feedback, and inventory levels. It's user-friendly and integrates easily with Teams for sharing insights.

Departments to Involve:

- Managers and Executives: To define which insights are most valuable.
- Data Visualization Team (or Analysts): To design the dashboards.

4.7Collaboration (Microsoft Teams)

Teams enables seamless communication and collaboration, allowing dashboards and updates from Power BI to be shared efficiently across the organization.

5. Implementation Strategy

In the next year, we'll move the existing NovinCharm's platform from in-house servers to ArvanCloud with more advanced pipeline using Kafka, and Power BI.

5.1 Implementation Timeline

1. ArvanCloud can be setup up approximately in 2 month.

We need to create an ArvanCloud account and design the new architecture including servers and database. A data migration from the SQL database to ArvanCloud will occur very seamlessly since the cloud database service will give us the same database that we have already (Mysql).

ArvanCloud offers scalable storage solutions and quick setup, with basic plans ready for immediate use. (https://www.arvancloud.ir/en/pricing)

2. Kafka Setup and Configuration (1–2 Months)

- Why it Takes Time: Kafka requires installation and configuration of brokers, topics, and partitions.
- Realistic Timeline: Small to medium businesses typically need about a month to complete these tasks, especially if the team is new to Kafka. Advanced configurations for resilience (e.g., replication, fault tolerance) may take additional time.

Integration with Other Systems (1–3 Months)

- Why it Takes Time: Kafka must integrate with existing tools like ArvanCloud (for data storage), CRM systems (for event streams), and Python scripts (for data processing). Developing connectors or APIs to enable this integration adds time to the process.
- Realistic Timeline: Integration depends on the complexity of the data flow. For example, setting up streaming pipelines to process customer data in real time requires custom configurations.

Testing and Optimization (2 Month)

- Why it Takes Time: After setup, the system must be tested to ensure data flows correctly, handles high loads, and avoids message loss. Fine-tuning configurations, such as retention periods and partitioning, ensures Kafka runs efficiently

(https://www.arvancloud.ir/)

3. Hiring additional programmer for Python. (2weeks)

5.2 Cost Estimation

ArvanCloud: (https://www.arvancloud.ir/fa/pricing/databases#business)

Free for basic storage (5 GB with 20 GB traffic per month). For larger plans:

Growth Plan (500 GB storage, 2 TB traffic): ~$9/month (€9).

Professional Plan (5 TB storage, 20 TB traffic): ~$99/month (€99)

Figure 3 Table of cost plans for ArvanCloud

Kafka:

Open-source and free. Implementation costs may include server hosting and configuration, roughly $500–$1,000 for initial setup (if using in-house IT).

Python:

Free (open-source). Training costs ~$200–$500 per person for 2–3 months (Coursera, Udemy)

Power BI:

Pro License: $10 per user/month ($120 annually per user). Premium options are available for advanced features.

Microsoft Teams:

Included with Microsoft 365 subscriptions (~$12.50/user/month for Business Standard)

5.3 Governance and Security Considerations

Cloud providers provide out of the box security solutions and protocols that can be configured easily via their customer portal, a secure and reliable alternative to traditional physical in-house servers. For business owners, this shift offers significant peace of mind due to the fact that cloud providers like ArvanCloud take on the responsibility of maintaining the infrastructure's physical and digital security. This allows business owners to focus on operations without worrying about technical risks. For example, DDoS protection, and secure authentication to safeguard data.

5.4 Legal and Ethical Considerations

When handling data in Iran, you must comply with global standards and local laws. Here are some key things to think about:

1. Iran doesn't have comprehensive data protection laws like the GDPR. But several laws cover data privacy and cybersecurity.

The Computer Crimes Law (2009): This law addresses data breaches and hacking. Companies must protect data from unauthorized access.

E-Commerce Law (2004): It governs how companies protect customer data during online business.

2. Ethical concerns:

Data consent: Users must agree to have their data collected or processed, especially personal information. Be transparent about how you use data.

Data Security: Data must be protected from breaches. Use encryption and restrict access to sensitive data.

Avoid discrimination in algorithms or analytics, particularly in sectors like hiring, finance, or service accessibility.

3. Compliance Challenges

While Iran lacks GDPR-like regulations, international partnerships may require compliance with GDPR or equivalent laws.

Ensure compliance with international sanctions, as some services may have restricted availability in Iran.

6. Conclusion

Moving NovinCharm's data pipeline to modern tools like will make it more efficient, scalable, and easier to make decisions. The plan will address current challenges by using automated workflows, advanced analytics, and interactive visualizations.

This can be done in 12–18 months at a low cost. By working with the IT team, data analysts, and managers, NovinCharm can upgrade their system to will the company make better decisions, work more efficiently, and grow faster.

7. References

Novin Leather. (n.d.). *Novin Leather official website*. Retrieved December 8, 2024, from https://novinleather.com/

ArvanCloud. (n.d.). *ArvanCloud pricing and services*. Retrieved December 8, 2024, from https://www.arvancloud.ir

Microsoft. (n.d.). *Power BI pricing*. Retrieved December 8, 2024, from https://powerbi.microsoft.com/en-us/pricing/

Coursera. (2024). *Python for data analysis and visualization*. Retrieved December 8, 2024, from https://www.coursera.org/

Udemy. (2024). *Learn Python for data cleaning and analysis*. Retrieved December 8, 2024, from https://www.udemy.com/