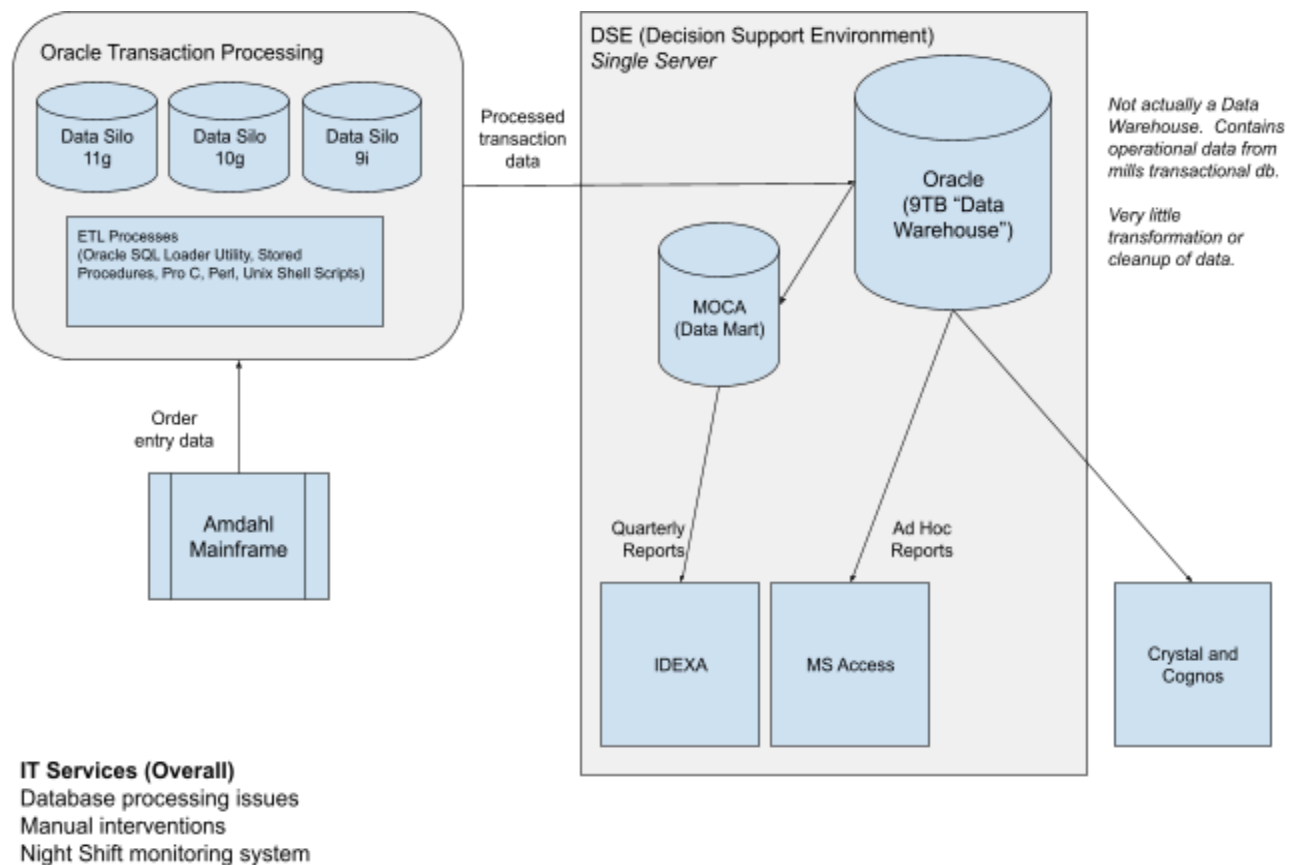


DAT 202 Group Activity #1

Date: June 26, 2024

Group Members: Colin Bowers, Munique Crespo, Qianxi (Emily) He, Jefferson Dos Santos

Based on the description provided, prepare a diagram of Lofasco's operational environment.



List some of the challenges that Lofasco is facing in managing its data and rank them (from greatest to least).

1. **Data quality and governance issues.** Firstly, because different business applications use inconsistent naming conventions, it causes confusion and errors in data handling. Secondly, customer data across different channels and data silos reduce the accuracy and data reliability. In addition, absence of tools and processes for data cleanup and consolidation result in poor data quality and inefficiency.

2. **System and infrastructure limitations.** Since many data silos are running versions of Oracle's RDBMS (11g, 10g, and 9i), there may be a lack of support and modern features. Furthermore, it may cause potential risk and complexity to rely on an Amdahl mainframe for order entry. As explained in the case, the decision support environment (DSE) experiences space issues and performance slowdowns, particularly when MOCA runs, it causes delays in critical reporting systems like IDEXA.
3. **ETL process complexity.** The ETL processes involve a mix of Oracle SQL loader utilities, stored procedures, Pro C, Perl, and Unix shell scripts, making it difficult to manage and maintain. In addition, frequent need for manual intervention to address issues during night jobs reduces efficiency and increases risk of errors.
4. **Inefficiency in reporting and analytics.** Ad-hoc reporting processes and lack of consolidated reporting tools make it slow and unreliable for the reports. Absence of standard enterprise BI reporting tools and reliance on older client-server applications like Microsoft Access result in slow query performance and user dissatisfaction. Also, there are minimal data analytic tools, which limit the access and ability to gain insights from data and make informed decisions.
5. **Risk of regulatory compliance.** There are concerns about the accuracy of MOCA data, especially from the US Department of Commerce, which pose a significant risk to Lofasco's business operations in the US. The lack of traceability of how cost figures in MOCA are delivered increases the compliance risk and trust issue of the data.
6. **Employee productivity issues.** Employees struggle with navigating through multiple data sources and inconsistent manual data entries, leading to inefficiency and errors. Due to the data issues, employees may also have low productivity.
7. **Strategic challenges.** The company faces uncertainty about how to leverage Big Data technologies (e.g. Hadoop cluster, DataBricks, etc). Furthermore, the investment in closed one-off solutions in the past creates more maintenance work. This fails to solve business problems or enhance productivity.

If Lofasco must pick one data enterprise solution to invest in, would it be (Explain your rationale for your choice):

1. **Robust data profiling / validation tools**
2. **Business Intelligence Solution for centralized reporting**
3. **Metadata Management tool**
4. **Big Data environment / Hadoop Cluster**
5. **or one of your selection**

We would recommend better Data Engineering infrastructure. Consolidating data silos into a single version of Oracle would lessen maintenance by the IT team. A more robust and modern ETL system, like Airflow, will be more performant (using less system resources), fault tolerant (reducing on-call issues during night shift) and automated (reducing manual intervention). Segrating systems to multiple servers will reduce the impact of long running processes like the IDEXA on other systems.

Investing in this first before the other areas listed will allow the company to gain better control of their labour allocation, operating costs and data quality. This will make future changes easier, cheaper and more effective.

How can a metadata management tool possibly benefit Lofasco?

1. **Enhanced Data Quality:** Enforces standardized naming conventions and improves data validation and cleansing processes, leading to more accurate and reliable data.
2. **Regulatory Compliance:** Provides improved data lineage, accuracy, reliability and increased transparency in data addressing concerns from regulatory bodies like the US Department of Commerce.
3. **Increased Productivity:** Centralizes metadata in a single repository and automates documentation, reducing the manual effort required from the already overloaded IT team.
4. **Efficient Ad-hoc Reporting:** Reduces time spent by analysts navigating numerous and inconsistent data sources.
5. **Consolidation of Tools:** Facilitates any future effort for consolidating reporting tools.
6. **Better Data Integration:** Identifies and consolidates duplicate data across silos, optimizing ETL processes and improving data flow.
7. **Technology Strategy:** Supports the migration to big data environments like Hadoop or DataBricks and eases the integration of new BI and analytics tools.
8. **Business Strategy:** Company data needs to be more efficiently and effectively understood, exploited and managed in order to remain competitive.
9. **Improved decision making:** with better data quality and consistency, the company can rely on accurate and timely data to make strategic decisions. The management tool could also facilitate advanced data analytics, which provide deeper insight and enable informed decision-making.

What key problems and challenges referenced in this Case Study can be addressed by the following:

1) a robust Data Governance organization

A robust Data Governance organization can significantly help Lofasco by addressing their data management issues, enhancing data quality, improving compliance, and supporting better decision-making. **Data Standards and Policies** should help create consistency to name conventions, data formats across all the systems. Establishing policies for data validation and cleaning would help to ensure the accuracy and reliability across different business applications and overall data quality in the processes. **Enhancing Compliance and Risk Management** to ensure data management practices comply with industry regulations and standards such as the required by the US department of commerce. Implementation of audit trails to maintain records available detailing changes and demonstrate compliance. **Optimize Data Management Process through streamlined ETL processes** to reduce manual interventions and automate routine in data monitoring and tasks to correct errors. Introduction of **data**

governance tools and platforms that provide features like data quality management and profiling, metadata management should help to ensure continuous improvement in data governance practices to handle large volumes of data more efficiently. Also, **training and awareness programs** to clarify any questions regarding data management and the importance of data governance and roles in the business. Implementation of a **metadata management repository to create a centralized source of data documentation** and data definitions, relationships and usage. A **data catalog** would help users and data owners to understand schemas and data available within the organization to improve data quality and utilization.

To conclude, a robust data governance organization can transform Lofasco's management practices leading the company to higher data quality, improve compliance and help to establish efficiency to processes and operations. Support IT and Business alignments and collaborations to foster better data initiatives and decision-making.

2) better data preparation

Better data preparation addresses all efficiency issues, problems with data validation, and consolidation. The advantages would be:

- Uniform data policies and procedures which will increase overall transparency, accuracy, reliability and trust in the company's data.
- Better data preparation can streamline ETL processes, reducing the need for manual intervention and the risk of errors during data operations. Also, the automation of ETL tools and delivery of reports
- Efficient data preparation can optimize data storage and processing, solve space and performance issues in the DSE and other systems.
- Standardizing data formats and definitions through a comprehensive data dictionary would create and maintain easy access to data analytics and ensure consistency in outcomes.

3) data quality processes

The implementation of the data quality process would address the Data accuracy issues, determine proper naming conventions ensuring the quality of data entered into the system and finally, it would fix the customer information complaints.

The benefits would be the enhancement of data integrity and accuracy which would lead to more reliable decision making and inputs from the employees.

High-quality data can improve the performance of ad-hoc queries and reporting tools. It can reduce waiting time and enhance user satisfaction.

Are there any security precautions can Lofasco implement to increase the privacy of their information (due to cross border trading).

To enhance the privacy and security of their information, Lofasco can implement several key measures:

- **Data Encryption:** Implement strong encryption protocols to protect sensitive information both at rest and in transit, ensuring data confidentiality.
- **Access Controls:** Enforce strict access controls with robust user authentication methods to limit data access to authorized personnel only, reducing the risk of unauthorized exposure.
- **Compliance with International Standards:** Adhere to international regulations such as ISO27701 for trading in the US and GDPR for Europe to ensure legal compliance and secure cross-border data transfers, thereby maintaining data integrity and trust.
- **Data Masking and Anonymization:** Utilize data masking techniques in non-production environments and anonymize personal data where possible to prevent unauthorized access and protect privacy.
- **Regular Security Audits and Assessments:** Conduct frequent security audits and assessments to identify vulnerabilities, ensuring adherence to security policies and regulations while proactively addressing potential risks.
- **Continuous Monitoring:** Implement continuous monitoring of systems and networks to detect security incidents promptly, mitigating potential threats and suspicious activities effectively.
- **Employee Security Training:** Provide comprehensive security training for employees to promote awareness and best practices in handling sensitive information, including conducting phishing simulations to enhance their ability to recognize and thwart attacks.
- **Data Backup and Disaster Recovery:** Perform regular backups of critical data and systems to facilitate data restoration in case of a security breach. Develop and maintain robust disaster recovery plans to minimize downtime and recover data swiftly in the event of a security incident.

By integrating these measures, Lofasco can significantly bolster its information security framework, safeguarding data privacy and integrity amidst cross-border trading operations.