White Paper

Digital Transformation with Data Warehousing

Addressing data challenges to inform successful business

By Ben Edwards

MACEDON
TECHNOLOGIES

Table of Contents

Introduction

Businesses today face difficult challenges when trying to make important decisions, often relying on great amounts of data. If that data is not readily available to the people responsible, making informed and timely decisions is almost impossible.

Some of the data challenges that impact decision-making for businesses include:

- **Poor Quality Data**
- **Unmanaged Quantity of Data**
- **Increasing Quantities of Data**
- **Receiving Bad Data**
- **Data Security**
- **Inconclusive Data Analytics**

Each of these issues can be solved using data warehousing. A data warehouse is a system that stores data from potentially multiple different sources in one place, with the intention that this data can be easily used to make important business decisions in the future. To ensure that these issues are handled, and that all data within a data warehouse is trustworthy, warehouses can rely on a number of different functional solutions. The main categories of solutions available include:

- **Data Integration**
- **Business Intelligence (BI)**
- **Security and Governance SQL Development and Management**
- **Application Development Interfaces**

Data Integration

The most commonly referred acronym that describes data integration with data warehouses is **ETL**. ETL stands for "Extract, Transform, Load" and is meant to describe the process of exporting data from other systems, modifying that data so that it adheres to standards defined by the target system with
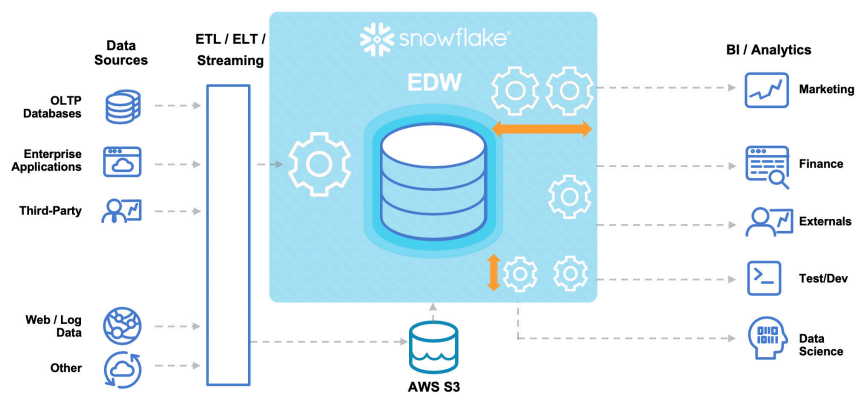
Snowflake image from AWS [2]

pre-built functions or processes, and then loading it into a database. The topic of performing ETL with a digital transformation is explained further in John Peacock's white paper on data migration.[1]

Data integration also encompasses more operations like data preparation, data migration, and data warehouse automation. The main idea behind data integration is to provide the means for a data warehouse to get data from other systems. By pulling data from other systems and making sure it is stored in an acceptable way in your warehouse, you can potentially solve problems of not having quality data, too much data, or bad data. There are many solutions to help perform the previously mentioned functions, including Alooma, Fivetran, and Informatica.

When considering data integration, a business should also consider the challenges previously mentioned, and whether or not a solution will help in solving the problem or could potentially make the problem worse. For example, if the ETL solution being used by a data warehouse is incredibly large amounts of data from an external database for storage in a data warehouse, the problems of holding poor quality or too large an amount of data can certainly arise. If ETL is performed in this sort of scenario, without first verifying that the ETL process will filter or improve data in a desirable way, then the mentioned issues can happen.

While workflow automation tools, such as Appian, can perform automated processes, it is not as well-suited as some products mentioned here when it comes to performing ETL and other data integration operations. Appian excels at process execution, and data kept in-process for automated ETL tasks may take up necessary application server resources that could be used for other important processes.

Data warehouses are valuable because they can help centralize ETL processes. Applications being built with tools that may not perform ETL efficiently can instead be built with a focus on data presentation, knowing that the data retrieved from the data warehouse is accurate.

Business Intelligence (BI)

BI tools are applications that collect and process large amounts of data from different systems. The main function of such tools is to amass data and make it available via search and then presenting results with visualization through reports, charts, dashboards, and other graphics. BI tools help enable analysis of, reporting on, and discovery of data to help businesses make informed decisions. Business intelligence can often overlap with data integration because many BI solutions have data integration components built in. Some popular BI tools include Tableau, Qlik, and ThoughtSpot.

BI tools help prevent inconclusive data analytics by helping users visualize data. With access to BI tools, analysts are able to find and share insights based on visualizations, which helps organizations make better decisions

with a more detailed picture of their ongoing operations. Data must be standardized in order to work with BI tools, typically, which helps make sure that visualization is accurate. Appian designers can build reports 5 and other BI-like components, but these are often not replacements for utilization of a BI-focused tool. Appian can present the same visualizations from other BI tools by embedding or installing plug-ins. However, the most important aspect of interaction with data visualizations is that the data is accurately reported from a source, like a data warehouse.

Business Intelligence Tools

Security and Governance

By utilizing security and governance tools, sensitive data can be maintained and protected from unsecure access. Such tools can help organizations achieve regulatory compliance as well. Many operations fall in the realm of security and governance, including instruction detection, data monitoring, data masking and risk assessment. Security tools may be limited with their interaction with data warehouses based on the type of connections and encryption required between data layers, but information about different data warehouse solutions and how they handle security is available through that solution's documentation.

Some popular tools include Informatica, Collibra, and Talend. If Appian interacts with a data warehouse that is adhering to specific security and governance guidelines, there may be some considerations that need to be made when retrieving or sending data. By default, Appian adheres to TLS when encrypting data at rest and uses SSL to secure interactions with an application server, so further steps may need to be taken if systems that Appian interacts with have stricter security requirements.

SQL Development and Management

Even while using a data warehouse, creating and managing databases well is incredibly important. Data warehouses can contain databases, schemas, tables, and other data components. Interacting with such components is often done through some form of SQL. The SQL used to develop and manage databases inside of a warehouse can often look and operate like relational SQL, but is used as a means to communicate similar functions between a warehouse and data storage layers.

By maintaining the data storage layers inside and outside of a data warehouse, problems could potentially be prevented before they even start. The common problems of bad data or too much data being stored in a database table can be prevented if administrators think about the data model being used ahead of time. Often-times, databases can have problems with data being insufficiently normalized, making it so that different field entities are being stored in one place where it may not make sense. Having administrators normalize such a database or tables can help reduce the amount of data stored in specific tables and allow for smoother operation. Preventing such issues is important when using relational databases because a poorly designed database may be difficult to use and maintain, may store incorrect data, and may even fail to function. Considering this is equally important if utilizing a data warehouse as a data source.
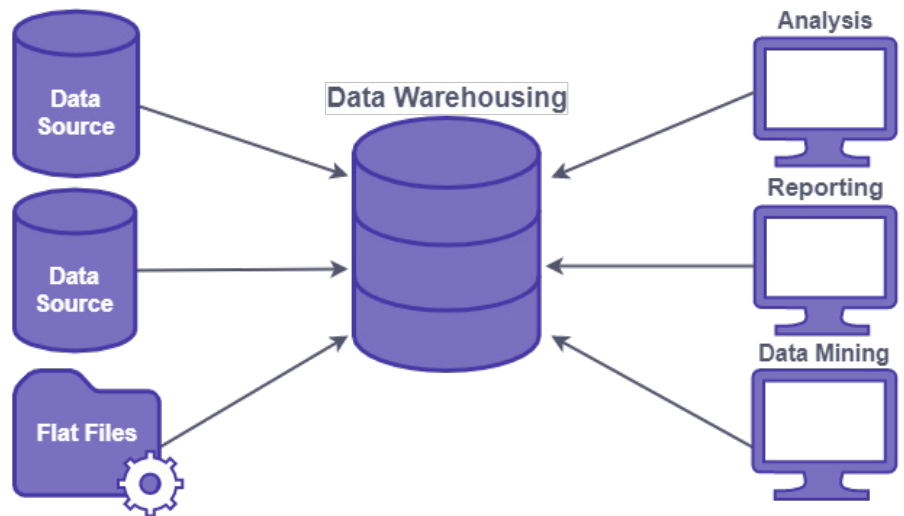
Application Development

When developing an application that interacts with a data warehouse, it is incredibly important to consider the mechanism which will let that happen. You must consider that the data model and application need to be compatible 6 with the chosen integration. Without these considerations, more issues could occur with application development; such as data inconsistencies, increased maintenance, or poor visual data representation.

A public web service API can be used to build integrations between applications and a data warehouse to retrieve and send data back and forth between a warehouse and external systems. Most popular warehouses with a public API available provide a RESTful pattern for integration. However, some data warehouse solutions, like Snowflake, do not provide a public API. In these cases, an alternative connection mechanism is required to receive data from and send data to a data warehouse. Alternative connection mechanisms could be built with programs that provide access to a data warehouse, like utilizing a JDBC connector.

No matter if an application is being developed to integrate with an API directly or will use a programmatic interface, the application should be built with retrieval of important data from the data warehouse in mind. Given there are a number of data warehouse solutions, different languages are supported by different types of warehouses and programmatic interfaces available. Many popular solutions support interfaces for popular languages like Java (using a JDBC driver usually), .NET, C, Python, and Go. Any plug-ins developed for Appian would be written in Java and utilize a JDBC driver, even if the solution was officially built into the product by Appian.

Data Warehousing Products

There are a number of cloud and on-premise data warehouse solutions, as well as software that integrates or works with data warehouses. When considering these solutions, there are a number of factors that a business should weigh when it makes the decision for what tools to use. On-premise solu-

tions have to be installed and maintained within an organization, while cloud solutions can be easier to handle because an organization can pay another company for usage and support of their tools.

Many data warehouses and other related tools are provided as "Software as a Service", or SaaS. This can be helpful for many organizations because SaaS solutions are often easy to get started with. If an organization uses a few SaaS tools to solve most of their needs, there are many benefits, in comparison to maintaining too many systems at once. Support and documentation for SaaS solutions are often better than having to piece together instructions for a number of different tools. For data warehousing, using a SaaS solution can also hide many of the back-end interactions that 7 are necessary to make sure that the data stored in the warehouse is the correct data that will be presented to business leaders.

Solution providers like Amazon Web Services (AWS), Microsoft, Google, or Oracle all provide different SaaS tools that can provide storage for a data warehouse, as well as many means to perform operations like ETL, integrations, or BI implementations. For large providers, there are a number of tools that integrate well with one another in the same ecosystem.

Cloud-based data warehouse solutions offer many advantages. A major advantage is having maintenance and support provided by a vendor, instead of inside of your organization. With a cloud solution, an organization would not have to worry about installation or configuration maintenance, which can help reduce reliance on having as many experts in networking, security and infrastructure. Another advantage of using a cloud solution is that they can be dynamically scalable, meaning the size and speed of a data warehouse could change depending on the amount of data storage needed and the needs of systems interacting with the warehouse.

Solutions from particular vendors are often able to interact well with other tools developed by those same vendors. For example, AWS offers Redshift which can interact with Quicksight for analytics or Aurora and RDS to retrieve data from operational databases. Through interacting with an S3 Data Lake,

other integrations are possible as well. As a slightly different type of solution, Snowflake offers what can be called a "Data Warehouse as a Service", or DWaaS. By setting up a cloud data warehouse, an organization has direct control of everything contained within and pays for their usage. Snowflake touts a large "ecosystem" that shows the amount of different partners and providers that provide solutions in categories previously discussed, that integrate directly with Snowflake.

Digital Transformation with Data Warehousing

Macedon is positioned well to help build digital transformation tools that work with data warehouse solutions. Whether that means building Appian applications, integrations for cloud database solutions or working with more specific data warehouse tools, the results will help enhance a business's data warehousing needs. Many database solutions discussed in this paper are unsupported by Appian natively, which means that often integrations or other custom solutions must be implemented. However, once a connection is made between Appian and an external system, the data received can be translated to Appian data types.

This translation is essential to let Appian applications take the data from data warehouses and present it to users for essential functions like reporting and processing. This would inevitably help reduce many of the pain points discussed earlier in this paper. By having all data in a warehouse and providing a connection between Appian and that warehouse, any user of an Appian system should be able to trust that the data from that warehouse is accurate and will help guide their organization better. Those users would also not have to worry about issues that could arise from having to integrate with multiple systems or databases and use a data warehouse as a source of truth.

There may be some difficulties incorporating a data warehouse in conjunction with a system like Appian, but there may be challenges with other types of data sources as well. After establishing a data warehouse 8 and connecting with it, the connecting system and its users would not need to worry about the issues discussed in this paper. Considering all of this, enterprises can implement a data warehouse and be confident that they are making a good choice in the long-term.

Conclusion

Data warehousing can provide a business with a consolidated source of truth for information in ways that can be easily visualized and reported on. With the focus on helping drive business decisions, data warehouses can be a viable way to gather data from multiple sources using methods so that users will not have to worry about common data sourcing issues. The problems that data warehouses hope to solve, as explained earlier, will become less of a concern for a business taking advantage of data warehousing. Through avoiding such problems and being able to make smart decisions, a business can expand their business reach and thrive by spending more effort on other

important tasks. When a business chooses to digitally transform with a data warehouse implementation, they can be confident that they are making a smart decision that will lead to longterm, data-driven effects.

About the Author

Ben Edwards is an experienced Appian Enterprise Architect, skilled in web application design, database technologies, integration implementations, CI/CD, and varied new technologies.

[1] John Peacock, Data Migrations (2020),
https://macedontechnologies.com/white_papers_lp/datamigration/

[2] "Snowflake's DataWarehouse Can Be Rapidly Deployed on AWS." Snowflake Data Warehouse, AWS, July 2019,
http://aws.amazon.com/solutionspace/financial-services/solutions/snowflakedata-warehouse/