Modern BI for All™: A Field Guide

PART TWO
 DATA AGILITY
Introduction

In Part 1 of our Modern BI for All Field Guide, we introduced modern BI principles and best practices including data agility, data literacy, and intelligent action. In this guide, we'll dive deeper into the foundational layer of Modern BI for All: data agility.

Data agility is about accessing and analyzing data at the speed of business, replacing queries that used to take weeks or months with sub-second response times, without compromising data governance or data security.

In an IDG Communications market research report on Data Integration and Exploration, 90% of data vision leaders perceive the integration of multiple data sources as critical to improving business outcomes. Business users and analysts still can't leverage 20-40% of dark data that sits outside of their cloud data warehouse and analysts spend 80% of their time chasing down this data to do their work.

Enterprises are using a variety of BI applications to solve this problem, creating headaches for IT. As a result, IT loses visibility, governance, and certification on data that gets pulled offline for analytics.

This means that data often gets put back into silos that data warehouses initially aimed to solve for. The key to solving these challenges is making all data accessible at cloud-scale with an extensible, bi-directional data integration framework. This includes the ability connect to any source, thus unlocking all data in an organization. This also suggests utilizing a data fabric from various cloud systems and other sources—so data from all sources can be combined, augmented, updated, and then written back out to the source systems or pushed into other platforms.

Data from the cloud and the internet now coexists with enterprise data. Many organizations have data sources and targets on-premises and in the cloud, because they have embraced big data, social media, the Internet of Things (IoT), SaaS applications, and cloud storage—some of which are challenging to ingest and manage in legacy data warehouse environments. To accommodate this hybrid data environment, future-facing organizations need to modernize and extend their cloud data integration infrastructures to fully support the web and the cloud with more robust governance and data lineage capabilities.

60% OF DATA LIVES IN ENTERPRISE DATA WAREHOUSES (ON AVERAGE). THE OTHER 40% IS DARK DATA.
Like shadow IT, dark data increases the risk of data loss and leaks, lowering governance and reducing control of data within the organization. In 2020, according to Forrester, 25% of global data and analytics decision makers said it took months to get analytics help from IT, while 8% said it took more than a year. This extended data lifecycle puts a strain on any business when decisions need to be made quickly.

Challenges with integrating disparate datasets

Building modern analytics products is dependent on integrating data from disparate and often fragmented sources. Ineffective data integration strategies, complexities within IT infrastructure, over-reliance on application integration, limitations in governance, and enabling data access all play a role in limiting the potential of an enterprise’s analytics suite.

According to Forrester, traditional data integration fails to meet new business requirements that demand a combination of real-time connected data, self-service, and a high degree of automation, speed, and intelligence. New and expanding data sources, batch data movement, rigid transformation workflows, growing data volume, and data distribution across multi- and hybrid cloud environments exacerbate the issue. While collecting data from various sources is often straightforward, enterprises often struggle to integrate, process, curate, and transform data with other sources to deliver a comprehensive view of the customer, partner, product, and employee.

As organizations look to modernize and become more digital and agile, the key factor to their success is how data is stored, how it flows, and how it's accessed throughout the organization to ensure the capabilities to make better informed decisions—faster.
How siloed data negatively impacts your business

Data is not accessible to business users
Analysts or power users need the ability to bring in their data for analysis. However, current architecture makes it challenging for IT to extend their back-end tech (ETL or iPaaS, cloud data warehouse, etc.) to these individuals. Limitations in flexibility and cost, security, governance, and ease of use are reasons they can’t extend these technologies.

ROI on data integration investments isn’t realized
Most CIOs today have invested a significant amount in their data warehouses but find it difficult to access, extract, transform, and put critical data in the hands of business users. They continue to rack up technical debt—spending countless hours building operational processes while complexity compounds. Data warehouse projects often aim to solve inefficient integrations by adding more business processes to an already abundant litany of jobs and IT data access requests.

Data governance is difficult
Once technical business users get access to data through analytical cubes, IT loses control of the data and has minimal means of maintaining governance once intelligence is extracted. The speed of business continues to accelerate, demanding that data, sources, and data-driven products (reports, insights platforms, and analytics services) be delivered for business use in minimal time.

Disparate tools for integration and BI
Organizations may use unified tools across the entire company, but often have disparate tools because they serve different purposes. For example, Marketing will use one tool, and Finance will use another.

Manual and time-consuming processes
Processes for importing, cleaning, and preparing data for further processing are usually manual and very time-consuming, hindering operational excellence and holding back progress for departments within organizations.

Low performance of data pipelines
Even if automated, latency and low performance of data pipelines negatively impact the speed of data processing and analysis.
Best practices for cloud data integration

Keep both IT and business users in mind

To properly build superior analytics practices, IT vision leaders need to get their data in the hands of business users—much faster. This access includes the ability to integrate all your data sources (including clouds, applications, servers, existing data warehousing, and big data analytics platforms) to build a unified source of truth at cloud scale, wherever your data resides and without the need to move sensitive or referential sources. It also means the ability to give users access to all data in real time for use in any BI or execution tool, while staying in full control of data governance. IT sets policies and user permissions with the flexibility and ease of use required for large collaborative enterprises, and to serve the needs of multiple stakeholder groups.

• Any modern BI tool should have the ability to connect to data from any source, whether it’s a data warehouse or lake, cloud system, on-premises system, or files—or all of the above.
• Whether you have data in the cloud, in on-premises systems, behind your firewall, in proprietary systems, in data warehouses or lakes, or a hybrid of all of those, a modern BI tool should be able to connect to all of that data quickly and securely.
• Look for systems that can work with your existing infrastructure tools and even make your existing data sources more complete.
• Modern BI tools do not require you to build your own APIs from scratch, and make it easy for non-IT users to connect data with pre-built cloud connectors and flexible connection options for all skill levels.
• Since many organizations use multiple cloud data warehouses, it’s critical to find a tool that allows IT leaders to optimize all their investments and make data available to business users from any and all cloud data warehouses. Multi-cloud data fabrics enable a native integration of BI tools into any cloud data warehouse to take advantage of existing infrastructure and deliver data to business users in a secure and governed way.
Look for tools that can join data from data warehouses, lakes, and other systems

A simplified architecture, advanced connector configurations, more elegant ETL orchestration tools, and sophisticated auditing processes will prove more valuable in the long-run in maximizing data architecture investments. The key is using a cloud integration solution that functions as a data fabric to weave together fragmented sources and can remain flexible when business users change the context of the data. Dark data can be brought into the light when contextualized in the way users value the information.

- Bring dark data to light through ETL data pipelines and orchestration tools. Look for more elegant ETL orchestration tools that can be used by both business users and IT users alike, with SQL capabilities for more advanced users and simple drag-and-drop interfaces for less experienced users.
- A modern BI platform allows you to build pipelines once and automatically apply your business logic to data updates. You can even get alerts when data transforms fail. All of these features combine to support a variety of use cases from the first mile of data transformation to the last mile of data product consumption for users.
- Data writeback capabilities, or bi-directional connectors, must be present to enable you to write cleansed and enriched data back to source systems.

Act before governance issues compound

- Cloud-based integration platforms with agile interfaces can compress development cycles by incorporating new data sources and users quickly and adding much needed governance and certification processes to the development of data products for business users. Implement cloud data integration solutions with core data governance systems that can help you manage governance across environments.
- Even highly unstructured data like social media posts, call centers, and other high volume and velocity unstructured documentation can be processed, transformed, and normalized under a single system. A connected architecture with pre-configured connectors enables this capability.
- Audit your data trail through tools that include a data lineage function so you can maintain control of your data while providing users the right kind of data access. With Data Lineage & Impact Analysis tools, you can easily see the downstream impact of changes to datasets on other datasets or areas of the system. You should also have a way to monitor your datasets and ensure they are up-to-date and running on time.
- Any modern BI platform should include content certification that allows you to set up review workflows for your datasets and visualizations in Domo so that everyone in your company knows which data has been certified.
Look for the ability to designate access to data down to the individual row level so you can easily filter data based on each user’s specific needs or security requirements.

Look for tools that allow you to provide access based on user attributes, such as department, region, or other identifiers.

Look for group management tools that allow you to easily manage large numbers of users and integrate with your SSO provider to make it easier to scale as your company grows.

**Weave data science into the data platform**

Data science workflows that are embedded into the platform deliver much more flexibility and resiliency—especially when data engineers prefer to work with tools they are used to. This includes ways to port scripts directly into the data pipeline, as well as using advanced and automated feature engineering to better understand what attributes drove outcomes.

**Use virtual systems**

Much faster than original ETL, virtualized systems allow the processing of millions of rows in minutes. Dataset views are completely virtual (no processing time) systems with MPP architecture, providing the ability to bring in raw transactions. Multiple processing units enable the speed and response times for generating complex queries on the fly.

**Insist on high performance systems**

Leverage systems with sub-second performance, massive parallel processing columnar architecture, and big data machine learning tools.
Conclusion

Data agility is about democratizing tools that once only lived in IT so data teams can go faster. Data agility is about unlocking all your data and getting it to work together—even if it wasn’t designed for this initially. It’s about connecting data from thousands of sources and multiple platforms, so that it no longer matters where the data comes from or where it lives. It’s about breaking down data silos, shining a light on dark data by connecting to any source to make data available if it wasn’t previously accessible, and combining disparate datasets to reveal new insights.

Data agility is the foundational layer of Modern BI. Getting to data value starts with integration and is critical for success. It’s about getting data within a single environment to start doing the work, but doing this quickly at cloud scale and record time. The need for speed is essential, along with the ability to scale and augment data in real time through sub-second query response times and big data requirements. Governance and management of data is more important than ever—these security measures have to be put in place in order to build trust across siloed organizations.

In part 3 of this e-book series, we will explore how the foundation of data agility supports the core of BI & Analytics through the modern BI practice of data literacy.

Sources
IDG Communications, MARKETPULSE RESEARCH: Data Integration and Exploration On Behalf of Domo, December 2020
Forrester, Balance The Surge in Self-Service BI with “Goldilocks” Governance That’s Just Right, September 2020