



FUNCTIONAL
ANALYSIS



EFFICIENCY



PERFORMANCE



USABILITY

BUILDING A MODERN DATA WAREHOUSE ON AZURE

I TABLE OF CONTENTS

INTRODUCTION

Business Transformation in a Data-Driven World.....2

CHAPTER 1

Key Trends Driving the Need for Change.....3

CHAPTER 2

How IT Can Lead Business Transformation with Data Warehousing in the Cloud.....8

CHAPTER 3

Guidelines to Building a Modern Data Warehouse in Azure..... 13

SELF-ASSESSMENT

Are You Ready for Data Warehousing and Analytics on Azure? 17

REFERENCES 18

INTRODUCTION

BUSINESS TRANSFORMATION IN A DATA-DRIVEN WORLD

Our world is driven by data. IDC predicts that by 2020 the digital universe will reach 44 zetta-bytes of data.¹ The challenge for business is how best to handle this explosion and exponential growth of data across both old and new data sources in order to uncover new insights and enable digital transformation for business growth and competitive advantage.

For over two decades, organizations have relied on a traditional data warehouse as a central hub and single version of the truth. This data management system consolidated relational data from multiple systems – such as ERP, CRM, and LOB systems – to provide reports and analysis.

The problem: **it is backward looking.** Reporting analyzes only what's already happened, leaving limited opportunity for business leaders to make forward-looking, data-driven decisions.

It is tantamount to driving a car forward while looking out the back window.

Data can be a vital asset, but only when properly managed. The volume and velocity at which new types of data are evolving require a dramatic change in our approach and effort.

This eBook will share the key trends driving the need to change, and how IT professionals can lead the way by building a modern data warehouse in the cloud that facilitates business transformation. What's more, we'll share our tips and best practices for building your data warehouse in Microsoft Azure, concluding with a self-assessment to help you determine your readiness.

“ Only 10% of organizations expect to have a highly profitable business unit specifically for productizing and commercializing data by 2020. (Gartner, 2016)



DATA WAREHOUSE DEFINED

“A data warehouse is a collection of data that is loaded from one or more data sources and is used to perform business intelligence tasks such as reporting and data analysis.”

Data warehouses are characterized by queries that scan larger numbers of rows, large ranges of data and may return relatively large results for the purposes of analysis and reporting.”

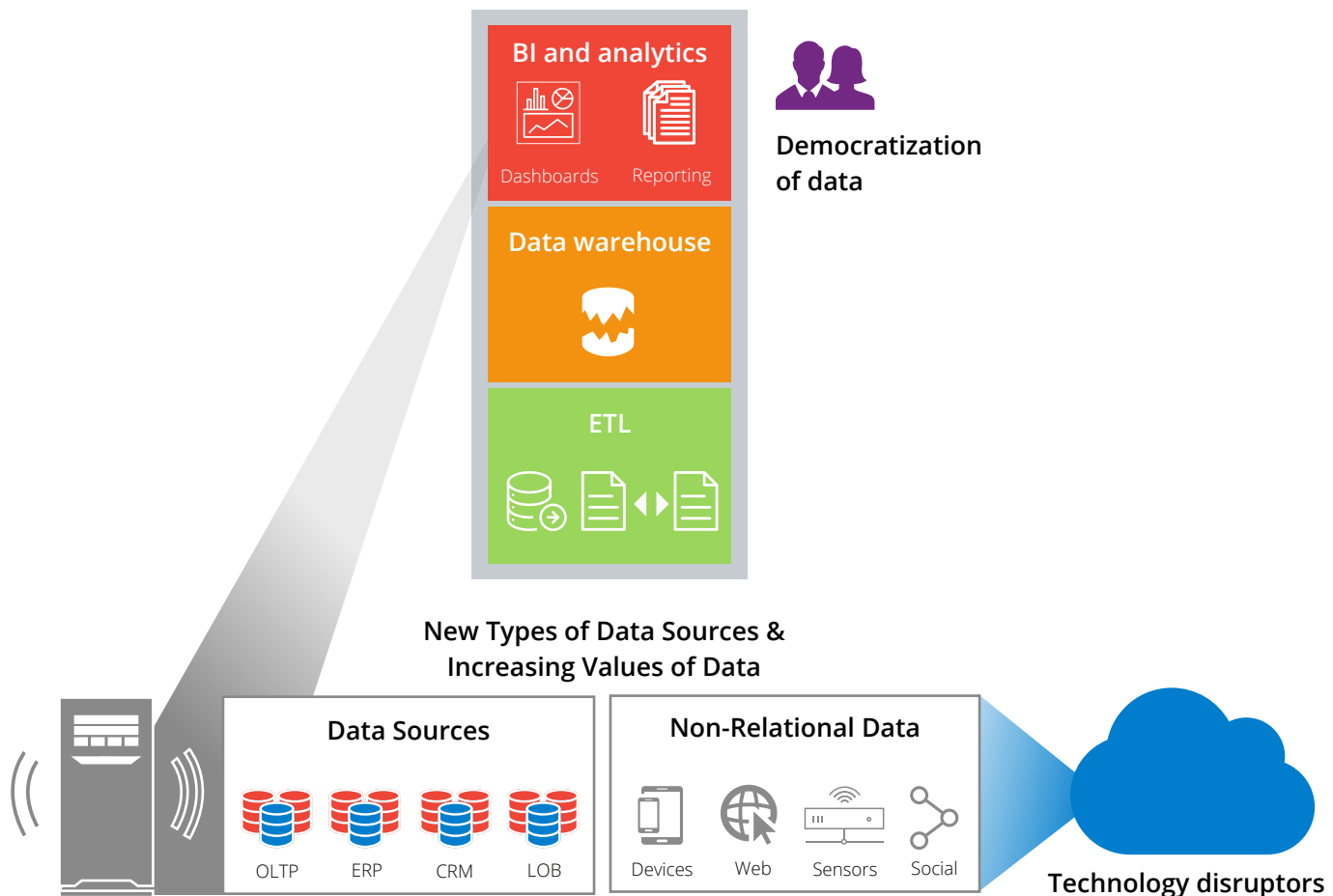
CHAPTER ONE

KEY TRENDS DRIVING THE NEED FOR CHANGE

Since the inception of data warehouses over two decades ago, the core business value of the traditional data warehouse has always been its ability to perform historical analysis from a complete source of data.

However, growing complexities are putting pressure on IT professionals to evaluate and evolve traditional data warehouses to meet the changing needs of the business.

KEY TRENDS DRIVING THE NEED FOR CHANGE:



Source: Microsoft, *The Microsoft Modern Data Warehouse*

1. New Types of Data Sources & Increasing Values of Data

Traditional solutions provide various levels and types of analyses of structured data, but were never designed to handle either the volumes or varieties of data that are currently available to businesses.

Reliance on symmetric multi-processing (SMP) technology makes adding more capacity expensive and time consuming. As businesses need to increase capacity, the process requires buying more powerful and larger hardware to mitigate the likelihood of the architecture experiencing performance issues.

With more than **85% of data volume coming from new, unstructured data sources**²

–such as mobile and social channels, IoT devices and sensors, web, video and other outside sources – it does not make economic sense to try and ETL (extract, transform and load) these non-relational data types into the traditional schema model.

Unstructured data is driving the world's data to double every two years,³ and as a result organizations need a data warehousing solution that can manage and combine relational and non-relational data types in order to drive advanced analytics and transformative insights.

2. The Democratization of Data

Data democratization means breaking down silos and efficiently connecting systems in order to provide access to data anywhere, anytime, and from any data source. This requires developing an IT platform that can support agile and flexible decision-making.

Every department is looking for ways to collect new data and put existing data to work. It's never been easier for a business (and individuals) to collect data from around the globe that they previously did not have access to. With the emergence of intuitive analytic and reporting tools, business units are enabled to analyze information and determine meaningful value through self-service.

IT can be a key enabler to democratization. IT departments have the opportunity to play a large role by building a platform that can deliver these transformative insights. It has become a business expectation. Leading digital transformation companies will strive to embrace a data-driven culture and use data as a strategic asset to drive intelligence.

x2 GROWTH

Unstructured data drives the world's data to double every two years.



3. Technology Disruptors

The Cloud

Cloud-first strategies have become a central part of many IT organizations. According to recent IDG research, 28% of organizations' total IT budgets will be dedicated to cloud computing in the next year.⁴

These cloud projects will be focused on IT modernization. Gartner noted that IT modernization is the top driver for public cloud adoption, followed by cost savings, innovation and agility.⁵

The movement toward cloud technologies means we now face a world of connected data, or data that is "born in the cloud". From videos and social feeds to cloud-born relational business data, from the movement of business applications to cloud-based platforms, the cloud is disrupting businesses and how IT should access, analyze, use and store data.

28% of organization's total IT budgets will be dedicated to cloud computing in the next year.⁶

Advanced Analytics

It's no longer enough to rely on historical analysis. Business leaders are looking for innovative ways to uncover new insights with actionable outcomes so they can remain competitive while identifying new business opportunities.

Advanced and predictive analytics are helping organizations determine what might happen, using a varied set of data sources. Advanced analytics and data science use the experimentation approach of exploring answers, which requires examination of data before it is curated in a schema. This allows data to drive insight in itself.⁷ Since the traditional data warehouse must work through a requirements-based model, where the schema and data collected is defined up front, this model is unable to deliver advanced analytics.



“ My first data warehousing project was in 2002. It took our team six months to deliver three to four dashboards across ten different subject areas. It was a very successful project during a time when the failure rate was high.

We had set the expectation for the CIO that this project would take six months. That wasn't a problem. Business users waited six months for the information they required. The traditional model worked well 20 years ago.

We are living in an era of digital transformation where democratization of data has shifted business expectations. Where it used to be ok to wait six months, that's no longer a viable option. Businesses want to see data and results immediately, at the click of button.

What's more, technology disrupters are providing more opportunities for IT to deliver flexible, agile, and self-service solutions so that users can derive value quickly, in a cost-effective manner.

Shifting from a traditional on-premises data warehouse to building a modern data warehouse in the cloud provides the agility, flexibility and elasticity businesses need to drive new insights from their data today.



Azhar Manzoor,
Cloud Solution Architect, Microsoft Canada

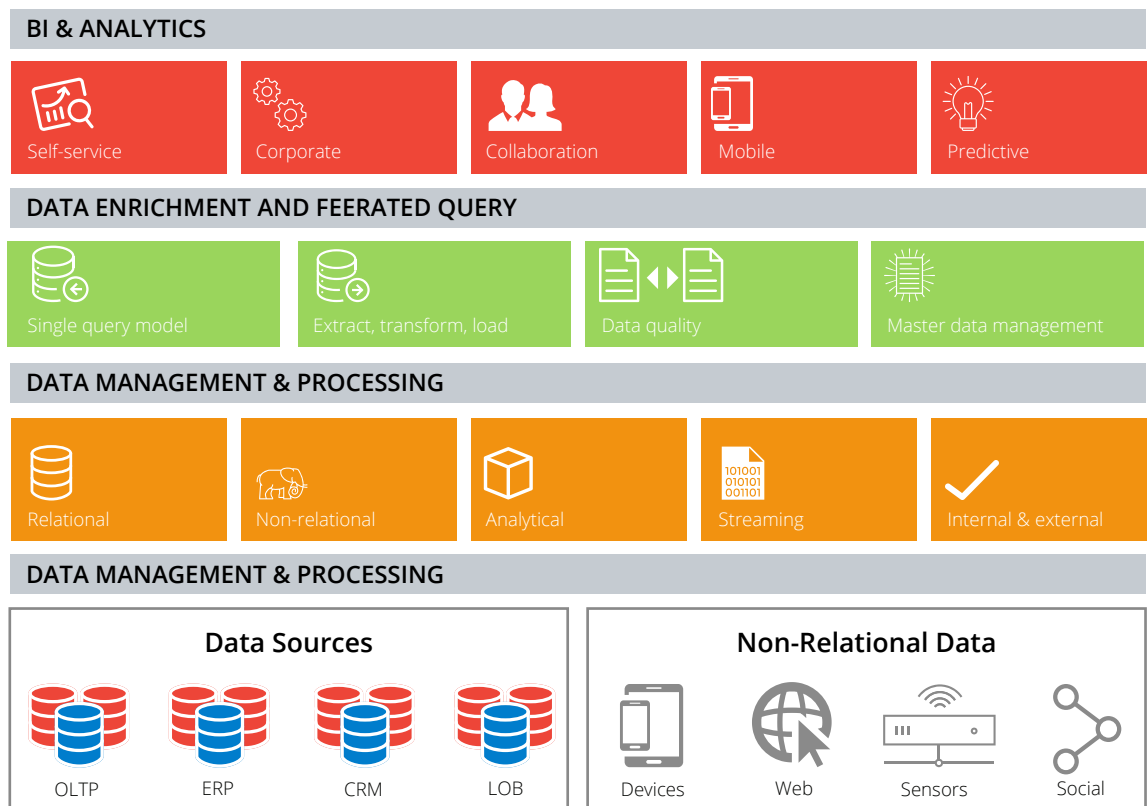


CHAPTER TWO

HOW I.T. CAN LEAD BUSINESS TRANSFORMATION WITH DATA WAREHOUSING IN THE CLOUD

A MODERN DATA WAREHOUSE

- Handles both relational and non-relational data in real time, providing an analytic engine for predictive analytics.
- Enriches data with ETL capabilities, or can prepare data through Big Data techniques. Credible and consistent data is supported through data quality and master data management services.
- Easily queries relational and non-relational data through a single federated query service.
- Supports a breadth of BI tools that allow business analysts to use what's already familiar, and allows business users to create and share analytics with teams.
- Supports data scientists while running experiments with data, predictive analytics modeling, and assisting in real-time decision-making.⁸



Source: Microsoft, *The Microsoft Modern Data Warehouse*

WHY THE CLOUD MAKES SENSE FOR DATA WAREHOUSING

ELASTICITY

Utilization remains a top concern for IT decision makers. Under traditional data warehouse solutions, the size of the system is based on the maximum load expected across the entire life of the project. In a data warehouse, a lot of data has to be moved, requiring additional compute power and storage. Yet, outside of peak periods for data loading, transformation and report generation, these systems are hardly utilized.

The cloud has unlocked computation and storage to make it much more elastic – by separating them. IT can rapidly provision and release resources to match workload requirements. As a result, the user only pays for what they use.





43% of organizations have the largest plans to move data analytics and data storage to the cloud.







- IDG



HOW IT CAN LEAD TRANSFORMATION - BUSINESS DRIVERS FOR INTELLIGENT ENTERPRISES:

During a webinar series on data warehousing and Big Data, Prem Prakash, Sr. Product Marketing Manager Data Platform at Microsoft, discussed five business drivers and included examples of leading companies that used data to achieve business outcomes:

 Predictive Maintenance: Fix problems proactively before they start. Energy consumption lowered by 35% 	 Data-Driven Services: Explore new business opportunities. Net profits increased 240% 
--	---

 Remote Monitoring: Improve visibility to make more accurate predictions. \$25.5 in customer savings 	 Inventory Management: Get the right products to the right places. Inventory processing reduced by 50% 	 Personalization: Offer customers what they want, when they want it. Customer insights cut from 24 hours to seconds 
--	--	---

With only 10% of companies thinking about creating a highly profitable business unit specifically for productizing data, there is an opportunity for IT to lead transformation and prioritize a data-driven culture. As a result, IT can help all departments create value that improves operations, customer experience and business models.

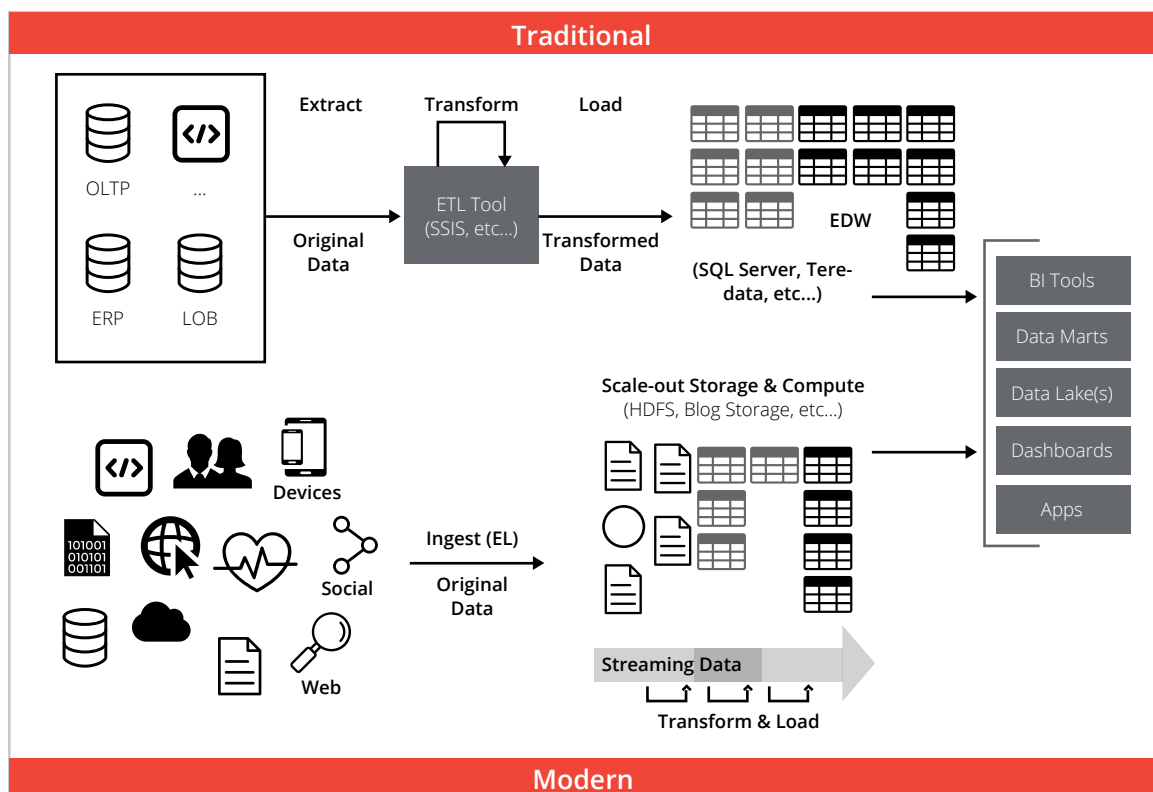
In the next chapter, we will explore how to build a modern data warehouse in Azure.

CHAPTER THREE

GUIDELINES TO BUILDING A MODERN DATA WAREHOUSE IN AZURE

The traditional data warehouse has served businesses well for many years, but its shortcoming is the time-to-value it delivers. A modern data warehouse meets a business's time-to-value requirements. IT can quickly add more analytics use cases – such as for reporting, advanced analytics or dashboards – as the business defines more requirements.

The main difference between the two models is that in a modern data warehouse all raw data is ingested first and stored in its raw form without a particular schema attached. Based on the analytics use cases, raw data is then transformed into separate data sets in order to fulfill the use cases.



Microsoft: Modern Data Warehousing with the Microsoft Analytics Platform System

HERE ARE 8 GUIDELINES TO HELP YOU BUILD A MODERN DATA WAREHOUSE IN AZURE:

1. INGEST DATA IN ITS RAW FORM INTO A DATA LAKE, SUCH AS AZURE DATA LAKE STORE OR AZURE BLOB STORAGE.

With increasingly diverse data varieties, the modern data warehouse requires a more robust storage mechanism that can store both relational and non-relational data. In **Azure Data Lake Store** you can store petabyte-size files and trillions of objects – 200 times larger than other cloud store objects. This provides flexible, unlimited storage designed for analytics and analysis.

3. USE AZURE HDINSIGHT CLUSTERS OR AZURE DATA LAKE ANALYTICS TO CREATE CURED ZONES BY TRANSFORMING RAW DATA AT SCALE.

HD Insight is a fully-managed Apache Hadoop offering that gives you optimized open-source analytic clusters for Spark, Hive, MapReduce, HBase, Storm, Kafka and Microsoft R Server. An IDC study showed that **HDInsight delivers 63% lower TCO** (total cost of ownership) than deploying Hadoop on-premises.¹⁰

Azure Data Lake Analytics is an on-demand analytics job service capable of running massively parallel data transformation and processing programs in U-SQL, R, Python and .NET over petabytes of data.

2. DEFINE RAW AND CURED ZONES WITHIN YOUR DATA LAKE STORE.

Eliminating the requirement to transform data before it is loaded allows you to store as many fields of data as available, and use what you need at a given time. With your data stored in its raw form in the data lake store, you can begin to transform it into cured zones.

A cured zone is where you will transform data for the specific uses case the business wants. Keep a copy of the raw data, as well as the cured data.

4. APPLY THE APPROPRIATE SCHEMA ON-READ FOR AGILE ANALYTICS AND DATA SCIENCE WORKLOADS.

Rather than using schema on-write like in the traditional data warehouse, you will apply the right schema on-read, at the right time, dependent on your different workloads. For example, a data scientist doing advanced predictive analytics will require a different schema to work with in comparison to the schema serving your dashboards. Therefore, the schema is very flexible.

“ **Best-in-Class cloud-BI users saw a 21% year-over-year increase in operating profit.**
(Gartner, 2016)

5. ADD MACHINE-LEARNING AND ANALYTICS IF NECESSARY USING HDINSIGHT / R SERVER CLUSTERS OR AZURE MACHINE LEARNING.

This is an optional component to add onto your data store to do advanced analytics. You're no longer bound to one database. Now, you can get data from your cured zone and use an additional **HD Insight** or **R server** cluster or **Azure Machine Learning** to do advanced analytics on scale.

7. FOR HIGH CONCURRENCY, EXTRACT AGGREGATES INTO AZURE SQL DATABASE AND ADD AZURE ANALYSIS SERVICES TO PROVIDE BUSINESS-FRIENDLY SEMANTIC LAYER.

It may be necessary to provide a more business-friendly and concurrent access to data. **Azure Analysis Services** delivers scalable, flexible and manageable BI semantic modeling capabilities.

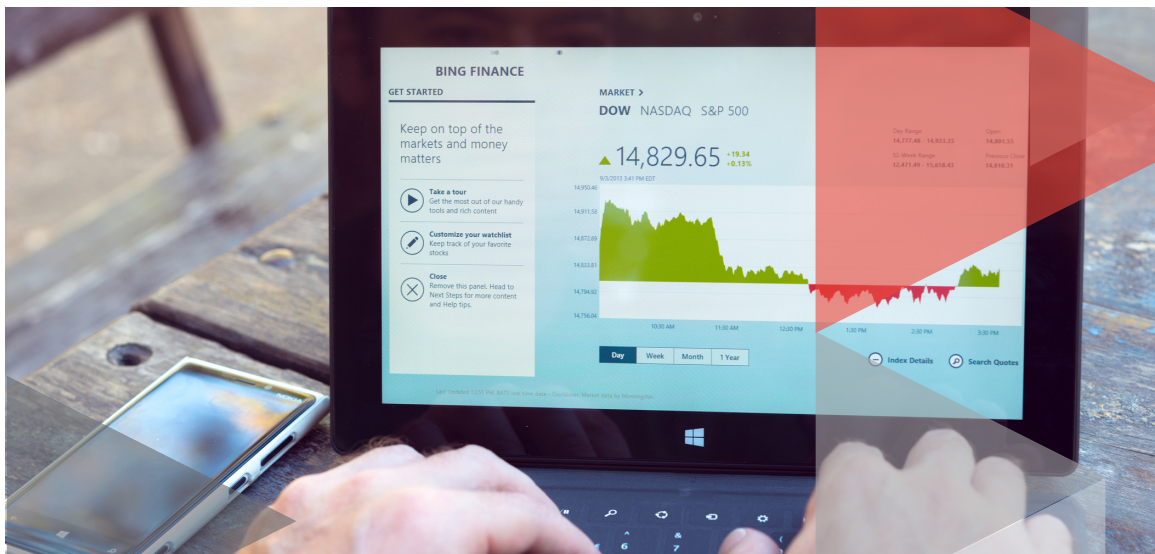
6. USE AZURE SQL DATA WAREHOUSE TO PROCESS LARGE AMOUNTS OF STRUCTURED DATA IN A DIMENSIONAL STAR SCHEMA FORMAT.

If you use structured data most of the time, make your data (stored in the data lake store) more available by setting up a **SQL Data Warehouse**. As a cloud service, it provides on-demand flexibility to give a relational schema on top of the data.

8. USE SELF-SERVICE VISUALIZATION TOOLS, SUCH AS POWER BI, TO VISUALIZE DATA AND OPERATIONALIZE THE DASHBOARDS.

Build interactive data visualizations with Power BI (or any other BI tool) to deliver actionable insights that support agile and flexible decision-making throughout the organization.

A modern data warehouse in Azure fulfills not only the need for historic reporting, but also meets all advanced analytics requirements in an agile and elastic way.



SELF-ASSESSMENT

ARE YOU READY FOR DATA WAREHOUSING AND ANALYTICS ON AZURE?

ASK YOURSELF THESE QUESTIONS:	YES	NO
1. Does our environment handle diverse data sources and a variety of subject areas?		
2. Can we handle excessive volumes of data (social, sensor, transactional, operational, analytical)?		
3. Can we support real-time or near real-time analytics to provide instantaneous insights?		
4. Can we support a variety of modern architectures to maximize scalability and performance?		
5. Have we leveraged cloud services for modernization, cost savings, innovation or agility?		
6. Has the organization applied data warehouse automated orchestration for improved agility, consistency and speed through the release life cycle?		
7. Is our organization running a bimodal business intelligence environment that supplements traditional reports and dashboards with ad-hoc reporting, data exploration and analysis and data science?		
8. If we asked our primary business sponsors, would they know where the data catalog is located to document business terminology?		
9. Do we use agile deployment models?		
10. Have we clearly defined how we certify enterprise BI and analytical environments?		

If you've answered "no" to any of these questions, it might be time for you to think about partnering with an expert to help you build a modern data warehouse in the cloud.

Optimus can be a seamless extension of your team throughout the discovery, agile delivery, transition and support process.

Get in touch today to see how Optimus can assist you with all your Azure and data analytics needs.

I REFERENCES

- ¹ IDC, The Digital Universe of Opportunities: Rich Data & the Increasing Value of the internet of Things, from <https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf>
- ² Microsoft, The Microsoft Modern Data Warehouse from download.microsoft.com/.../Microsoft_Modern_Data_Warehouse_white_paper.pdf
- ³ Microsoft, Webinar: Data Warehousing & Big Analytics from <https://info.microsoft.com/data-warehousing-and-big-data-analytics.html>
- ⁴ IDG, 2016 Enterprise Cloud Computing Survey, from https://www.idgenterprise.com/resource/research/2016-idg-enterprise-cloud-computing-survey/?utm_campaign=Cloud%20Computing%20Survey%202016&utm_medium=Press%20Release&utm_source=Press%20Release
- ⁵ Gartner, Gartner says worldwide public cloud services market to grow 17 percent in 2016 from <http://www.gartner.com/newsroom/id/3443517>
- ⁶ IDG, 2016 Enterprise Cloud Computing Survey, from <https://www.idgenterprise.com/resource/research/2016-idg-enterprise-cloud-computing-survey/>
- ⁷ Gartner, Big Data Business Benefits are Hampered by 'Culture Clash', from <https://www.gartner.com/doc/2588415/big-data-business-benefits-hampered>
- ⁸ Microsoft, The Microsoft Modern Data Warehouse from download.microsoft.com/.../Microsoft_Modern_Data_Warehouse_white_paper.pdf
- ⁹ Microsoft, Webinar: Data Warehousing & Big Analytics from <https://info.microsoft.com/data-warehousing-and-big-data-analytics.html>
- ¹⁰ IDC, The Business Value and TCO Advantage of Apache Hadoop in the Cloud with Microsoft Azure HDInsight, from <https://azure.microsoft.com/en-us/resources/business-value-and-tco-advantage-hadoop-with-hdinsight/>