



WWW.ENGAGINGDATA.CO.UK

AN INTRODUCTION TO DATA WAREHOUSING

A Data Warehouse has become a critical component of modern business intelligence, providing a centralised repository of all your data from different sources. It is simpler and easier for Analysts and Business Users to access and analyse all the information needed to make informed decisions.

Data hugely benefits your organisation, providing it is not siloed, stale or standalone. In this case, it can create or be part of the problem. Centralised Data will undoubtedly provide a solid foundation for data-driven decision-making and help address many issues your business faces.

Yet, despite its many benefits, a poorly designed or maintained Data Warehouse can cause serious problems for your business. It can become a hindrance rather than help, which we're sure, you don't want, and we certainly don't want for you!

So, is your Data Warehouse killing you? Or maybe you don't even have one.

Not to worry though, it happens to a lot of people. They don't know where to start or what they're doing when it comes to creating a Data Warehouse; they've decided they want to use data to grow their organisation effectively and become data-driven, and that's a great thing to want to do. You're not alone!

Data projects can be a very complex and challenging feat.

A Data Warehouse, in its many different forms, is a crucial tool for you to get your data projects started and for them to thrive. But you are probably wondering where to start, getting frustrated and finding it challenging to deal with. You downloaded this PDF, didn't you?

Here are the 10 Reasons Why Your Data Warehouse is Killing You.

1. YOUR DATA WAREHOUSE IS TOO SLOW.

One of the first questions you must ask yourself is "When do I need my data for our stakeholders?". Their answer will inevitably be "ASAP" or some silly time in the morning. You'll look to your developers, whose response will be, "It will take as long as it takes based on all the data you have requested".

Now, if we look at a slow data warehouse's impact on a business, we need to consider the consequences of not providing the warehouse on time.

A slow Data Warehouse is delaying decision-making. Slower data analysis can lead to missed opportunities or suboptimal decisions on outdated or incomplete data.

Being slow will also impact costs and efficiency. Waiting around for your data for any data team is frustrating and, as a result, will often lead to decreased productivity. Costs will likely increase, and the waiting around will inevitably lead to project overrunning.

A Data Warehouse could be running slow for many reasons, and in most cases, it's probably various issues as opposed to one quick fix. It could be down to data volumes or data quality, but it's just as likely that it's poor database design. This could include inefficient queries and inadequate indexing.



2. YOUR DATA WAREHOUSE IS TOO EXPENSIVE.

A Data Warehouse does require a certain level of investment; that's no secret.

However...

If your Data Warehouse is costing you an arm and a leg and is not delivering ROI, then something is fundamentally wrong!

Many different components make up the cost of a Data Warehouse: the storage platform, the transformation pipeline and, notably, the individuals who make it all work!

When these components are neglected, they will damage efficiency and drive up the costs of running your Data Warehouse.



3. YOUR DATA WAREHOUSE IS DIFFICULT TO MAINTAIN.

Your Data Warehouse should not be difficult to maintain!

Our golden rule is that if someone calls you at 2am, interrupts your beauty sleep, and asks you to fix a problem, could you do it?

The answer should be YES. If you can't, then you have a serious issue!

Maintaining a good service ensures that end users develop trust in the data, the time that the data is made available, increased adoption, and supports a data-driven decision-making culture. If it's really good, it can handle data quality too!



4. YOUR DATA WAREHOUSE IS UNDOCUMENTED.

What documentation? Where is it? Oh wait, we don't have documentation...

An undocumented Data Warehouse is of no use to anyone!

Without proper documentation, there is no understanding, and it will be very challenging for users to recognise the structure of your Data Warehouse, how it's organised and how to use the system effectively.

Imagine navigating a city without a map or road signs, and it's dark!

You're going to get lost!



5. YOUR DATA WAREHOUSE IS OUTDATED OR HAVING DATA REFRESH PROBLEMS.

Back in my day...

Data platform architecture designed in the 1990s is simply not ready to solve business problems in 2023.

With the ever-growing volume, velocity & variety of data, traditional Data Warehouse solutions are becoming outdated.

A legacy Data Warehouse requires a disproportionate degree of management as you spend lots of time performance tuning, monitoring and resource provisioning.

Sound familiar???

An old, sorry, legacy Data Warehouse can significantly hinder an organisation's ability to use and analyse data effectively.



6. YOUR DATA WAREHOUSE IS NOT SCALABLE OR FUTURE PROOFED.

Watch your development team run for the hills, when you mention Data Archiving!

It's a developer's worst nightmare, and for a good reason!

Is your Data Warehouse set up to handle increasing amounts of data or a growing number of users without experiencing performance issues?

If not, it's only a matter of time before you face a problem.

For any business anticipating growth (so everyone), a fully scalable Data Warehouse solution is a no-brainer. You can future proof your business with a scalable data warehouse that can support new data sources, incorporate new data types and implement new analytics and reporting tools.

We can't think of any reason not to invest in a more Scalable solution.



7. YOUR DATA WAREHOUSE IS NOT SECURE OR FLEXIBLE.

We regularly see that secure data means flexibility goes out the window.

Why is this happening?

This should not be the case if you have the right process, security measures and tooling.

Your Data Warehouse needs to be secure; otherwise, you will face problems with breaches, unauthorised access, data integrity and compliance.

Likewise

It also needs to be flexible. Otherwise, you will face challenges adapting to changing business needs, data silos and an inability to integrate with other systems.

A modern Data Warehouse needs to be both. You don't have to sacrifice one for the other.

So, don't!



8. YOUR DATA WAREHOUSE IS LOCKED INTO TECH STACK.

It happens! You've invested significant time, money and energy into a particular set of technologies, programming languages or tools.

It was cost-effective and easy as it provided a stable foundation.

But now you are committed to a tech stack and cannot easily switch to another technology without having to unpick all the work you've done and invest substantially more than you wanted.

Being stuck in a 'Tech Stack' can limit innovation and flexibility and make it harder to keep up with new trends and technologies.

If transitioning away from a stack is part of your roadmap, then be prepared for a challenging process, but with careful planning and execution, it should be worth it.

Always have an exit strategy!



9. YOUR DATA WAREHOUSE IS NOT SUITABLE FOR ALL DATA TYPES.

"Hey, we have some social media data that would enrich the Data Warehouse, and it would be great to have that data in our marketing dashboards," said the Marketing Manager.

Data Warehouses are generally optimised for structured data types as they can be stored in tables and columns and easily queried and analysed.

Initially, that'll probably do, but over time you start to get requirements for unstructured data, and if you haven't planned up front for handling unstructured data, it will be challenging to integrate.



10. YOUR DATA WAREHOUSE DOESN'T HAVE ALL YOUR SOURCE DATA.

Measure what matters & only store the data that delivers the facts.

If you're not utilising all your relevant data, you're not delivering proper insight.

You don't need all flat files or tables from every system loaded into your data warehouse.

Just bring in the relevant data to build a full picture! Not having the complete picture increases the risk of missing a potential issue or making the wrong critical business decision.



11. YOUR DATA WAREHOUSE IS NONEXISTENT!

(WE'VE INCLUDED AN EXTRA REASON SINCE YOU'VE MADE IT THIS FAR!)

WhAt Is A dAtA wArEhOuSe???

If you don't have a Data Warehouse, hopefully, this document will help you navigate the issues with creating a Data Warehouse for your organisation.

You can avoid falling into the pitfalls that have befallen many before you.

You're welcome.



BUT... WHAT DOES GOOD LOOK LIKE?

Stop letting your Data Warehouse kill you, this shouldn't be happening!

A great Data Warehouse should be designed to support the analytical needs of your organisation and provide high-quality, reliable data which enables you to make informed decisions.

A great Data Warehouse should have the following characteristics:

SCALABILITY Your Data Warehouse should be scalable to handle increasing amounts of data as your business develops and grows.	DATA QUALITY It is necessary that the quality of your data is superior. Eliminating data inconsistencies and errors. If there's bad data going in, there'll be bad results coming out!	INTEGRATION Your Data Warehouse should integrate data from various sources into a single repository, creating a single source of truth.
PERFORMANCE Fast performance is a necessity; it allows you to get results quickly and efficiently.	SECURITY No one wants a data breach - your Data Warehouse's security must be robust and protected.	FLEXIBILITY Adjusting to your changing business requirements and analytical needs - flexibility and adaptability is a must for your Data Warehouse.

* This is not an exhaustive list. We could add a lot more information, but we wouldn't want to bore you too much!

this isn't you! you're not bored!



BACK TO BASICS...

A Data Warehouse is simply two things:

- A large storage area
- A couple of processes that perform tasks such as Extracting, Loading and Transforming - your data wrapped up something called an ETL (or ELT, depending on the tool you're using).

Easy right, so how does it get so complex?

Well, we could start talking about design methodologies like Data Vault and throw in the odd buzzword like Lakehouse or Contextualised Views and now it sounds complex! But ultimately it is the same thing.

We grab some data from one place. We convert some codes into descriptions, apply some source system logic, if needed, some business rules, and save it in a table for people to use!



DATA WAREHOUSING CHEAT SHEET

ETL/ELT PROCEDURE-

- \Box Select and Extract
- □ Pre-integration Quality and Integrity Check
- □ Integration: Transformations according to Mapping
- \square Post-integration Quality and Integrity Checks
- \Box Load Data into Warehouse

REQUIREMENTS -

- \Box Business Q&A with data values being supplied by BI
- \Box Terminology of these Q&As
- Create Business Matrix
- \Box Business Rules (collected by Analysts)
- Business Data Model: translate BO's to facts and their dimensions
- $\hfill\square$ Dummy Prototypes of BI Tools facing the Data Platform

DESIGN -

- \Box Candidate Data Sources identified by users
- \square Priority list of identified sources
- Extract sample data from sources
- \Box High-level data model supporting all the BI tools
- □ Transform & Load sample data into model schema

DEVELOPMENT -

- \sqcup Complete Warehouse design
- \Box Mapping source data to all Warehouse Data
- Comply mapping with Business Rules
- \Box Analyze source tables in databases & archived files
- \Box Contact owners of data about Quality & Integrity issues
- □ Implement mapping in ETL/ELT algorithms
- □ Coding and Testing ETL/ELT algorithms
- \square Test BI tools created by app development team

BEST PRACTICES -

- ☐ Identify Business Requirements: Before designing a data warehouse, it's important to understand the business requirements and the data that will be needed to support those requirements.
- Choose The Right Design Methodology: The choice of methodology will impact the performance and complexity of the data warehouse, so it's important to choose the right design methodology for the specific use case.
- □ Implement Data Governance: Data governance ensures that data is accurate, consistent, and compliant with regulations and policies.
- ☐ Monitor and Optimize Performance: Monitoring the performance of the data warehouse and optimizing queries and indexes can improve performance and reduce costs.
- □ Continuously Improve: As the business requirements change and new data sources are added, it's important to continuously improve the data warehouse to ensure it remains relevant and effective.



WWW.ENGAGINGDATA.CO.UK

WHY ENGAGINGDATA?

Our approach is simple.

Focusing on providing efficient and effective solutions for your specific requirements, we aim to help you save time and effort by streamlining your data processes and reducing the amount of manual work and effort required.

We understand the importance of building flexible and scalable systems that can easily adapt to new data sources and your evolving business needs - therefore our goal is to enable you to:

- Spend less time focusing on processing data
- Easily build and maintain New Data Processes
- Confidently get your Data Projects started.

Providing you value by maximising the impact of your data resources and reducing unnecessary complexity, the Engaging Data approach enables you to do more with less effort, by automating your data processes.

"I've been involved with teams so preoccupied with fixing or debugging data pipelines that they can't build new or improved data assets.

I believe that data teams should do the complex work they are superior at and not waste time with tedious, repetitive tasks.

Spending a small amount of time upfront, you can avoid detrimental problems and set yourself and your team up for success.

Engaging Data has a wealth of experience helping companies build new data teams and supercharging productivity from existing ones."

- Simon Meacher, Managing Director, Engaging Data

Every client's needs and requirements are unique.

Working closely with you, **we're here for a good time not a long time**, delivering you measurable results and optimizing your data processes for **long-term success** in a **short amount of time**.

WEAR YOURSELF OUT WITH A



WAREHOUSE

 \int_{X}^{2}

BOOK A CALL NOW -

🖾 office@engagingdata.co.uk

🜭 (+44) 0204 5056 566

🛞 www.engagingdata.co.uk

in Engaging Data Limited

WWW.ENGAGINGDATA.CO.UK