Cloud Data Warehousing Playbook for SMBs



BE DELIGHTED WITH YOUR DATA



Contributors



Weelin Lim

Chief Product Officer, TrueCue

Weelin has worked in the data and analytics industry for the last 20 years. In that time, he has delivered numerous large and small-scale BI and data analytics platforms across a variety of industries, including healthcare, finance, media and energy.

Having delivered major solutions for Lloyds of London, Citigroup, EMI Music and the NHS, Weelin is the creator and Chief Product Officer of the TrueCue Platform, a new SaaS Cloud Data Warehouse Automation Platform for SMBs.



Richard Birtwistle

Senior Manager, TrueCue

With over 20 years of management consulting experience, Richard's strengths are in defining strategies to complex business problems, engaging the business in the solution and leading it through to implementation.

From delivering a £35m analytics business case to a FTSE100 company, to developing solutions to model capacity and future demand across acute care trusts, leading to identification of savings in excess of £100m per annum, Richard develops innovative solutions to enable companies to effectively analyse their revenue and costs.



Ramiro Medina

Senior Data Warehouse Consultant, TrueCue

in

in

in

Ramiro is an experienced BI professional, who developed his expertise at IBM and MicroStrategy before joining TrueCue.

Having designed data warehouses for several multinational companies, Ramiro is currently deployed as a Technical Architect in an ambitious, cross-organisational Public Sector engagement. Equipped with an M.Sc. in Information Systems (Lund) and an MBA (London Business School), Ramiro leverages his interdisciplinary background to provide solutions that maximise business value without compromising on technical aspects.

Contents

- 1. A word from our Managing Director
- 2. Executive Summary
- 3. Introduction
- 4. Data Maturity: The Challenges for SMBs
- 5. The Case for Data Warehousing
 - 5.1. An Introduction to Data Warehousing
 - 5.2. The Data Warehouse Approach
 - 5.3. The Benefits of a Data Warehouse
 - 5.4. The Challenges in Building a Data Warehouse
- 6. The Cloud for Data Warehousing
- 7. How to Building a Cloud Data Warehouse
- 8. The TrueCue Platform: Leading the Way in Cloud Data Warehouse Automation

9. The Future of Cloud Data Warehousing for Small and Medium Businesses is Now

10. About TrueCue





Small and medium sized businesses (SMBs) are the backbone of the UK economy.

SMBs account for 50% of the total revenue generated by UK businesses, and employ 44% of the UK's workforce 1. In order to compete, develop and grow, there is an increasing need for them to be able to work with, and analyse, data. This capability can help to improve operations, deliver better products and services to customers, and understand the competitive landscape in which they operate.

Handled correctly, data can create a discernible source of competitive advantage.

However, the ability to work with large volumes of data has traditionally been the preserve of large enterprises, those with deep enough pockets to employ specialist teams and invest in the technology to store, manage and analyse it.

Fortunately, times are changing. There is a new breed of modern BI technology which is more accessible than ever before - both in terms of time and cost. Increasing use of AI and automation means that software can do much of the heavy lifting, rather than expensive teams of specialists. Smaller business can now manage data estates that would previously have required teams of experts.

For the first time, SMBs can gain access to enterprise-grade data solutions.

The transformative effect of this is not to be underestimated. Companies that are able to manage data and produce insightful analytics are proven to outperform those that don't. They are not only better at acquiring new customers, but better at retaining them. They grow faster. They are more profitable.

If you are an SMB looking for a way to give your business the edge, then this playbook is for you. It introduces the concept of modern data warehousing and explains how it serves as the foundation for good analytics. It covers options. It gives tips. It warns of the pitfalls.

Most importantly, it sets out clear actionable steps you can take to build this capability within your organisation. Faster and at lower cost than ever before.

I hope you find it useful.

Ganoo Dan (

James Don-Carolis Managing Director, TrueCue

1 UK Department for Business, Energy & Industrial Strategy, gov.uk.

A word from our Managing Director -James Don-Carolis

Executive Summary

SMBs have invested increasingly more time and money in their data and analytics initiatives over the past five years, but despite this Gartner recently reported 2 that **73% of mid-sized organisations are still labelled as having "low data and analytics maturity"** and are falling short of achieving valuable business outcomes from their investments.

This downfall stems from the organisation's reliance on IT departments to determine their data strategy, rather than fostering a culture of data literacy, and working together with the business to democratise data across the organisation.

To put it bluntly, if you want to drive maximum business value from your data and analytics strategies and initiatives, simply procuring analytics technologies and hiring consultants is not enough, especially in the current climate.

Organisations who are striving to achieve greater data maturity must establish and nurture an organisation-wide culture of data and analytics, and align their data strategy with their broader corporate strategy. This will be critical for organisations looking to remain competitive beyond COVID-19, with virtually all organisations facing some degree of industry and business model disruption — often created by digitalisation.

Data and analytics are essential enablers for responding to that disruption with transformation, whilst we also know that data-driven organisations are not only better at acquiring new customers, but better at retaining them. They grow faster. They are more profitable. To put it simply, data and analytics maturity is a key competitive differentiator, and is arguably now a necessity for every business to prioritise, as we become more reliant on the digital economy.

So, how do you become more mature?

Well, it is crucial to first understand the maturity journey which an organisation needs to embark upon, and the 5 stages pertaining to this journey. The first phase involves manual preparation of reports, driven by ad-hoc requests.

On average, data workers are spending seven hours per week manually updating formulas, pivot tables, cell and sheet references, and smaller organisations do not usually have the resources to facilitate this. To address this issue going forward, employers should introduce more efficient tools, both in data visualisation and preparation, as well as training staff to utilise these correctly.

The second phase consists of generating insights. Successful organisations will train departmental analysts with agile data visualisation tools to build dashboards and provide a consistent capability, as well as introducing best-practice guidelines for usage and standardisation via templates.

Automation, robustness and security of the processes are key to the third phase, with central data teams responsible for standardising approaches, while bringing data functions into a single location to help with inconsistent processes. Incorporating an effective data management strategy at this stage will ensure the automation of repetitive processes that can greatly improve

2 3 Questions That Midsize Enterprises Should Ask About Data and Analytics. Gartner, 2020



the efficiency and productivity for any organisation, relieving internal resources (both human and machine), and reducing manual labour and errors.

The fourth phase involves enterprise maturity. The enterprise data platform is designed to support the overall data and analytics strategy, provide a robust, governed, consistent and scalable capability, and service both standards and ad-hoc analytics as well as supporting advanced analytics such as machine learning.

Cloud strategy is also core to this stage, leveraging all the benefits and flexibility cloud resources offer. Automation at this stage can deliver many benefits such as time - and resource - savings, standardisation and flexibility.

And finally, in the fifth phase, corporate strategies should be analytics-led. Organisations that utilise both internal and external data in a systematic way to drive decision making across the business are proven to outperform those that don't. Creating a position of leadership that has the responsibility and ambition to prioritise data within their organisation's culture is key to achieving data maturity at this utmost level.

Alas, this journey is by no means easy or straight-forward.

The journey is all the more challenging for SMBs, who do not have the same funds or resources as their larger enterprise counterparts. So, it is understandable why they often struggle to derive maximum value from their data, and find it challenging to leverage their information assets, to become more data driven.

However, with advancements in accessible and affordable cloud technologies, such as cloud data warehouse automation, improving your data and analytics capability is now very much a reality for organisations with smaller budgets and less technical resources, and the transformative effect of this is not to be underestimated.

We know from our research that data-driven organisations are **23x more likely to acquire customers, 6x as likely to retain customers, and 19x as likely to be profitable as a result of building this capability** 3. Small and mid-sized organisations must therefore take the necessary steps outlined in this playbook, to ensure they achieve data maturity, which is not only crucial to remaining competitive, but for their survival, as we accelerate into the digital future.

Introduction

We are living in an increasingly data-driven age, with access to unprecedented amounts of information, and a digital economy that is showing no sign of slowing.

This accelerating digital landscape means that for many small and medium-size businesses, digital transformation is no longer a matter of choice — it is an existential necessity and, as Gartner quite rightly asserts ⁴, it is impossible to be a digital business without being a genuinely data-driven organisation.

Before COVID-19 struck earlier this year, a study by the SMB Group, found that 48% of small and medium businesses were planning to engage in activities to help them adapt and transform their organisations for a digital future, with 36% already implementing their digital transformation plans ⁵.

However, since the crisis began the race to innovate has inevitably intensified, with companies having to establish technical infrastructure in haste, as they frantically try to transform their business models, engender trust with their customers and elevate their digital presence.

Indeed, a company's ability to compete beyond COVID-19 will increasingly depend on its ability to make faster, more informed decisions.

This is where data can be the fundamental difference between staying afloat and sinking.

Disruptive business conditions resulting from the recent pandemic have forced many organisations to cease normal office operations and direct their employees to work from home. This shift in behaviour has made data even more critical. Keeping both their customers and employees informed, connected and enabled to continue communicating and collaborating is a top priority for business leaders.

⁴ How to Craft a Modern, Actionable Data and Analytics Strategy That Delivers Business Outcomes. Gartner, ⁵ SMB Digital Transformation: Why, When, Where and How? SMB Group, 2017

"Only 32% of companies realise tangible and measurable value from data." Accenture



'||;

Achieving this, however, requires consistent and reliable access to the flow of data in and across the organisation — a challenge smaller businesses have been struggling with for many years.

And, despite Gartner recently reporting that SMBs have invested increasingly more time and money in their data and analytics initiatives over the past five years, these organisations are still falling short of achieving business outcomes for their investments ⁶.

In fact, according to Gartner's IT Score for Data & Analytics tool, 73% of midsize enterprises are classified as having "low data and analytics maturity". Whilst a meagre 9% of midsize organisations from the same report stated that they were "very effective" at managing information assets strategically, with a further 21% stating that they are "very ineffective" ⁷.

WHY IS THIS?

Well, an organisation's data and analytics initiative will be meaningless unless they articulate who will use the information to make decisions, what business outcomes will result, and how those outcomes will advance organisational goals.

Despite this, only 30 percent of organisations align their data and analytics strategy with their broader corporate strategy 8. This has resulted in a resounding 44% of data

B Global AI Survey: AI proves its worth, but few scale impact Global AI Survey: AI proves its worth, but few scale impact. McKinsey, 2019 workers' time being wasted every week, having spent almost half their week searching for, preparing and integrating data 9. This is invaluable time which a smaller organisation simply cannot afford to waste, with their already limited data and analytics resources. It can also create a culture of uncertainty around data accuracy and make data analysis even more challenging. In addition, all this effort is spent on preparing and analysing only about 1/3 of their organisation's data, with between 60%-73% of the entirety of an organisation's data never being analysed 10.

Furthermore, as we all know, with more data comes more complex data management, to be able to effectively handle, utilise and realise the true value of that data to the business.

The reality is that many SMBs have never appropriately planned to manage the volume or variety of data that they acquire, nor do they have the capabilities to do so from both a technology and skills standpoint. This inevitably leads to the creation of data silos. These silos are essentially collections of data that are held under control of a single person or department, and are inaccessible by others within an organisation.

With multiple departments working on different models and software solutions, data silos begin to rapidly grow as the amount of data within departments increases in isolation.

⁶ Midsize Enterprises Need to Get More Value From Data and Analytics Investments

 $^{^7}$ 3 Questions That Midsize Enterprises Should Ask About Data and Analytics. Gartner, 2020

⁹ State of Data Science and Analytics Report. IDC, 2019

 $_{\rm 10}$ Closing the Data-Value Gap - How to become data-driven and pivot to the new. Accenture, 2019

This is where data management becomes a pivotal prerequisite for achieving transformative levels of maturity for an organisation's data and analytics capabilities.

Data management has always been a linchpin in the success of any mission-critical digital and IT led initiative, with data management technologies typically used to understand "how data is retained and controlled" and a more recent shift towards "how data is used and accessed".

Of course, larger organisations have long had the privilege of leveraging such technologies, in particular data warehouses, which play a crucial role in how their data is organised, made available, integrated, shared and governed across the business.

However, with the advancement of cloud technologies such as cloud data warehouse automation platforms, achieving more measurable business value from data and analytics has become viable — something which has traditionally only been available to the few, larger enterprises, is now available to the many organisations who are still struggling to extract valuable insight from their data.

We are living in an era in which humans and machines work collaboratively across the flow of data within the organisation — it is the era of harnessing human creativity augmented by machines. So, there has never been a better opportunity for SMBs to capitalise on this, and cost-effectively modernise their data architecture to have the flexibility to handle current and future use cases, which deliver high business value. For many SMBs in the UK, COVID-19 has understandably caused a crisis that has affected both their lives and livelihoods. But amidst the crisis many business owners have demonstrated great resilience, digging deep to ensure the survival of their business, and, if anything, the pandemic has shown that rapid change is both possible and pivotal for business survival.

This is one of the reasons we have written this playbook; to provide business leaders with an edge on how to leverage advancements in data warehouse technologies, as a foundational strategic asset for accelerating their organisation's data and analytics maturity, and supporting their digital transformation initiatives.

The time has come for SMBs to acknowledge that the future will be quite different from the past, and to build on both the new and pragmatic ways of working with their data.

Ways which have the potential not only help them to survive but to also thrive in whatever world awaits us in the new normal.

"Data-driven organisations are 23 x more likely to acquire customers, 6 x as likely to retain customers, and 19 x as likely to be profitable as a result."

McKinsey



Data Maturity: The Challenges for SMBs

As Robert Grossman quite rightly alluded to in his "Framework for evaluating the analytic maturity of an organization" 11, it is rare in today's world, for an organisation to develop software that is critical to its products, services or operations without a software methodology being used; on the other hand, it is relatively common for an organisation to build analytic models that are critical to its products, services or operations without using any analytic methodology.

This is one of the fundamental reasons why SMBs in particular are struggling to create a modern, actionable data and analytics strategy, that delivers the business outcomes and value that they crave and need.

Regardless of their size, most businesses thrive on data and the analysis of it, yet many of these same organisations are hampered by treating data and analytics as supportive and secondary to their business initiatives.

This results in several key challenges, which unless addressed at a foundational level, will continue to restrict an organisation from accelerating their data maturity.

"By 2023, 75% of all databases will be on a cloud platform." Gartner



11 A framework for evaluating the analytic maturity of an organization. International Journal of Information Management,

SO, WHAT ARE THE CHALLENGES HINDERING DATA MATURITY?





A lot of these challenges cannot be addressed overnight, nor can they be resolved with investment in technology alone. Instead, they are alleviated as an organisation advances along their Data Maturity Journey.

We consider there to be 5 stages to an organisation's Data Maturity Journey, and whilst different organisations will find their own pace as they progress along this, there are ways to accelerate progress.

Before we get to discussing that, we need to understand what these 5 stages are, where SMBs usually find themselves on this journey, and more importantly where they should be aiming to arrive at.



Stage 1: The Static Reporting Phase

Situation:

This phase involves a lot of manual preparation of reports, driven by ad-hoc requests. Report preparation is mostly reactive to requests, rather than proactive to requirements, and the tool of choice is usually Excel. All organisations would have started here, and even organisations that have progressed in general from this stage may have departments and teams still working like this.

Challenges:

Some organisations find it difficult to make the initial push out of the static reporting phase, because the constant workload and requests makes any planning time to progress quite difficult.

Additionally, since most of the work here is done manually by individuals, they hold key information about the processes they follow. These processes aren't often documented well, if at all, leading to key-person dependencies and a high-risk of single points of failure, as well as inconsistent and unchecked results. If the person responsible for running processes is unavailable, there is difficulty in generating the data required for reporting and different approaches are taken. This is both extremely time-consuming and introduces inconsistencies to the process which could lead to different results.

Progress:

One of the biggest challenge posed at this stage lies with the tools being used, which typically demand heavy manual effort.

Obviously when using Excel, the likes of Excel macros can be created to help with some of the manual and repetitive effort, but you will still have the same issue where these are only understood and maintainable by the Excel macro builder. The quickest way to address this would be to introduce more efficient tools, both in data visualisation and in data prep as well as training your staff correctly on these.

Stage 2: Generating Insights

Situation:

In this stage, the data preparation and analytics effort is often departmental-based, with analysts within those departments preparing reports and answering ad-hoc requests. Quite often, these are still Excel-led efforts, but it's common for some teams to have started using BI tools to deliver some of the reporting and even to help with some basic data prep processes.

There will be sets of standard, but often static reports that now have a process behind their preparation that is followed regularly. Some of these processes are now coded into Excel macros to introduce automation, whilst there will also be some formation of reporting data sets existing in data silos and will usually be department and even report/dashboard specific.

"Data workers are spending 7 hours per week on average manually updating formulas, pivot tables, cell and sheet references."

IDC

'||;

During this stage, there will not often be an overall organisation-wide data strategy, and there might not be a designated data leader within the organisation.

Challenges:

The second stage of the journey, typically consists of pockets of good practice and different approaches within departmental teams, but there will still be key-man dependencies and labour-intensive effort which significantly hinders the organisation's ability to be agile. Inconsistencies and unchecked results will still exist, as will information silos.

In this stage, departments will use data visualisation tools that allow analysts to build dashboards to answer more questions, more easily, reducing the effort required to answer standard questions. This stage would also introduce some self-service ability for people to answer those repeating or similar questions.

Often, progress at this stage is different within teams, with some teams still stuck in stage 1 whilst others will be developing in stage 2. The teams in stage 2 will be developing skills in more capable and modern BI tools like Power BI and Tableau and will also be exploring the data prep capabilities to help them deliver data in structures more ideal for consumption by those BI tools.

In addition, the more advanced teams will be exploring and using data preparation tools which their analysts can use to start bringing consistency to their data prep approaches, but this knowledge will still reside in key individuals.

Progress:

The best way to make progress to the next stage is to train your departmental analysts with strong, agile tools that provide a consistent capability as well as introducing best-practice guidelines for usage and standardisation via templates and approaches using those tools.

Introducing the automation capabilities of data prep tools as well as the modelling capabilities of visualisation tools like Tableau and Power BI will greatly improve the teams' data understanding and maturity, whilst reducing manual labour and errors.



Stage 3: Repeatable Analytics.

Situation:

In this stage, analytics is still usually departmental-based, but organisations will be using data prep tools to build pipelines to prepare and shape data correctly for reporting. Within departments, there will be concentrations of expertise and knowledge required to build and manage these processes.

Organisation-wide analytics will often then be collated and compiled by a central analytics team, whilst there is also likely to be an organisation-wide data and analytics strategy, and a more capable tool-set to support driving data analysis. Usually, these are agile, self-service tools that can be used by analysts within departments.

There will also be dashboarding standards and approaches, to ensure end-reporting looks and feels consistent.

Typical tools being used at this stage will be BI and dashboarding tools like Power BI and Tableau, as well as analyst-led data prep tools such as Alteryx, Trifacta and Tableau Prep.

You can also expect the introduction of automation and agile tools, but inconsistent practices and calculations across teams will still occur.

By now, there will be much more analytics demand which can only be supported by the right tools and skills within the teams.

Challenges:

Data and report processing responsibilities still lie within departments and individuals, which may result in inconsistent approaches, but there should be a standard, capable toolkit that analysts use which can effectively support these demands.

Central data teams with the responsibility for standardising approaches and processes will be formed, bringing this data function into a single location, to help with the problems of data silos and inconsistent approaches.

Progress:

Automation, robustness and security now become more of a focus as the organisations mature and think about the enterprise-wide needs of governance and scale.

At this point, organisations review their current tool-sets and approaches to consider whether they still are fit for the business' needs. This is where a lot of the agile tools which were originally used, show deficiencies in automation, scalability and capability within a maturing data platform strategy. Agile tools that were accelerators at department level to enable analyst teams to work quickly and efficiently may not be the ideal tool for a central data and analytics team looking to bring that maturity to the whole data provisioning process.

Having a data leader to define a data strategy, with the required backing from the leadership team is now becoming vital to align the outcomes with the wider corporate strategy, and secure the level of investment required to effectively execute.



Stage 4: Enterprise maturity

Situation:

At this stage, you will have an organisation-wide, common analytics infrastructure which is consistent, automated and integrated, as well as department-based analytics capability where needed.

The enterprise data platform will be designed to service both standard and ad-hoc/self-service analytics, with the underlying strategy being to ensure standardisation and consistency, reduce effort and repetition and prevent cottage-industries of departmental data prep and data silos. It will also support advanced analytics such as machine learning.

There will be an organisation-wide data strategy and also a Data Strategy Leader in the leadership team.

Challenges:

Stage 4 brings Enterprise quality but at a cost, with potentially a much more technically complex data platform, heavy development and maintenance effort, potentially expensive hardware, software and teams.

This stage is often only accessible to larger organisations that can afford it. However, by this stage, data value is central to an organisation's strategy.

The biggest challenges here are around the technical capability needed to progress to this stage from Stage 3. Usually, if the skills and experience do not exist within the current staff, you would have to create a new team to develop this and to maintain the solutions and platforms afterwards.

However, data governance or legal reasons may mean that organisations in certain industries have requirements to achieve this level of compliance and security. GDPR, Freedom of Information and Information Governance are examples of regulations that need an enterprise-level approach to data.

Fitting this requirement into your cloud strategy is also core to delivering this, because the cloud now makes this aim much more accessible from a cost and capability perspective. In this stage, agile tools can still provide a valuable capability with ad-hoc and experimental tasks.

Progress:

Treating data as a valuable asset and defining the overarching data strategy looking at how best to extract additional value from the data, once the enterprise-level data capability is in place will be vital.

Understanding what AI, Machine Learning and other more-advanced data analytics capability can provide, as well as determining the foundations to deliver those capabilities will enable good progress.

Stage 5: Digital Transformation

Situation:

Organisations here use both internal and external data in a systematic way to drive decision making across the business, both top-down as well as bottom-up. This is where the business and analytics strategies are aligned and the analytics is used to select strategic initiatives, led by a Chief Data and Analytics Officer at the helm.

Data is treated as a business differentiator and data from all sources are made available for analytics to be performed on it. Advanced analytics, such as AI and Machine Learning will be used to drive real value.

Challenges:

The challenge to reaching this stage is often in the drive and commitment of the leadership to prioritise data within their organisation's strategic approach.

Progress:

Creating a position of leadership that has the responsibility, ambition and permission to deliver this is key.





An Introduction to Data Warehousing

Today's data is big, fast, complex and ever-changing.

Whether it's the questions being asked of the data, or the data required to support decision making, organisations expect the right data to be delivered to the right teams, within the right context, at the right time – sometimes, days and hours is no longer acceptable.

This heightened expectation and dependency means that whilst data is an invaluable asset, it can just as easily become a liability if not used correctly, in the right context and within the right time frame.

With all the noise around big data, unstructured data and data lakes, understanding where data warehouses fit into the overall data platform is vital for organisations to ensure they leverage this well-established and proven approach successfully.

As a foundational building block from which an organisation can develop their data and analytics capability, a data warehouse is primarily responsible for bringing disparate, structured data sources together into a business-focused data model suited for analytics.

The data warehouse is by no means a new methodology in the world of data analytics, having originally arrived on the scene in the early-1980's.

However, unlike many other data management innovations which have failed to satisfy the demands and requirements of the modern data-driven business, three decades later, data warehouses are only going from strength to strength, mainly driven by the fact that all businesses have structured data that needs reporting on which would benefit by bringing into a single, consistent repository for organisation-wide analytics.



The Data Warehouse Approach

Data warehousing follows a structured approach with the end goal of enhancing decision making across the business.

Adopting a four-step process, these well-established data management approaches take the data on a journey, from its raw state, into a centralised, trusted and organisation-wide analytical foundation.

1. Extracting data from silos

Raw data from your company's transactional systems, such as your CRM and your accounting suite, is extracted and landed into a single, centralised repository.

This step lays out the foundations for your data warehouse. Rather than being spread out across databases with disparate technologies, all data is now accessible in a single platform.

2. Cleansing, correcting and linking data

It is very likely that, when looking across your CRM and other transactional systems, you will find values that are invalid or inconsistent. This is particularly true if you are gathering data from legacy systems. These values should be flagged as erroneous and either removed or handled with care. Owners of the data sources should be notified of these errors, and a correction can be issued in the source.

In some cases, invalid values will be evident and able to be captured with simple validations; for example, emails that don't follow a valid pattern. In other instances, it will be necessary to link data across sources before determining whether a value is valid.

3. Creating a business-focused data model

3.1. Conforming the model

Building a data warehouse is not only a technological feat, but also an endeavour that will challenge stakeholders across your company to arrive at common data definitions.

Even simple questions such as "what is a client?" may not be straight forward to answer. The areas closer to operations may claim that if it is in the CRM system, it is a client. However, the areas concerned with management accounting may apply additional criteria; for example, filtering out records without at least one purchase in the last three years.

A shared terminology will make sure all stakeholders can derive value from your data warehouse.

In the previous example, it's likely the business will opt for the more comprehensive definition and consider every client in the CRM, but at the same time, flag records with at least one recent purchase to suit management accounting needs.

'||;

These definitions should feed into the data dictionary, the central document for all data terms.

With a shared terminology agreed, different versions of the reference data extracted from across the organisation will need to be consolidated. This may need to happen every time data is extracted.

For example, if data is being extracted from a CRM and an accounting solution, both systems will have their own client records. To reap the benefits of the data warehouse, a single, combined registry must be produced. In data warehousing methodology, the process of standardising this is technically dissimilar but conceptually equivalent reference data is known as "conforming" the data model.

3.2. Calculating KPIs

Management without KPIs is often equated to driving in the dark, and KPIs such as ARPU (Average Revenue Per User) can help set the direction of an entire company. However, it's likely that different areas across the business will have slightly different definitions of what a user is, and even of what revenue is. For example, differences in revenue recognition can stem from following different accounting standards, i.e. IFRS vs GAAP.

For a data warehouse to be successful, the definition of company-wide KPIs must be unequivocal and agreed on by the main stakeholders involved. With those definitions agreed, KPIs can be updated in a predictable way every time new data is extracted. As with other shared data definitions, the KPI definitions should be stored in the central data dictionary.

4. Loading governed data

Data warehousing is an ongoing process that helps the company provide the latest information to its decision makers. Hence, data needs to be constantly kept up to date.

Establishing when new data becomes available is key to align expectations across the company; normally, a daily refresh schedule will be agreed, but this is completely dependent on when data updates.

Besides establishing when, it's imperative to determine how data becomes available, and to whom.

A data warehouse will contain some of your company's most confidential information, likely spanning areas such as Finance and HR. Access to data should be tightly controlled, and ideally be role-based and integrated to your existing IT infrastructure.

Once set up, your data warehouse can act as a trusted data source for all types of analyses. Often, data warehouses are associated with visual, self-service BI reporting tools such as Power BI, or Tableau. The simplicity and business orientation of data warehouse models make them perfect to work with such tools. However, data warehouses can support and enhance all types of analyses, including ad-hoc queries and even machine learning algorithms.

The Benefits of a Data Warehouse

For over 30 years, the prime objective of data warehouses has been to manage large volumes of data, providing a central source of data that is well-understood, quality assured and easy to analyse.

A solid data warehouse implementation will exhibit the following characteristics and reap the associated benefits.



Governed and Secure

Access to data will be governed and secured to reduce data risk and to ensure proper usage. Data access will be logged for auditing purposes.



Automated and Timely

A data warehouse relies on repetitive processes ran at regular intervals; it must be able to supply the right information at the right time. These processes should be fully automated and not reliant on manual interaction.



Correct and Trusted

Due to the data validation logging and full audit or transforms, the data warehouse data can be proven to be correct. Any data quality issues are highlighted and feed back to data owners.

These processes ensure that the data warehouse is a trusted source of valid data for the organisation. Ensuring trust in the data is vital in ensuring the data warehouse delivers analytics value to the organisation.



Enterprise-wide

The data warehouse will provide access to data originating from across the organisation in a single, conformed and corrected location. It will act as a single "golden source" of data for analytics.



Centrally managed

A centrally managed data warehouse frees up analyst resources across departments and allows data policies to be implemented consistently across the company. Central responsibility prevents ambiguity of data ownership.





Consistent

The data warehouse will hold standardised definitions for reference data and KPIs, helping eliminate the noise created by inconsistent definitions and interpretations.

A
H

Expandable

A properly designed data warehouse will be able to expand to accept new data sources, without needing to re-engineer existing assets. This will allow the company to easily and progressively extend its analytics coverage.



Based on proven techniques

Data warehousing is a tried and tested solution for organisation-wide analytics. It is underpinned by mature, methodology-driven design and implementation techniques. Making use of these proven techniques will maximise the value delivered by the data warehouse.

"Companies struggle with building the right level of trust in their data. Data governance is hard and is often seen as boring by non-data folk. Learning how to introduce the foundations of trust-in-the-data-system requires letting the audience experience the data system themselves.

This is why cloud data warehouse automation platforms are so exciting. Using the power of AI to democratise data, they give people the trust and belief people need in their data, empowering them to achieve more."

> Phil Harvey Senior Cloud Solution Architect Data and Al, Microsoft

Challenges in Building a Data Warehouse

Due to the inherent complexity, designing, building and maintaining a data warehouse solution can be challenging. Data warehousing projects often exhibit most, if not all, of the challenges listed below, each of these which has the potential to undermine the value that your company can derive from a data warehouse.



Complex

Data warehousing projects can be complex and have many moving parts. The cross-organisational nature of these projects increases the value they deliver, but also their complexity. Discovery phases often drag on, eroding trust in the project and its benefits.



Technically demanding

A dedicated and experienced technical team is needed to develop a data warehouse, from data modellers and platform architects to developers and QA.



Inflexible

Making changes to the specifications once delivery is underway will require extra effort. Agile methodologies are often applied to surmount this challenge, but often increase overall development costs.

Moreover, in data warehousing, implementing agile is seldom straightforward, as it is sometimes difficult to divide the project into features small enough to fit into agile sprints.



Expensive

The costs of delivering a data warehouse can quickly pile up once development, hardware, software and support costs are included.

Data warehousing projects typically run for many months and sometimes longer than a year. The business often needs to wait until the final third of the project to start seeing results.





BAU Effort

Once the data warehouse is up and running, maintaining it in a business as usual state whilst introducing changes to the model can be very challenging; minimising disruptions requires careful planning and testing. Seemingly small changes can demand big efforts, which frustrates IT and business stakeholders alike. It is not uncommon for data warehousing teams to be flooded with change requests, which severely limits their ability to tackle new requirements and continue growing the data warehouse.



Storage and compute-intensive

Data warehouses have a comparatively high requirement for data storage compared to transactional data systems, given that they're designed to store historical data spanning many years. They also have compute-intensive data processing requirements; some operations, such as linking data across sources or aggregating historical data to calculate KPIs, are notably demanding.

The storage requirement will continue to grow over time and, as more data is added to the data warehouse, compute requirements will also increase.

Ensuring the platform is specified to meet these changing requirements can be challenging as it must be able to cope with the increasing workload and storage.



Increased processing times

As the data warehouse grows and evolves, the need for computing power quickly increases, often surpassing everyone's expectations. When different sources are linked, compared and combined, processing times grow quicker than data volume; linear growth in the amount of data can drive exponential growth in processing times, unless more resources are provisioned.

Depending on the data load window, processing power may need to be increased for certain data loads to ensure the data can be processed and made available to end-users within the required timescales.

The Cloud for Data Warehousing

Data warehousing has been re-thought and reborn in the cloud for the modern, data-driven organisation, and more importantly for the smaller sized organisations for whom this capability has previously been out of reach.

By using the cloud, there can be much less infrastructure and software maintenance and resources can be obtained within minutes. Data security and disaster recovery options can be built-in and all of these features can be managed remotely.

The flexibility and ease with which storage and compute can be increased means that organisations no longer need to purchase hardware based on future predicted usage, but rather, purchase the storage and compute they require when they need it, and increase it accordingly over time.

In addition, the ability for certain cloud services to be temporarily scaled up to deal with increased workloads and then scaled back down after the processing has completed is a pattern that is highly suited to data warehouse usage patterns which can be highly sporadic.

All of these reasons make the cloud a great fit for data warehousing but understanding the different options within the cloud platform you choose is vital to ensure you derive the most benefit from it.

"There are three key ingredients which we are seeing come together, that are really propelling momentum in the technology space - cloud, data and AI.

It is the combination of these three core ingredients; where we are able to ingest data at scale from any source, of any format, into a unified data estate, where we have access to essentially limitless compute, limitless storage, in a de-coupled fashion so that they can be scaled independently of one another, and where we are able to use AI to reason over that data, in order to unlock those deeper, more meaningful insights.



Organisations that are harnessing those three ingredients effectively, will be well positioned to capture a competitive edge."

Taygan Rifat - Partner Technology Strategist, Microsoft



Cloud services models:

The 3 established cloud service models are:

- 1. Infrastructure-as-a-Service (laaS)
- 2. Platform-as-a-Service (PaaS)
- 3. Software-as-a-Service (SaaS).

In this section, we will summarise each type of service in relation to data warehousing, along with their various benefits and drawbacks.

All of the cloud approaches deliver their capability over the internet, rather than physically. In general, the differences between the different cloud models can be neatly summarised using the following diagram which shows areas of responsibility between the client and the cloud provider:



1. laaS:

Infrastructure as a Service involves running your servers on cloud virtual machines, rather than on your own hardware. It is often the first step that organisations consider when looking to develop cloud maturity. This approach allows organisations to quickly create new servers with designated specifications without having to worry about purchasing or maintaining hardware. It also allows a degree of scaling of the server to provide flexibility with storage and compute.

The organisation would still be responsible for installing and patching operating systems and maintaining any other server requirements such as software that is needed to develop and run a data warehouse platform.

Benefits:

Most organisations with experience of on-premise data warehouses and servers consider this the simplest cloud move if they are looking to simply migrate existing data warehouses into the cloud with the least amount of effort or learning. It brings the benefit of SOME cloud flexibility with scaling servers without having to worry about pre-sizing hardware, reduction in maintenance effort for hardware software O/S and typically requires the same skill sets that managed the existing systems. This model also allows you to have new servers commissioned and up and running quickly.

Drawbacks:

The PAYG model in IaaS is limited and scaling events still need planning and coordination. You cannot separate storage and compute to scale these efficiently for typical data warehouse processing patterns easily and you will still need to manage much of the typical infrastructure necessities such as disaster recovery, firewalls, high-availability etc. Although you do not need to purchase hardware, in the long run, you will likely pay more for your IaaS capability in comparison to the like-for-like on-premise capability.

Suitability:

laaS suits organisations that have a well-established, in-house IT capability and data warehouse solutions that they simply want to move to the cloud to move away from physical hardware, but keep as close to the original solution as possible.



'||;

2. PaaS:

Platform-as-a-Service provides a package of services and capabilities to develop, run and manage applications without the complexity of maintaining the underlying infrastructure and operating systems.

PaaS offerings completely take over the responsibility for infrastructure of servers, storage, networking, and middle-ware, to deliver a complete platform of services such as database management systems, messaging systems, storage and others.

Users of a PaaS service simply look after the implementation of their developed applications. Organisations remain responsible for configuring and developing with the PaaS tools to build their cloud data warehouse.

Benefits:

This version of cloud service allows organisations to separately procure and scale storage and compute services as required, without the need for any server configuration. It provides much more flexibility in terms of independent scaling events and services without the need to procure servers.

For data warehousing, the ability to flexibly and quickly scale storage and compute is important due to the nature of the data warehouse processing requirements. At times, there will be a need for much heavier compute than others and storage will always grow for data warehouses. PaaS also provides high resilience and disaster recovery, which can be increased easily if required without the need to manage clusters or backups.

Obtaining PaaS resource for ad-hoc or POC work is also very simple and would not require any software installation or planning for where it would run.

PaaS allows much more opportunity to optimise spend at a lower granularity, saving on hardware and software support cost and effort and allowing granular scaling of resources for specific use-cases.

Drawbacks:

This model requires a different way of thinking when it comes to budgeting and estimating cost. Due to it's PAYG model, it does not require heavy CAPEX outlay but it is more difficult to calculate the true OPEX since it is meant to allow flexible usage.

The PaaS services may not be identical to the equivalent on-prem service and may require slightly different approaches to achieve the same outcome.

There will be a learning curve for developers to learn the PaaS approaches compared to the lift-and-shift approach required for IaaS.

Suitability:

PaaS suits organisations wanting to gain more benefits from the cloud such as flexibility and scaling, especially when delivering new data warehouse solutions. It delivers a much more modern cloud data warehouse capability with future-proofing built in. For data warehousing, there are a number of different levels of PaaS offering, ranging from requiring a full build using PaaS resources, to PaaS data warehouse automation.

These offerings are potentially suitable for different organisations with differing skill-sets and priorities.

3. SaaS:

Software as a Service provides a completely managed service for delivering the required solution. In this scenario, the organisation's solution would be completely hosted and managed by the SaaS provider with little to no interaction with any component services. Users simply interact with the SaaS solution online.

Benefits: This is the simplest and quickest way to obtain the functionality required, with no responsibility for development, support or maintenance of the application service, or previous experience on developing data warehouses.

Drawbacks:

There is no control over features and functionality, which may be a problem if an organisation's use-case is atypical and not covered in existing functionality. There may also be lock-in where it is not possible to leave the service without losing the capability entirely.

Suitability:

Organisations that want to obtain the capability of data warehousing but do not want to run a data warehouse team, or do not want to develop this expertise within the organisation. Careful assessment of the offering is important to understand the approaches on offer.



"Companies that struggle with gaining the benefits from their big data and disparate silos are at risk of becoming extinct. To become data-centric, stakeholders must be able to understand their current situations through a cloud-enabled holistic view of their company's data."

> Forbes Technology Council

How to build a Cloud Data Warehouse

Once you have concluded that a cloud data warehouse is right for your business, the next question is how to build one.

Implementing a cloud data warehouse requires much more than just selecting your technology stack. Whether you choose Azure or AWS as your platform and within them select various components such as Snowflake, Synapse or Redshift, Data Factory, Glue, Data Lake, you will still need to consider the following:

- **Platform architect** how all the components fit together, their responsibilities and secure integration between them.
- **Data warehouse methodology** there are a number of well-established and well-proven methodologies. Understanding and following one will be important to maintain data model consistency.
- **Design features and functionality** as previously mentioned, there are a number of expected features and functions a data warehouse platform should provide.
- **Implement coding patterns and templates** this ensures development happens consistently across the development team.
- **Manage source control, testing and deployments** following good development practices to ensure delivery of quality solutions.





When considering all of the above, there are four main options an SMB can take to build a data warehouse:

1. Build your own

We have seen many organisations start down this route under the belief that it is the cheapest. However, this option takes the longest time to deliver, requires a dedicated team of expensive data warehousing experts and has the highest chance of over-run. The main advantage of this approach is that it provides the most flexible and custom solution and may be suitable where there is already an existing data warehouse team and the data warehousing requirements have very unique aspects.

2. Partner

Similar to build your own in that this is a highly flexible approach, yet has the advantage that you do not need to run a permanent and expensive in-house team and the main benefit of leveraging your Partner's huge depth in experience. This is more expensive initially than building in-house, and potentially leaves you with a continued dependency on consultancy resource in BAU. This option would suit organisations which have complex data requirements that a custom solution is the only option, but do not have an experienced in-house team to do the work.

3. Data Warehouse Automation

Data Warehouse Automation (DWA) brings many benefits to delivering a data warehouse; from the acceleration of output and standardisation of approaches, to the inclusion of all expected features. Having this capability is hugely beneficial for inexperienced, under-resourced teams to ensure they have a much higher chance of delivering a data warehouse solution successfully.

There are a number of automation tools in the market to help you build your warehouse. However, many of these tools are designed for enterprise organisations that wish to build and maintain many data warehouses, and the tools are priced accordingly.

Furthermore, the tools are usually aimed at technical developers and frequently focus on the data modelling aspect of the warehouse, not automating the underlying cloud platform and services. This option usually suits enterprise-sized organisations that require a greater output from their data warehouse teams.

4. PaaS Data Warehouse Automation

This takes Data Warehouse Automation a step further, offering it as Platform as a Service which provides end-to-end (PaaS) the architecture management as well as the data warehouse automation capability in a single, integrated solution, rather than having automation software for the data warehousing and still having to plan and manage the data platform components separately.

When developing solutions in the cloud, it is important to optimise resource usage for your use case. This is one of the big advantages of cloud resources, but done incorrectly, it could actually make your solution more expensive to run in the medium to long term. This is something that PaaS DWA should do for you. **Automation**

TrueCue Platform: Leading the Way in Cloud Data Warehouse

As a team of data warehouse practitioners who have delivered many data warehouse solutions over the years, we recognised the need for an automated data warehouse platform aimed at all organisations and not just enterprises.

We created the TrueCue Platform to be the simplest and quickest way to deliver cloud data warehousing for every organisation, regardless of technical experience and capability.

By automating the data modelling, data pipelines and cloud services orchestration, the TrueCue Platform provides a robust and scalable PaaS solution, based on the Azure cloud and implementing best-practice data warehousing methodology that any organisation can use.

The cloud resources are automatically optimised to ensure you achieve value from your cloud spend.



Benefits to the Organisation:

Along with the cloud benefits covered earlier in this book, the TrueCue platform provides additional benefits that specifically tackle the typical data warehouse challenges as well as bringing faster Return-on-Investment (ROI) for your cloud data warehouse platform:



Accelerate Cloud Maturity:

We accelerate an organisation's cloud data maturity by delivering the foundational data warehouse fully up and running rapidly so that organisations can concentrate on the analytics required, build on solid foundations, all without needing an expensive and experienced technical team.



Work Agile

We enable agile data warehouse working due to our ability to auto-configure and build models, data and pipelines and deliver fully working portions of the model quickly and independently of other parts.

Our publishing process takes care of the processes typically required to deploy and release code and changes into environments without the typical heavy DevOps and Regression testing efforts needed. This directly tackles the BAU challenge whereby even small changes to a data warehouse process can take a lot of effort to scope, document, develop, test and deploy once the data warehouse is in use.



Accelerate Team Output:

We greatly accelerate a data warehouse team's output, removing this bottleneck to achieving your analytics goals and allowing the team to deliver value much earlier in the data warehouse development life cycle.

This reduces the frustration often felt by business teams that do not fully appreciate the effort involved in making even small changes to a data warehouse, especially once business-as-usual has been established.



Azure Expertise out-of-the-box:

As a fully managed Azure platform, the architecture, scaling, deployment and scaling of services within the Azure platform are automatically managed. This ensures that the cloud services are used in an optimal manner, providing confidence that all the advantages of the cloud are being leveraged.

TrueCue Platform: Leading the Way in Cloud Data Warehouse Automation

In Summary:

Automation at our core:

TrueCue provides the simplest and quickest way to deliver a fully functional and productionised cloud data warehouse running with your data and modeled for your business, all in Azure, following best-practice data warehouse patterns.

Accelerate Cloud Maturity:

We enable rapid cloud maturity acceleration, bringing you fully into the cloud, with all its benefits, and allowing you to take advantage of the additional analytics offerings within Azure, such as Power BI and Azure Machine Learning.

Enterprise Capability:

Our platform delivers all the features and functionality expected of a secure, scalable and robust enterprise data warehouse platform, regardless of your organisation's size, so that you can benefit from all of these without needing to develop them.

	TRUECUE TE	IANT: Monument	Home	Feed Configuration	Schedule and Monitor	File upload View Buil	der Weelin Lim	© [→	
	Welcome to the TrueCo The TrueCue portal will allow you to main	tain, monitor and manage	your Data Warehous	se					
	FEED CONFIGURATION							_	
	37 Dimension		7 <u>+</u> †‡ F	act	> 5	🔗 Mapping			
	FEED CONFIGURATION								
	25 Active Schedules		1 Inact	tive Schedules					
l	D CONFIGURATION								
	0 1 Dimension		22 🌔 F	Processed					
	FEED CONFIGURATION								
	Process Admin		Data	Quality Controller					

'||;

Cloud Data Warehousing Playbook for SMBs The Future of Cloud Data Warehousing for Small and Medium Businesses is Now

The Future of Cloud Data Warehousing for Small and Medium Businesses is Now

Still not convinced that you need to start taking your data management seriously?

Consider that this year, COVID-19 has created a 'new normal' for the entire world. According to Visa's Back to Business study 12, nearly eight-in-10 consumers worldwide (78%) have changed how they pay in order to reduce contact, with more than two-thirds of SMBs (67%) having had to try a new approach to payments, from launching an e-commerce site or changing POS technology, to keep their business on track. How are you effectively managing all of this newly acquired data?

Add to that, your company is most likely working in part or completely remotely, so measuring productivity and customer satisfaction just became all the more challenging.

How do you take the data from video conferencing apps and your customers' sentiments from social media platforms and harness it in combination with your ERP data? The world and the way we work has changed forever.

All of the data you are increasingly amassing, coupled with world events such as pandemics, tsunamis, nuclear meltdowns and war, means that without the capabilities which a cloud data warehouse brings, you will be at a significant disadvantage, and considerably more exposed to risk.

12 The Visa Back to Business Study, 2020

New technologies and other world events will always serve to create new data challenges for small business owners. The question is, will you be ready? Remember, the ability to gather data from disparate sources and analyse it for strategic decision-making is the key to the gold mine. Your job is to get there before the competition does.

So, whilst embarking on your data analytics maturity journey may seem like a daunting challenge, the reality is, that this is a journey which every business, regardless of size, must inevitably embark on.

Doing so with a cloud data warehouse on-board will make your journey much less of a bumpy ride.





About TrueCue

TrueCue's mission is to empower business leaders with truth and certainty from their data. They provide high-impact consultancy and data-driven solutions underpinned by their cutting-edge technology platform – a PaaS data warehouse automation platform built for the Microsoft Azure Cloud.

Combining analytics, consulting and technology expertise to create data-driven solutions, the TrueCue team supports organisations on their end-to-end data and analytics journey – covering data management, visualisation, operational reporting, advanced analytics and data science.

Their solutions are widely used across the public, private and third sectors, with clients including the NHS, Nuffield Health, Imperial College Health Partners, Local Government, EDF, HSBC, and GlaxoSmithKline, to name but a few.

So, whether its understanding patient care to optimise treatment strategies, driving customer insight to create the next sales campaign, or understanding donations across a charitable trust, their solutions have lasting value which drive performance improvements, so that you can take action on what matters most, with confidence.

TrueCue are a Microsoft Gold Partner, Tableau's-longest standing EMEA Partner, and Alteryx EMEA 2019 Partner of the Year.

TFECUE

THE TRUTH IS WHAT DRIVES SUCCESSFUL DECISIONS.

But, most businesses have no idea what that truth is.

They know the threads which will lead them to the truth are out there somewhere, but they're muddled, hard to find and often contradictory.

We provide a single source of the truth to business leaders craving certainty.

We do this by providing best-in-class technology to filter out the noise and untangle the knots.

Our people help companies identify the right questions and work with them to find the long term, sustainable answers.

Our philosophy is about providing real value, so that working with us isn't a flash in the pan,but something that is as useful today as it will be in a decade.

And this all happens at the speed of modern business.

