

Cloudification of on-premises Data Warehouses

www.hexaware.com

Table of Contents

Introduction 2

Data Warehouses – Where and Why are they used 2

On premise Data Warehouse 2

Cloud Data Warehouse 3

Key Metrics: On-premises vs. Cloud 3

Comparing Key Metrics: On-premises vs. Cloud 3

What benefits does Azure provide 3

How do you migrate to cloud? 4

Key Metrics for Data Warehouse Migration to Cloud 4

AMAZE™ for Data & A.I. – Value Offering 5

Closing comments 5

References 5



Introduction

Businesses that have successfully migrated to cloud over the years have realized tremendous growth in their revenue streams and operations, while simultaneously making their business more agile and responsive to the demands of a constantly changing environment [1]. Driving these digital capabilities has resulted in cloud migration market to grow at roughly 30% [2]. The rate of adoption has been accelerated by the pandemic where the world is looking for a more sustainable solution.

In the case of data migration specifically, by 2022, 50% of new system deployments will be based on a cloud data management ecosystem and by 2023, 75% of all databases will be on a cloud platform [15].

However, many large organizations run the risk of an unsuccessful data migration to cloud due to a host of problems ranging from defining the right cloud architecture to operational inefficiencies in the migration process [16]. There are also budgetary overruns that render the objective of migrating to cloud seem meaningless. As a result, companies face delays which translate into unmet business potential or in some cases cloud repatriation. Understanding the data warehouse landscape and its challenges therefore becomes important to fully realize the benefits of cloud migration.

On premise Data Warehouse

Data Warehousing usage became very prominent during late 80s that saw a rapid advancement in late 90s and early 2000s to become a core part of the Information Technology group across large enterprises.

Some of the vendors like IBM Netezza, Teradata started offering customized hardware to manage data warehouse architectures within state-of-the-art machines, to be on top of the list of priorities for CIOs and CTOs [5].

Benefits of having Data Warehouses on-premises is that it allows a greater degree of control vis-à-vis usage, cost of company's hardware & software and access grants, negligible network latency if any, information security and governance (such as GDPR).

However, since the hardware capacity is prepared for feeding the peak and ad hoc demands, utilization of resources isn't optimal which translates to **higher CapEx and OpEx.**

Additionally, maintenance and support of both hardware and software adds to the total cost of ownership. The IT infrastructure team has full responsibility to ensure that the underlying infrastructure stays up and running efficiently, reliably, and securely [6].

Data Warehouses – Where and Why are they used

Data Warehouse is a system for collecting and managing data from company's operational databases (such as ERP systems) and external sources, making it easier for business leaders to analyze data over a specific period [3]. However, the data entering a data warehouse must be structured and formatted.

Data Warehouses make use of ETL tools (Extract Transform and Load) which extract the data from various data sources, transforms it in the staging area and then finally, loads it into the Data Warehouse system. The users can access the processed data in the Data Warehouse through Business Intelligence tools, SQL clients, and spreadsheets [4].

Data Warehouse on cloud differs from Data Warehouse on-premises in that it is a platform-as-a-service (PaaS). A business pays for the storage space and computing power it needs at a given time and scalability is a simple matter of adding more cloud resources.



Cloud Data Warehouse

Data Warehouses on cloud stand to gain where their on-premises counterpart fails [18]. CapEx costs are cut straightaway substituted by the pay-as-you-go model. Meaning, customers have the flexibility to add and remove resources as per their needs, and the bill is generated only for the period that the resource is used for.

Typically, cloud data warehouses have several servers that can balance the data load, increase processing speeds, and serve multiple areas of the business in different geographies, thereby resolving latency issues.

Concerns around data security are also addressed with use of virtual private network (VPN) and cloud encryption services like multi-factor authentication making transporting of data across regions and resources incredibly secure [7].

Key Metrics: On-premises vs. Cloud

Whether to keep Data Warehouses on-premises or on-cloud boils down to two Key Metrics:

1. **Total Cost of Ownership (TCO)** – Broadly divided into server, storage, compute, and personnel costs, it helps you to precisely calculate economic impact during the life cycle of an IT project(s). By migrating to cloud, you can add/remove/modify resources according to your need which frees up unnecessary costs, thereby reducing overall TCO.
   (to estimate the cost savings you can realize by migrating your workloads to Azure, click here)
2. **Downtime per year -** The amount of time a server or application is online and accessible to end user provides for a greater SLA and therefore lesser downtime. Cloud users get to experience SLAs of the order of 99.99%, compared to 99%-99.9% for on-premises. This translates into time attributable to support and maintenance.
3. **Time spent on data analytics** – Businesses today are run on time-sensitive queries and reports, demanding companies to reduce the time spent on non-value adding services such as support and maintenance. Because legacy infrastructure is complex, we often hear that businesses continue to invest in hiring people to manage those outdated systems, even though they're not advancing data strategy or agility [8].

Comparing Key Metrics: On-premises vs. Cloud

To put it into perspective, we compare the key metrics for a US Enterprise looking to migrate from Teradata Vantage (on-premises Data Warehouse) to Synapse (cloud Data Warehouse).

Key Metrics	On-premises	Cloud
TCO (over 5 years)	$26.83 Mn	$13.14 Mn
Downtime per year	8.76 hours	52.56 minutes
Time spent on data analytics	15%	100%

These kinds of savings are similar across the top three Cloud Service Provider (CSPs) – Amazon's AWS, Microsoft's Azure and Google's GCP – that dominate the market with a combined share of ~60%. However, each has its own USP. We look at what makes Azure special in terms of its service offering [9].

What benefits does Azure provide

With presence in over 51 regions around the globe and 7 on their way, Azure is expanding its services far and wide to address user concerns related to availability and latency. Policy measure pertaining to different geographies are also thereby taken care of.
(to know which regions does Azure serves in, click here)

Many Windows and Linux based software are also cloud compatible which translates to cost saving in the form of single license fee (for both on-prem and cloud) and the amount of re-platforming required. Additionally, it also offers a TCO savings of the order of 40%-50%.

Having the option of a hybrid cloud architecture allows customers to choose which data and applications they want to keep on-premises and which ones they want to migrate to cloud. Benefit of such a structure is that they have a greater degree of control compared to a public cloud [10][11].

Microsoft Azure is the most secure CSP network with over 90+ security certifications that ensures only authorized users get access to the data. Access to customer data is denied and least privilege is granted in case of performing task such as auditing and log access. The access-control requirements are established as per the CSP's Security Policy.
(to know how Azure Security Policy works, click here)



How do you migrate to cloud?

In principle, there are two major ways in which one can migrate legacy system to Data Warehouses on cloud: **Rehosting and Re-platforming**. The quickest and easiest approach is to rehost (Lift and Shift) legacy Data Warehouse technologies to next generation hybrid, cloud, MPP, or open source databases. However, organizations cannot simply lift and shift their Data Warehouses as this would be suboptimal use of cloud economics. It would also mean that some of the cloud-native benefits such as continuous real-time deployment may not be possible [19].

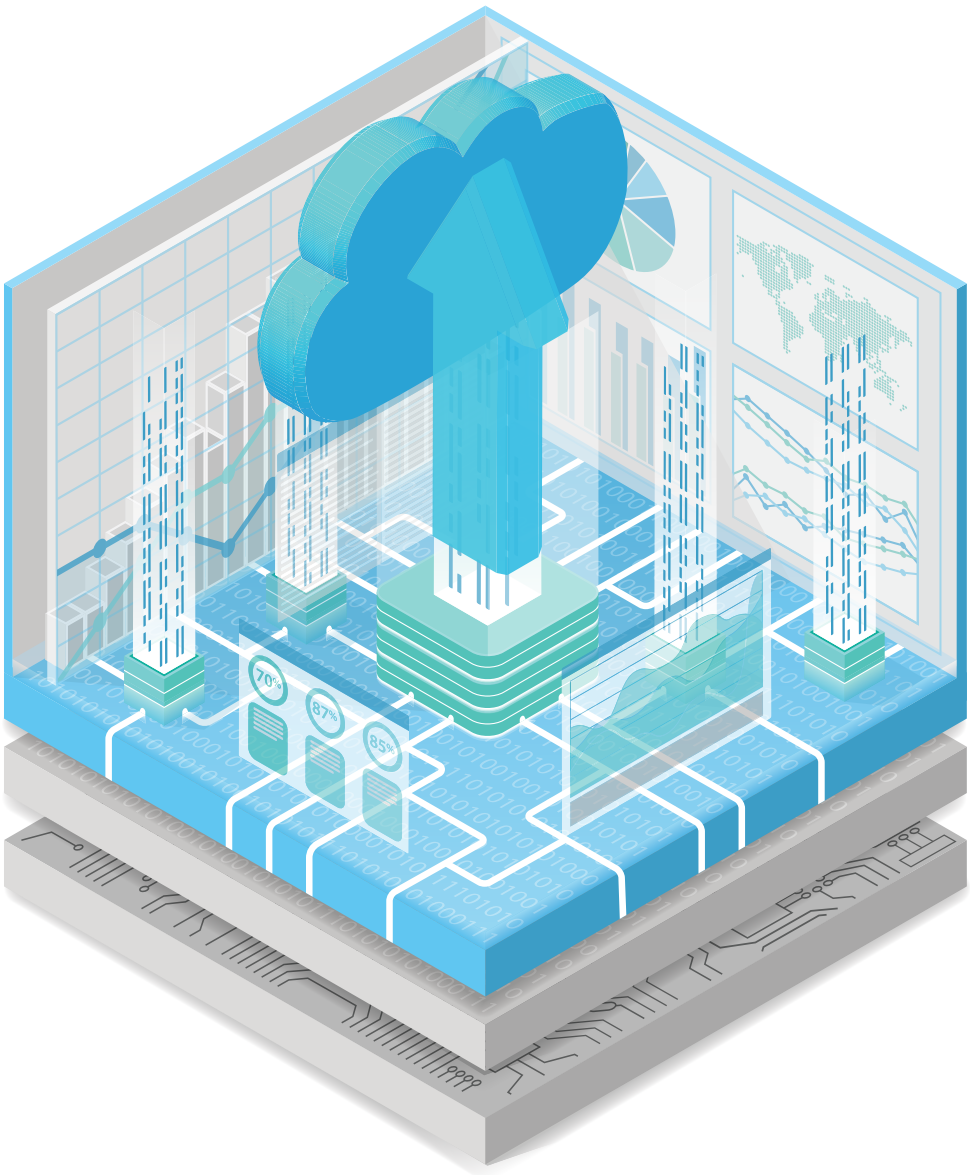
Re-platforming on the other hand is the most popular approach as it is optimized for cloud use, whether you will rip-and-replace the warehouse's primary platform or augment it with additional data platforms. In this, the data engineers migrate data schema, data pipeline and visualization tools to cloud [12].

However, certain processes such as code assessment can be automated to reduce the implementation cost, TCO, and human intervention. In doing so the processes can be finished in hours which would have otherwise taken days, weeks or even months to do manually.

Key Metrics for Data Warehouse Migration to Cloud

Migration can be done either manually by the customer or with the help of a Managed Service Providers or MSPs. Following key metrics help understand which option of the two is better:

1. **Cost of Migration** – Most third-party cloud Data Warehouse migration tools available in the market come with a high license cost and don't offer customization. Additionally, the level of automation provided by such tools further adds to the total cost [13]
   (to know more on Azure Migrate Pricing, click here).
2. **Duration or Time to Market** – The time it takes for a customer to migrate their Data Warehouses from on-premises to cloud. Most solutions in the market automate schema and report migration but offer a manual approach for ETL code conversion. End-to-end automation reduces both time and probability of an error occurring during migration [14].



AMAZE™ for Data & A.I. – Value Offering

Customers while manually migrating Data Warehouses to cloud, typically face a higher cost of migration and a schedule overrun, among other perils (eg: data security). With AMAZE™, however, they can realize obvious benefits within those metrics.

AMAZE™ for Data & A.I. is a cloud re-platforming service which takes account of both the aspects of automation and necessary human expertise to deliver across all key metrics requisite for a successful cloud migration. Again, putting into the context of the **US Enterprise:**

Key Metrics	Manual Migration	AMAZE™	Improvement
Cost of Migration	$98k	$39k	60% reduction
Time to Market	6 months	3.5 months	40% faster

While it is reasonable to assume that roughly 60%-70% of the work can be automated, some key processes such as configuration of data and analytics for cloud, creating new environment dependencies, testing and validating data need human intervention.

The entire re-platforming (cloud migration) process is divided into three phases: **Data Capture & Assessment, Readiness & Build, and Migration to Azure.**

In the **Data Capture & Assessment phase**, customer's IT landscape is analyzed, and a deeper insight is generated into the existing Data Warehouses and analytics environment (Schema, ETL scripts and Visualization reports, hereby referred to as data ecosystem). A cloud-readiness report then gets created which identifies the changes required to make the application cloud ready while highlighting the exact lines of code that need to be changed. (use details from ppt)

AMAZE™ makes use of Metadata extraction while assessing the landscape. It ensures that no customer data is exposed (or manipulated), thereby improving the security of the process. This phase is almost fully automated (95% and above) owing to AMAZE™'s automation-based discovery tool.

Next in the **Readiness & Build phase**, based on the assessment results and the services required by the customer, the best cloud architecture (hybrid/single) is chosen. Then with the help pf object mapping data ecosystem gets modified to suit the new architecture and the apparent functional gaps are filled. Again, 60%-70% of the process can be automated to reduce cost and increase speed.

Lastly, the **Migration to Azure phase** sees a phase-wise implementation of data ecosystem migration (data schemas, objects, and pipelines) and conversion (reports and dashboards).

In each of these phase-wise migrations key metrics, benchmarks, guidelines, and best practices are agreed upon mutually. Though this way the process takes more time than the conventional "lift-and shift" approach, but the risks are mitigated and learnings from initial phases can be implemented in the following phases.

The biggest value add of this approach is that the entire migration process happens with zero downtime and since no externally licensed product is used for migration, 50%-60% cost savings is realized in customer's data ecosystem TCO.

Closing comments

Additionally, analytical capabilities of the customer can be enhanced by making use of cloud native development of applications and open source projects.

Doing business using cloud Data Warehouses is no longer a thing of the future as every organization looks towards adopting cloud for far reaching benefits. It therefore makes sense to invest in technologies such as automation and real time analytics to achieve a successful migration of legacy Data Warehouses or systems at low cost & risk and high speed & return.

References

1. TechRepublic.com. 2020. Report: 95% Of Businesses Have Migrated Critical Applications to The Cloud. [online] Available at: https://www.techrepublic.com/article/report-95-of-businesses-have-migrated-critical-applications-to-the-cloud/
2. Mordorintelligence.com. 2020. Cloud Migration Market | Growth, Trends, Forecasts (2020 - 2025). [online] Available at: https://www.mordorintelligence.com/industry-reports/cloud-migration-services-market
3. Talend.com. 2020. What Is A Data Warehouse? Definition, Concepts, And Tools. [online] Available at: https://www.talend.com/resources/what-is-data-warehouse/
4. Oracle.com. 2020. What Is A Data Warehouse? [online] Available at: https://www.oracle.com/in/database/what-is-a-data-warehouse/
5. Diyotta.com. 2020. Data Warehouse: Definition, History, And Evolution. [online] Available at: https://www.diyotta.com/data-warehouse-definition-history-and-evolution/
6. Stitchdata.com 2020. On-Premises Vs. Cloud Data Warehouses: A Comparison. [online] Available at: https://www.stitchdata.com/resources/compare-on-premises-and-cloud-data-warehouse/



7. Oracle.com. 2019. 6 Benefits of a Cloud Data Warehouse. Available at: https://blogs.oracle.com/database/cloud-data-warehouse-benefits

8. Google Cloud Blog. 2020. 5 Reasons Your Legacy Data Warehouse Won'T Cut It. [online] Available at: https://cloud.google.com/blog/products/data-analytics/5-reasons-your-legacy-data-warehouse-wont-cut-it

9. Intellipaat.com. 2020. AWS Vs Azure Vs Google - Detailed Cloud Comparison. [online] Available at: https://intellipaat.com/blog/aws-vs-azure-vs-google-cloud/

10. SaaSplaza.com 2020. What Are The BENEFITS Of Microsoft Azure? [online] Available at: https://www.saasplaza.com/blog/what-are-benefits-microsoft-azure

11. Microsoft.com. 2020. Rethinking The Business Benefits Of The Cloud. [online] Available at: https://azure.microsoft.com/en-in/resources/videos/rethinking-the-business-benefits-of-the-cloud/

12. dbseer.com. 2019. 3 Options For Migrating Applications To The Cloud: Rehost, Replatform, And Refactor. [online] Available at: https://dbseer.com/3-options-migrating-applications-cloud-rehost-replatform-refactor/

13. Google Cloud Blog. 2020. Data Warehouse Migration Challenges And How To Meet Them. [online] Available at: https://cloud.google.com/blog/products/data-analytics/data-warehouse-migration-challenges-and-how-to-meet-them

14. Transforming Data with Intelligence. 2019. Plan Carefully When Migrating to A Cloud Data Warehouse. [online] Available at: https://tdwi.org/articles/2019/11/07/ta-all-plan-carefully-when-migrating-to-cloud-data-warehouse.aspx

15. Thanaraj, R., 2020. Cloud Data Warehouse: Are You Shifting Your Problems To The Cloud Or Solving Them? Gartner Research.

16. Friedman, T., 2009. Risks and Challenges in Data Migrations and Conversions. Gartner Research.

17. Friedman, T., 2009. Best Practices Mitigate Data Migration Risks and Challenges. Gartner Research.

18. Smith, D., 2017. Cloud Strategy Leadership | Gartner Insights on How and Why Leaders Must Implement Cloud Computing. Gartner Research.

19. Business Reporter Production. 2020. Migrating Legacy Applications to The Cloud. Forbes.

About Hexaware

Hexaware is the fastest growing next-generation provider of IT, BPO and consulting services. Our focus lies on taking a leadership position in helping our clients attain customer intimacy as their competitive advantage. Our digital offerings have helped our clients achieve operational excellence and customer delight by 'Powering Man Machine Collaboration.' We are now on a journey of metamorphosing the experiences of our customer's customers by leveraging our industry-leading delivery and execution model, built around the strategy— 'Automate Everything™, Cloudify Everything™, Transform Customer Experiences™.'

We serve customers in Banking, Financial Services, Capital Markets, Healthcare, Insurance, Manufacturing, Retail, Education, Telecom, Professional Services (Tax, Audit, Accounting and Legal), Travel, Transportation and Logistics. We deliver highly evolved services in Rapid Application prototyping, development and deployment; Build, Migrate and Run cloud solutions; Automation-based Application support; Enterprise Solutions for digitizing the back-office; Customer Experience Transformation; Business Intelligence & Analytics; Digital Assurance (Testing); Infrastructure Management Services; and Business Process Services.

Hexaware services customers in over two dozen languages, from every major time zone and every major regulatory zone. Our goal is to be the first IT services company in the world to have a 50% digital workforce.

NA Headquarters
Metro 101, Suite 600,101
Wood Avenue South,
Iselin, New Jersey - 08830
Tel: +001-609-409-6950
Fax: +001-609-409-6910

India Headquarters
152, Sector – 3
Millennium Business Park
'A' Block, TTC Industrial Area
Mahape, Navi Mumbai – 400 710
Tel : +91-22-67919595
Fax : +91-22-67919500

EU Headquarters
Level 19, 40 Bank Street,
Canary Wharf,
London - E14 5NR
Tel: +44-020-77154100
Fax: +44-020-77154101

APAC Headquarters
#09-01, One Finlayson
Green, 1 Finlayson Green
Singapore-049246
Tel : +65-63253020
Fax : +65-6222728

Australia Headquarters
Level 3, 80 Mount St
North Sydney NSW
2060, Australia
Tel : +61 2 9089 8959
Fax : +61 2 9089 8989

Safe Harbor Statement
Certain statements in this press release concerning our future growth prospects are forward-looking statements, which involve a number of risks, and uncertainties that could cause actual results to differ materially from those in such forward-looking statements. The risks and uncertainties relating to these statements include, but are not limited to, risks and uncertainties regarding fluctuations in earnings, our ability to manage growth, intense competition in IT services including those factors which may affect our cost advantage, wage increases in India, our ability to attract and retain highly skilled professionals, time and cost overruns on fixed-price, fixed-time frame contracts, client concentration, restrictions on immigration, our ability to manage our international operations, reduced demand for technology in our key focus areas, disruptions in telecommunication networks, our ability to successfully complete and integrate potential acquisitions, liability for damages on our service contracts, the success of the companies in which Hexaware has made strategic investments, withdrawal of governmental fiscal incentives, political instability, legal restrictions on raising capital or acquiring companies outside India, and unauthorized use of our intellectual property and general economic conditions affecting our industry.

© 2020 Hexaware Technologies limited. All rights reserved.