DIGITAL TWINS

Big Data Strategy

"Let's suppose that a caterpillar goes into a cocoon and comes out just as a faster caterpillar. It's interesting, but not impressive because that's just a change. But the fact is that a caterpillar goes into a cocoon and emerges as a butterfly. That's the transformation, a real transformation. Isn't that what we all want?"

Team Members:

Srinivas Madhu Kumar Atyam

Nisha Pardeshi

Saranya Murugan

Arjuna Menon

TABLE OF CONTENTS

LIST OF FIGURES

I. What is Digital Twin? Describe the technology or content area.

A digital twin is a digital representation that mirrors a real-life physical asset, or a process, or a system, wherein data is transmitted continuously to the digital twin so that it can exist simultaneously with the physical entity. There are four broad categories of digital twins:

i. A process digital twin that provides a virtual view of an entire business process, such as the car manufacturing process in an automobile company.

ii. A component digital twin is used to represent a major sub-component of an asset that has a significant impact on its operation, such as the engine in a jet.

iii. An asset digital twin that provides a virtual rendition of an entire asset such as a wind turbine.

iv. A system digital twin digitally portrays a collection of interrelated assets together, such as an oil refinery.

All these types of digital twins allow us to analyze the data they capture to forecast new opportunities for a business. They foster a culture of data-based decision making to implement improvements which is far better than the practice of making decisions based on one's gut feelings.

Various scenarios can be run on these digital twins and the results obtained from these simulations can be studied to determine the best course of action to be taken on the real assets to optimize their performance, increase return on investment (ROI) or improve various other characteristics. Therefore, they are the basis for new products as services business models. Digital twins are also used to create a virtual prototype of a product to test whether the actual product creation is feasible or not. Due to this use, digital twin technology is also referred to virtual prototyping.

Thus, overall, it can be said that digital twins technology impacts how the design phase, build phase, and operations phase of a product are constructed in a life cycle.

The prevalence of the Internet of Things (IoT), which is the network of various devices created through the use of sensors, has enabled the digital twin to become a cost-effective modelling tool. Digital twins technology is listed among Gartner's Top 10 Strategic Technology Trends for 2018. In this rapidly evolving world, holding to the pace of technology is a must. As said by Thomas Kaiser "Digital twins are becoming a business imperative, covering the entire lifecycle of an asset or process and forming the foundation for connected products and services. Companies that fail to respond will be left behind."[1]

The term "digital twins" was coined by Dr Michael Grieves at the University of Michigan in 2002. However, this concept was actually first practiced by NASA in the 1960's when their scientists faced challenges to operate, maintain or repair systems that were not in close proximity or that were in space. NASA's famous Apollo 13 rescue mission in 1970 made use of a digital twin of this mission's aircraft to test various possible solutions to the problems encountered at ground level before conveying the chosen best solution to the astronauts in space. NASA now uses digital twins to develop next-gen aircrafts and roadmaps and their ultimate goal is to use this technology to create, test, and build their equipment in a completely virtual environment.

[1] https://www.forbes.com/sites/bernardmarr/2017/03/06/what-is-digital-twin-technology-and-why-is-it-so-important/#471aaf8b2e2a

II. What are the latest trends in Digital Twins?

Digital twins have a wide variety of applications ranging from improvements in asset performance and utilization, better maintenance of equipment such as turbines, jet engines, and locomotives, and designing and repairing of products such as airplanes.

The growth of the IoT, adaptation of cloud-based platforms that are cheap, and the increase in need for cost effective solutions is fueling the growth market of digital twins.

A few of the latest trends of this technology in various industries are listed below:

- General Electric (GE), in partnership with Noble Corporation, launched a digital drilling vessel to enhance the drilling experience.
- General Electric also uses digital twins of its wind turbines for real time maintenance and configuration changes during operations.
- Digital twins of smart cities (i.e. cities whose conditions are tracked in real time using sensors) are used to optimize traffic, smart grids, etc., and implement a host of new services.
- Descartes Lab Inc. is building a digital twin of the entire earth by combining geospatial data (data associated with a certain location) it gathers from satellites, sensors on vehicles and a variety of other sources. It plans on using this model to answer baffling science questions, and to relate its internal processes to real world conditions.
- Tesla creates a digital twin of every vehicle it manufactures, and it regularly provide software updates to its customers' cars based on the data it receives from the twins.
- Bombardier, the world's leading manufacturer of trains and aircrafts, partnered with Siemens to optimize its processes for developing, producing, and supporting products, through the use of a product development cycle digital twin.
- Three dimensional digital twins of automation systems are used by students at the Aarhus school of Marine and Technical Engineering to program automatic production lines.

- Boeing has used digital twins of its airplanes' development process to attain a 40% improvement in the quality of the parts and systems it uses to manufacture commercial and military airplanes.
- Microsoft's Azure platform has aided in creating digital twins of smart hotels, which are used by the management to optimize various activities such as room bookings, and power consumption. It also has a smartphone app that can be used by customers to adjust room settings as required.
- In the healthcare industry alone, there are a plethora of uses for digital twins and a few of these are as follows:
    i. Philips provides personalized healthcare by creating a digital twin of every patient through the use of each person's medical history and current condition. This twin can also predict the future needs of the patient to a reasonable degree of accuracy.
    ii. Digital twins of various organs and even the entire human body can be created, on which various procedures are simulated and studied using virtual reality. This has the potential to decrease mortality rates and improve patients' outcomes by providing optimal procedures tailored to each individual's unique features. These digital twins can also process real time data during the actual operation to provide best alternative solutions if unforeseen circumstances occur during the procedure.
    iii. General Electric engineers are planning on building bandage-sized sensors that monitor a person's vital details such as sugar level and heart rate so that these can be connected with a digital twin that analyses this data to alert doctors regarding any anomalies. This could even provide diagnostic recommendations with the integration of artificial intelligence.
    iv. Digital twin of an entire hospital helps the operations team create a safer environment along with expedited patient care.

III. How can big data, business intelligence, and analytics accelerate development in Digital Twins?

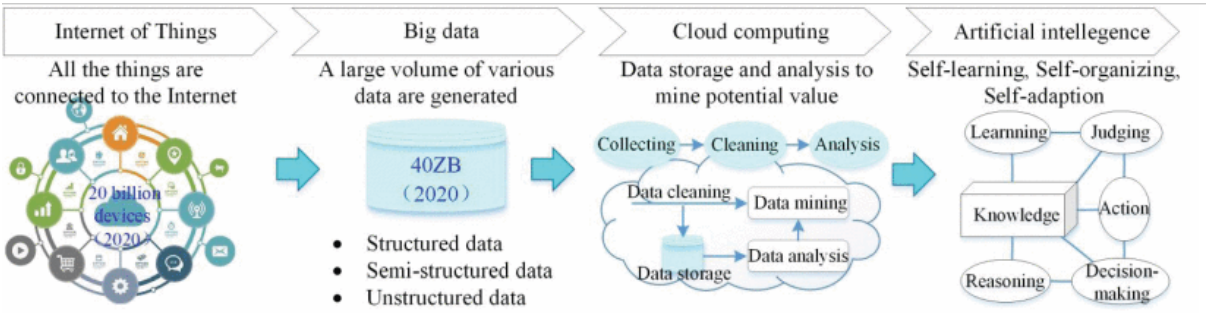
A digital twin can provide solutions to companies with a little data available, but for this technology to give meaningful insights to the business user in most cases, the digital twin should be connected to data with large scale and high velocity. The idea of digital twin technology revolves around the utilization of big data. Enterprises globally invest in big data platforms like IoT- based sensors, data management infrastructures, and on data scientists. These serves as the critical ecosystem for a digital twin to capture data, transform it, and perform predictions in real time to provide data driven decisions, automate business process, increase collaboration, and create new business models.

Digital twins relate to the rapid development of information technology and cyber-physical systems integration. Business intelligence, big data and analytics tools provide a faster and effective mechanism to implement digital twins effectively on a broader scale. A big data technology, Hadoop ecosystems, provides cheaper infrastructure to store real time big data that is usually collected passively through the IoT devices. Hadoop ecosystem hosts many open source stream processing software platforms like Apache Kafka, Spark, NiFi and Flume. Continuous streaming of both structured and unstructured data from end user to end user is possible by using Apache Kafka, Nifi and Flume. Fast real time processing and modeling of data is made possible by sophisticated systems like Apache Spark. Blockchain along with various software platforms provide a perfect combination for data security as this data cannot be modified due to the network architecture. Business intelligence (BI) tools like Tableau, PowerBI, Qlik Sense, Online Analytical Processing, and Google Analytics, enable businesses to collate, analyze and visualize data easily, which enhances the decision-making quality of the managers which helps them to come up with better strategic plans. Cloud platforms provides a total cheap infrastructure for saving big data and technologies required to transform, analyze,

make models, and offer insights in real time efficiently. Business analytics in general uses predictive analytics to analyze patterns and connections and provide insights.

A digital twin requires collection of real time data that is mostly acquired from IoT devices' sensors. The cloud and ecosystems like Hadoop accelerate the process of saving, transforming and processing the data. Various simulations of physical entities need to be replicated. Analytics plays an important role in this scenario and it also provides efficient mechanisms to get the best possible outcome for quick decision making. When the scenario is replicated from cyberspace to the physical entity, the accuracy of the digital twin can be improved and thus can be automated for further processes. Business intelligence tools help in getting insights which will enhance the decision-making power thus making timely decisions. Big data, analytics and BI together form a very powerful tool for implementing and enhancing the digital twin system in real time. The generic process is as shown in the below **figure-1.**

Figure-1Role of bigdata technologies in digital twins



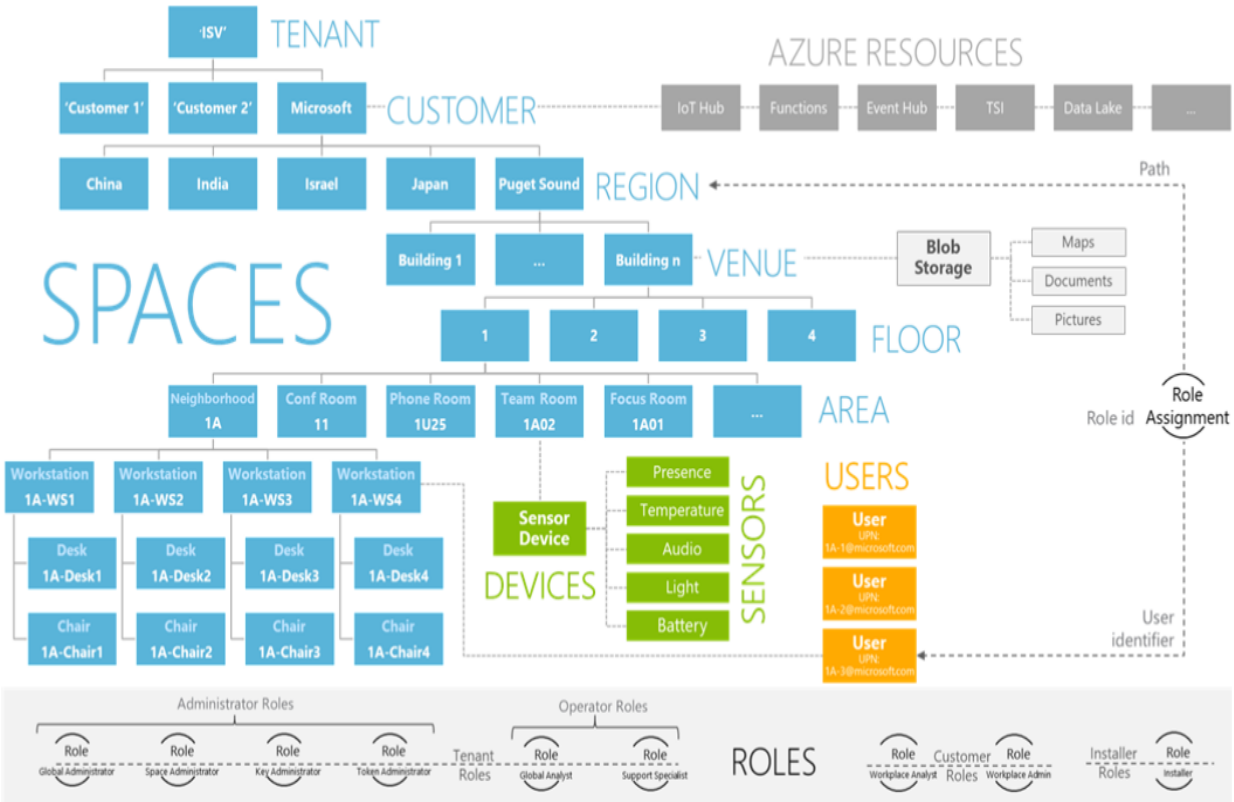
Microsoft Azure makes use of big data, business intelligence, and analytics technologies. It provides a platform for building extensive digital models and solutions that can be applied to any physical environment. With Azure digital twins, we can query data from physical space rather than from IoT sensors. It applies to all types of environments such as schools, hospitals, warehouses, stadiums, factories, banks etc.

Regardless of the real-life scenario, the corresponding digital instance of the object can be created using Microsoft Azure digital twins.

Azure digital twins has the following key capabilities:

- **Spatial intelligence graph:** It is a virtual representation of physical environment. Azure digital twins is an Azure IoT service, which eases the development by organizing domain specific concepts to models. These object models make up an ontology. A smart building's ontology may comprise of regions, venues, floors, offices, zones, conference rooms, and focus rooms. Many customizations and different scenarios exist based on combinations of models and ontologies. With models and ontology one can populate a spatial graph. The below figure-2 represents the spatial intelligence graph for a smart building's ontology.

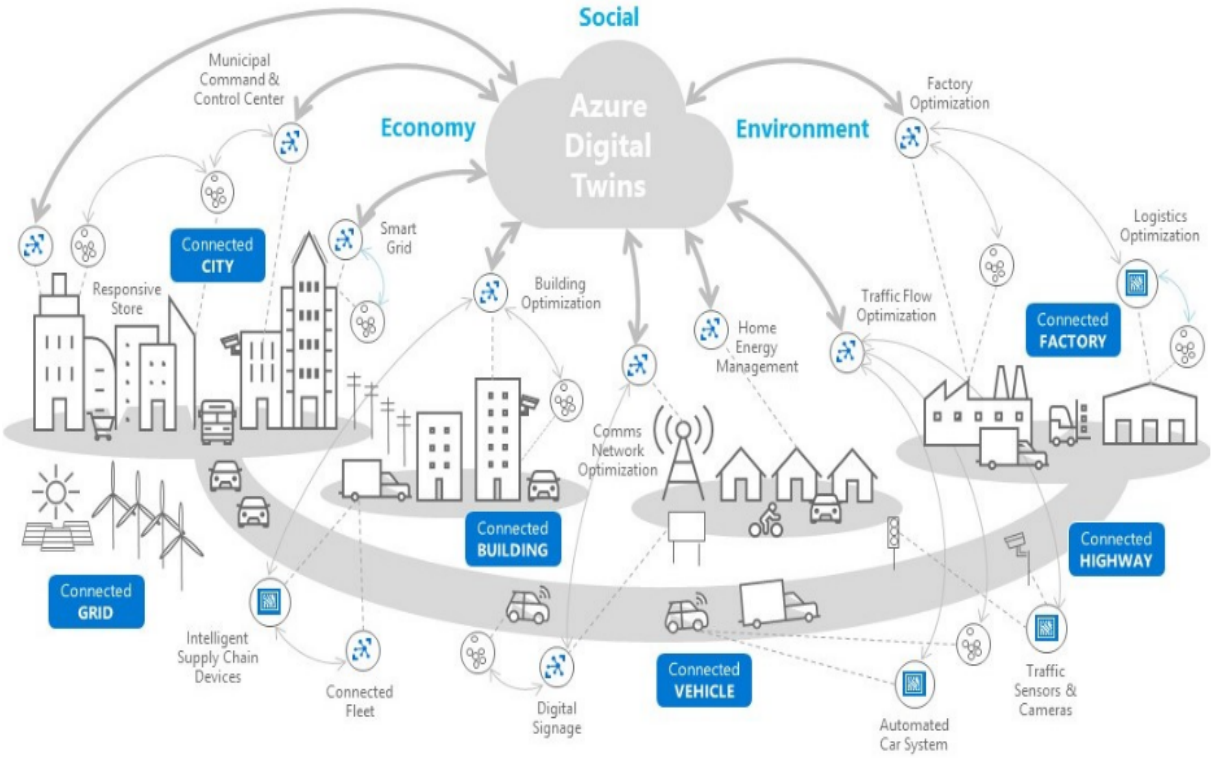
Figure-2 Microsoft Azure digital twin platform



- **Digital Twin Object Models:** They are predefined device protocols and data schemas.
- **Advanced compute capabilities:** Users can customize functions to generate events or notifications based on data from sensors. For example, based on the number of people, the temperature of a room is automatically controlled.
- **Ecosystem:** Azure digital twins instance can be connected to various stable Azure cloud services like Azure Stream Analytics, Azure AI, Azure Storage, Azure Maps, Dynamics 365, Office 365 etc.

One of the the smart hotel digital twin solution called Hotel 360^0 created using Azure digital twins makes use of Power BI to provide a simplified customizable dashboard containing information about the hotel to management. Power BI also provides a mobile application interface to the customer providing various services such as requesting ice to be sent up to one's room. This solution also makes use of the spatial intelligence graph feature of Azure to provide the vertical hierarchical view of the hotel, which consist of various levels that include a hotel overview, hotel floors, rooms per floor, etc. It provides different information at each level, such as the power usage and cleaning schedule at the floors level, and the temperature and lighting levels at the room level. It uses the extensible digital twins object models to differentiate between various room types such as conferences rooms and occupants' rooms by using extended properties such as balcony size, and whether a bed is present or not. The advanced compute capabilities this solution implements are user defined functions wherein, say, if a customer wants a particular lighting level and temperature to be set when s/he enters a room, the code required for this to occur can be written and then sent to the digital twin to get the desired result in the smart hotel. Hotel 360^0 also makes use of role-based access control in Azure to provide this service to various independent hotels in a secure fashion.

Figure-3 Microsoft Azure digital twin platform applications[2]

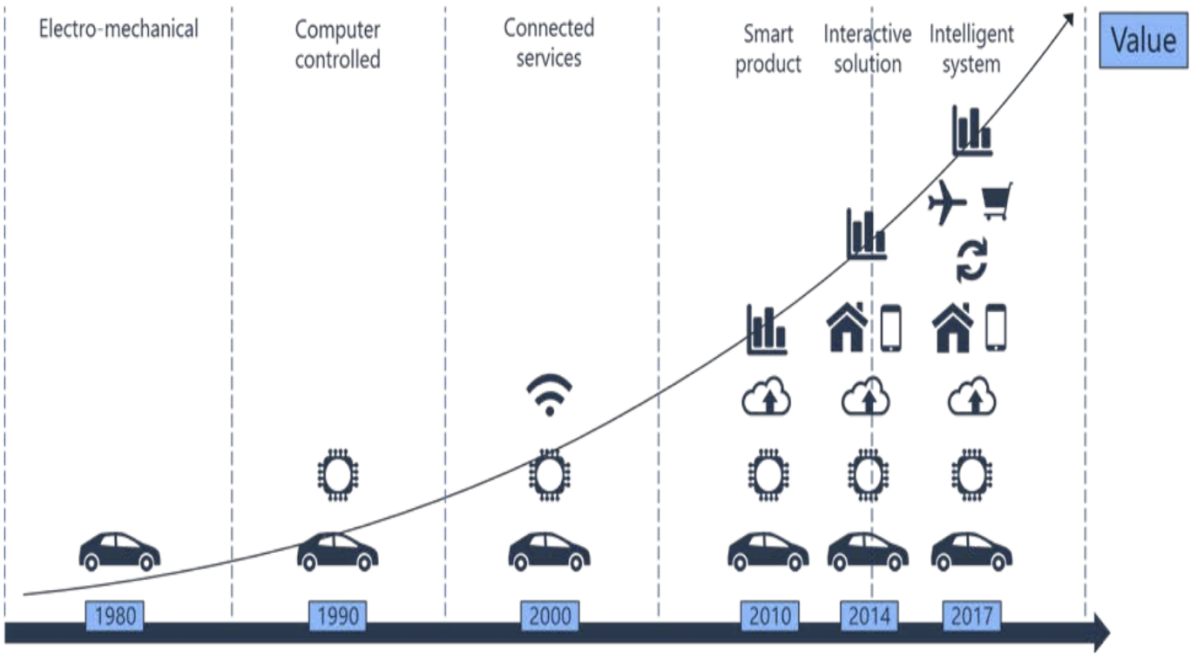


[2] https://blogs.microsoft.com/iot/2018/09/24/announcing-azure-digital-twins-create-digital-replicas-of-spaces-and-infrastructure-using-cloud-ai-and-iot/

IV. What strategies should IT firms adopt to develop Digital Twins?

Information technology (IT) firms adapting to the right developing technologies always add value to their business market. Hence, IT firms adapting to digital twin technology have added value by providing an excellent service platform for end users to optimize business operations.

Figure-4 IT firm technology adoption[3]



In the above picture, it can be seen that with the evolution of advanced technologies from year to year, IT firms that adapt to these advancements have added value to their businesses exponentially. One of the main technologies IT firms adapted for optimization of products or services to the fullest is the provision of cheap cloud infrastructure and cheap IoT devices. These devices can be connected to literally any physical system to collect data passively and transferred to its digital twin which can be simulated and processed in cloud/cyberspace. This provides a vast scope for the IT industry to take advantage of this technology. Organizations have entered the era of cyberspace with support from the IT firms which provide a platform

[3] https://info.microsoft.com/rs/157-GQE-382/images/Microsoft%27s%20Digital%20Twin%20%27How-To%27%20Whitepaper.pdf

and tools to implement and utilize digital twins' capabilities. A detailed process of strategies that IT firms should adopt to develop digital twin is explained below.

IT firms initially need to test the digital twin technology platform they need to capture adequate amount of data for creating the digital replica of the physical entity. The data is captured from IoT devices and is stored in either the cloud or in cheap commodity machines such as Hadoop clusters (HDFS) for better scaling and accessibility.

Strategies that IT firm should adopt are as follows:

- **Define clear objectives:**

  The goal of what the firm is trying to achieve should be clear throughout the organization and it should be aligned with the organizational strategies and goals.

- **Define a quick win use case:**

  Building a small quick win use case with limited functionality to start with is very essential for acquiring funding and demonstrating immediate results. Selecting a use case with narrow scope after experimenting and exploring the opportunities available will help in evaluating the suitability of technology and getting better insights.

- **Finding the right expertise:**

  Understanding the value of data and digitization is not enough. A fundamental step is to find the right expertise needed to implement digitization.

- **Find the right partner:**

  The company should look for the right partner to provide better assistance, training, and consultations in improving the implementations of its open source solutions. There are lots of different platforms provided to store data, as well as a lot of analytical tools and techniques available to perform analysis. Not all solutions can be provided by a single tool or technique. Proper emphasis should be given to the choice of right tools, techniques and

partnerships to solve integration issues that may arise. Finding the right partner and building a perfect ecosystem is of utmost importance.

- **Good data governance:**

  The company should provide appropriate data architecture, metadata, data quality assurance methods, and proper data stewardship. Organizations should ensure that they adhere to the policies and acts related to data privacy and protection in their respective countries. They must also ensure that legal contractual agreements for appropriate data integration is employed within the organization.

- **Developing the toolkit:**

  Development is one of the most critical and challenging stages. Hence, highly specialized data scientists, research and development teams should be involved. They must have deep understanding and knowledge on statistics and machine learning technology. At this stage, three-dimensional (3D) machine models are developed, with the addition of mechanical processes and functionalities required, which are used to generate unique digital twins for different assets. This process includes gathering real time information of the physical assets' conditions and operating environment. Data collected with the help of sensors and radio-frequency identification (RFID) are used to develop machine learning algorithms for optimization of system operations. One of the strategies is moving performance of all these activities into the cloud platform as it provides the best infrastructure for saving, transforming, and analyzing big data. Various analytical tools can be used to perform simulations in a particular scenario. Outputs of the simulations could help determine an asset's behavior in a particular setup or environment, which in turn can help in improving its performance and reducing risks associated with that asset's operations.

- **Generating a digital twin:**

  After the integration of the tools developed in the previous stage, one can create a unique digital twin for each physical asset. At this stage specialized statisticians and developers

make sure that the data collected, and the models created to simulate the reality are of best quality. Data collected from the IoT devices and sensors must be sufficient enough to accurately replicate the behavior of the physical entity.

- **Operating the digital twin:**

  Once the initial version of digital twin is created, it can be deployed and run in production. Digital twin, with the help of sensors, are enriched with data from the physical machine, augmenting the raw data with insights provided by modeling. Operating the digital twin in production is an integral part of the data processing chain in analytical systems. All the machines in the industry needs to be monitored continuously in real time to identify its current operations, performance, and its inefficiencies so that necessary corrective and preventive measures can be implemented, which helps in reduction of cost and time, better operation, improved decision making, and increase in revenue of the organization. More sophisticated digital twins can also learn to self-adjust the gradual subtle changes in the physical asset.

- **Taking the digital twin to the cloud:**

  The amount of digital content being generated, transformed and shared with the twin in today's world is astounding, as we use live data to view and analyze the system performances. Due to the scale and velocity of data that is transferred between the physical asset and virtual twin, we need to have cheaper data storage system such as the cloud so that multiple vendors across all industries can use it with reduced cost and without the need to create separate data warehouses to store, transform, process, and analyze the captured data. Storing the data in cloud provides more sophisticated techniques to access high volume and velocity of data at a low cost and enables all the industries to avail this digital twin technology.

- **Connecting the digital twin to mobile apps:**

In this digital world, humans are eager to know results as and when they are generated since it helps them to improve their decision making process, optimize operations and increase the possibility of achieving greater and quicker solutions. Digital twins can be connected to mobile applications to monitor the data regularly and frequently, which leads to faster decisions with greater accuracy due to the precision, velocity, and scale of the data available. Mobile is one of the most used and important platforms in this era and hence, launching/connecting with mobile devices will earn more customers for the digital twin technology.

Microsoft Azure uses Power BI and machine learning tools to produce digital twins of smart hotels called Hotel 360^0. These twins are connected to the customers' mobiles, which aids them in viewing their room, floor and other details without being in the hotel. Moreover, customers can adjust their preferred room temperature and lighting levels to appropriate values using the mobile application provided. This is one of the strategies to attract customers by providing them with an innovative and easy to use mobile application.

- **Focus on measurement:**

  Precision of the digital twin should be measured in order to collect feedback and use it for further optimization of operations. It is the key factor in determining the accuracy, usefulness and value created for the organization through the use of digital twins.

- **Communication integration:**

  Digital twins nowadays are integrated with richness in voice and vision making human life better and simpler. Humans doesn't have to struggle interpreting data from graphs. Instead machines are programmed to communicate information about their status such as their operations, performance, and a lot more. Ex: GE has developed a digital twin of a wind turbine that can provide the current status of its performance and also provides suggestions for improving and optimizing its operation through the use of a voice. It also contains voice recognition and analysis to understand instructions spoken by the user.

V. What strategies should business adopt to take advantage of Digital Twin?

Digital twins technology helps businesses improve their customer experience by providing a better understanding of customer needs, promoting innovations, extending the life of assets, providing end-to-end visibility of products in real time, uncovering operational inefficiencies, deploying preventive maintenance, reducing maintenance costs, and providing a host of other benefits.

Implementing digital twins used to be expensive but it is now cost-effective due to the development of the IoT which reduced the storage costs and processing power associated with the use of digital twins technology. The sensors widely used in IoT devices for collecting real time passive data are very cheap and readily available.

Hence, utilizing digital twins is now a necessity for most organizations to remain competitive in their markets. This is because competing companies are either already implementing this technology or planning on doing so in the very near future due to the low cost and amazing benefits provided by digital twins.

Gartner predicts that organizations will first simply implement digital twins, and then evolve them over time by improving their capabilities with regards to data collection and visualization, analytics and so forth with the ultimate goal being to respond to business objectives. However, this should not be the case, and companies should follow a proper strategy to fully take advantage of this technology.

Figure-5 Business Strategies to adopt Digital Twin

The important points to consider while building a digital twin strategy are:

- **Keep it simple:** The technology leaders of a company should properly know their business goals in order to clearly define the objectives of their digital twins strategy. Using this knowledge, build digital twins of the complexity needed to meet these objectives. Do not develop overly complex twins that do not provide the required insights, or which provides unnecessary extra information. Thus, to put it in the words of project management, avoid the problem of scope creep during the project of developing digital twins.
- **Readiness assessment:** Study the company's businesses thoroughly to ensure that the IoT network present here is ready to embrace digital twins technology. And if it is ready, determine where best to begin implementing this technology.

- **Developing the right expertise:** A company should make sure that it has the expertise required to properly implement and maintain the digital twins technology either inhouse or finding the right external partner to provide this expertise.
- **Small quick win use case:** Before a large-scale implementation of digital twins technology, a small use case should be presented to show the feasibility and benefits of this technology in order to garner more support and funding for its implementation.
- **Focus on core objectives:** There are three types of strategies based on McKinsey's three horizons of growth and digital twins should be used first and foremost in horizon 1, which is maintaining and defending the core business. Once this has proved to be successfully implementable, it can be used, if needed, in horizon 2, which is nurturing emerging businesses, and finally, if required, it can be used to create genuinely new businesses once the company has found that it can efficiently implement and maintain digital twins in horizon 2.
- **Concentrate on measurements:** The digital twins strategy should be focused on generating value and tangible return on interest (ROI). Key performance indicators (KPIs) should also be developed to measure the progress of the digital twins initiatives undertaken in the company, and to measure the return on interest being obtained.
- **Integrate a single digital twin:** A company first needs to implement a single digital twin before moving on to more complex twins, if needed. When a single twin is implemented, it will be associated with various Internet of Things devices and hence, several endpoints should be properly defined here. For example, a digital twin for manufacturing equipment should have the following endpoints:
    i. Sensors that help gather information about the equipment in real-time.
    ii. An application that tracks the status of the equipment, such as one's laptop or smartphone.

iii. Suitable technologies that ensure that data gathering and sharing follow regulatory compliances.

- **Incorporate compound digital twins:** Compound digital twins are implemented in a company that requires them to meet complex business objectives that cannot be met through the use of single digital twins alone. Single digital twins should first be implemented to form the basis for building compound digital twins. A compound digital twin has several layers, each of which addresses the different business needs of different parts of the company (Ex: customer layer, manufacturer layer, etc.). Leverage advanced technologies like artificial intelligence to help monitor the connected complex Internet of Things devices well. Select the best manufacturers who can provide products which when connected with sensors will best fit the compound digital twins. Also ensure that a technical support team is present that can analyze the complex data gathered for and from the twins.
- **Personalization/ Customization:** To improve customers' experiences and satisfaction, one can implement digital twins that produce products based on the preferences of each individual. Such customization of solutions attracts more customers and hence, increases the customer base which in turn improves the business. Customers may actively or passively need to participate in their data capture based on the business objective. Ex: Precision surgery in the healthcare industry can be aided by the use of digital twins of the patients. The digital twin makes use of an individual's DNA structure and electronic health records information to create a digital replica of that person which contains their unique personal characteristics. This replica is used to analyze the various options available for performing a particular operation to treat an ailment on that patient and selecting the optimal method based on the results obtained from these simulations. Advanced technologies like virtual reality can also be used to more easily understand the simulations of these surgeries on the digital twin.

- **Address ethical and privacy concerns:** To obtain the highest value from digital twins, a company must openly address any privacy and ethical issues brought up by its customers, employees, partners, or any other parties affected by the digital twins strategy implemented. The company should provide transparency in its data capture, utilization, and sharing techniques. It should identify potential areas where its customers or its own data could drive value but could also be at risk and should provide adequately reasonable solutions to these problems.

The above strategy is just a generalization and to truly maximize the value offered from digital twins, chief information officers (CIOs) need to work with business leaders to develop specific economic and business models that consider the benefits in comparison to the development and maintenance costs of these digital twins.

REFERENCES:

1. Marr, Forbes (2017, March 06 Published). Article title. Retrieved from URL. https://www.forbes.com/sites/bernardmarr/2017/03/06/what-is-digital-twin-technology-and-why-is-it-so-important/#471aaf8b2e2a

2. Herald Keeper, Market Watch (2018, September 05 Published). Article title. Retrieved from URL https://www.marketwatch.com/press-release/digital-twin-market-2018-recent-trends-industry-analysis-opportunities-size-segmentation-competitive-analysis-and-forecasts-2018-09-05?ns=prod/accounts-mw&ns=prod/accounts-mw

3. Hoof, Microsoft Internet of Things (2018, September 24 Published). Article title. Retrieved from URL https://blogs.microsoft.com/iot/2018/09/24/announcing-azure-digital-twins-create-digital-replicas-of-spaces-and-infrastructure-using-cloud-ai-and-iot/

4. Seto, Gerard, Thurman Kshirasagar, Microsoft Azure (2018, December 13 Published). Article title. Retrieved from URL https://docs.microsoft.com/en-us/azure/digital-twins/about-digital-twins

5. Eniram, How do you build a digital twin (2018, August 27 Published). Article title. Retrieved from URL https://www.eniram.fi/how-do-you-build-a-digital-twin/

6. GE Channel. (2016, November 16). *Minds+ Machines: Meet A Digital Twin* [Video file]. Retrieved from https://www.youtube.com/watch?v=2dCz3oL2rTw

7. Philips Channel. (2018, April 20). *Philips Digital Twin Concept* [Video file]. Retrieved from https://www.youtube.com/watch?v=H6JzPCbyVSM

8. Microsoft Developer Channel. (2018, November 29). *Smart Hotel 360, a demo powered by Azure Digital Twins* [Video file] Retrieved from https://www.youtube.com/watch?v=urcmaFVQnF4

GLOSSARY:

Cyber-physical system: It is a mechanism that is controlled or monitored by computer-based algorithms, tightly integrated with the Internet and its users.

Machine learning: The scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence.

Artificial intelligence: The intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals.

Information technology: The use of computers to store, retrieve, transmit, and manipulate data, or information, often in the context of a business or other enterprise.

Business intelligence (BI): Comprises the strategies and technologies used by enterprises for the data analysis of business information. BI technologies provide historical, current and predictive views of business operations.

Big data analytics: The often-complex process of examining large and varied data sets (big data) to uncover information including hidden patterns, unknown correlations, market trends and customer preferences that can help organizations make informed business decisions.

Apache Hadoop: A collection of open-source software utilities that facilitate using a network of many computers to solve problems involving massive amounts of data and computation. It provides a software framework for distributed storage and processing of big data using the MapReduce programming model.

Apache Kafka: An open-source stream-processing software platform developed by LinkedIn and donated to the Apache Software Foundation, written in Scala and Java. The project aims to provide a unified, high-throughput, low-latency platform for handling real-time data feeds.

Apache NiFi: Software project from the Apache Software Foundation designed to automate the flow of data between software systems. It is based on the "NiagaraFiles" software previously developed by the NSA, which is also the source of a part of its present name – NiFi.

Apache Flume: A distributed, reliable, and available software for efficiently collecting, aggregating, and moving large amounts of log data. It has a simple and flexible architecture based on streaming data flows.

Apache Spark: An open-source distributed general-purpose cluster-computing framework.

Computer cluster: A set of loosely or tightly connected computers that work together so that, in many respects, they can be viewed as a single system. Unlike grid computers, computer clusters have each node set to perform the same task, controlled and scheduled by software.

Blockchain: A growing list of records, called blocks, which are linked using cryptography. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data. By design, a blockchain is resistant to modification of the data.

Tableau Software: Provides software applications for fast analytical and rapid-fire business intelligence. Tableau Desktop is a data visualization application that lets you analyze virtually any type of structured data and produce highly interactive, beautiful graphs, dashboards, and reports in just minutes.

Power BI: A business analytics service by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards.

Qlik Sense: A BI and visual analytics platform that supports a range of use cases, including centrally deployed guided analytics apps and dashboards, custom and embedded analytics, and self-service visualization, all within a scalable, governed framework.

OLAP (Online Analytical Processing): The technology behind many BI applications. OLAP is a powerful technology for data discovery, including capabilities for limitless report viewing, complex analytical calculations, and predictive "what if" scenario (budget, forecast) planning.

Google Analytics: A premium web analytics service offered by Google that tracks and reports website traffic, currently as a platform inside the Google Marketing Platform brand.

Cloud computing: It makes computer system resources, especially storage and computing power, available on demand without direct active management by the user. The term is generally used to describe data centres available to many users over the Internet.

Business analytics: It refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning.

Predictive analytics: Encompasses a variety of statistical techniques from data mining, predictive modelling, and machine learning, that analyze current and historical facts to make predictions about future or otherwise unknown events.

Cyberspace: It is the notional environment in which communication over computer networks occurs.