Digital Twin Architecture and Standards

Authors:

K. Eric Harper
Senior Principal Scientist
ABB Corporate Research, US
eric.e.harper@us.abb.com

Dr. Christopher Ganz
VP Digital Research and Development
ABB
christopher.ganz@ch.abb.com

Dr. Somayeh Malakuti
Senior Scientist
ABB Corporate Research, Germany
somayeh.malakuti@de.abb.com

INTRODUCTION

Digital Twins are key components in an Industrial IoT (Internet of Things) ecosystem, owned and managed by business stakeholders to provide secure storage, processing and sharing of data within an architectural tier. Industrial IoT is an integration exercise rather than a development challenge, bringing many vendors and technologies together. Digital twins enable flexible configurations of applications and data storage, especially to integrate third parties. An architecture based on digital twins is one alternative for managing this complexity.

We propose six sets of operations to characterize digital twin interactions within the Industrial IoT ecosystem:

1. Digital twins are discoverable, can be queried to determine their capabilities and composed to provide industrial solutions.
2. An information model abstracts a digital twin, with discoverable object types that can be browsed by other components and interactively, supporting underlying data repositories that evolve according to real world lifecycles.
3. Key-value pairs are created, read, updated and deleted in column stores with possible configured side effects that can modify or enhance the value contents. Data source ingest is performed using create operations and application access is performed using read operations.
4. Applications within an ecosystem tier subscribe to notification events published when digital twin transactions occur, triggering actions to retrieve and process the affected content.
5. Digital twin contents are securely synchronized in bulk between connected tiers, using the network bandwidth to its best advantage to consolidate related content in centralized storage without losing ownership.
6. Authenticated users are authorized by the owner to configure and manage the digital twin properties using a separate set of operations.

An integrated information model, separate from those representing each digital twin, forms the basis for all interactions, including design, orchestration, execution and administration.

DIGITAL TWIN CAPABILITIES

The Digital Twin concept first appeared for industry in 2003. The meaning of the term has evolved, and this powerful metaphor can be extended to include a comprehensive set of possible capabilities, as shown in Table 1. These capabilities create value throughout the lifecycle of industrial assets, as shown in Table 2.

Feature	Functionality
Document management	All documents (drawings, instructions, etc.) associated to equipment throughout its lifecycle
Model	Digital representation of the equipment that can mimic properties and behaviors of a physical device
3D representation	Properties of a physical device (measured or simulated) mapped to a 3D digital representation
Simulation	Representation of a physical device in a simulation environment to study its behavior
Data model	Standardized data model for connectivity, analytics, and/or visualization
Visualization	Graphical representation of the object either on a supervisory screen or personal device
Model synchronization	Alignment of a model with real world parameters (potentially in real-time)
Connected analytics	Algorithms and computational results based on measured properties of a physical device

Table 1: Digital Twin Features

	Plan	Build	Operate	Maintain
Document management	PLM	PLM	Operation instructions	Service record
Model	Physical properties predict		Optimization	Diagnostics
Simulation	Design simulation	Virtual commissioning		
3D representation	Design drawings	Manufacturing instructions		Service instructions
Data model	Engineering data	Production data	Operational data	Service data
Visualization			Operational state display	health status display
Model synchronization			Real-time movement	Model inversion
Connected analytics			Operational KPIs	Asset health KPIs

Table 2: Digital Twin Features and Use Cases

New industrial assets can be designed using simulation tools and physical models to precisely predict behavior. Physical properties (electromagnetic, thermal, pressure, stress, etc.) are mapped to the design model to optimize the device's performance. This approach requires knowledge of the environment and its effects.

Digital twins are composable, where components interact with each other in the physical world. In discrete processes, components are reasonably decoupled which allows the combination of separate behavioral simulators to build a larger system. Components interact and influence each other in continuous processes. Equipment needs to be modeled in one common simulation tool with a standardized model format.

MOTIVATION FOR DIGITAL TWIN

Digital twins combine data and processing. The necessary data capabilities for Industrial IoT processing are provided in four consecutive phases: data generation, data acquisition, data storage and data consumption. [1] Data also flows in the opposite direction for set point control to the production process, optimization re-calibration and customization directives for specific deliverables.

A heterogeneous ecosystem for processing comes into play in all these phases and data flows. Process measurement is associated with its equipment type, converted to engineering units and validated for accuracy. Data is acquired using many different protocols and temporary repositories. Each component vendor has their own (legacy, hosted) platform for historical data and applications—for example, analysis that interprets the measurements without exposing proprietary algorithms. These results guide business decisions and continuous process improvement.

The keys to success for Industrial IoT are to create value for end users and find business models that allow various ecosystem players to co-exist and successfully co-evolve. [2] Distributed data stores and analytics are essential components that make this ecosystem possible. One example is shown in Figure 1, including use of a Distributed Control System (DCS). Industrial IoT can be organized in tiers or layers, with each layer able to operate autonomously based on the available data and services.

[1] Hu, H., Wen, Y., Chua, T.S. and Li, X. 2014. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. In Access, IEEE, vol.2, no., pp.652-687, DOI= http://dx.doi.org/10.1109/ACCESS.2014.2332453.

[2] Toivanen, T., Mazhelis, O. and Luoma, E. 2015. Network Analysis of Platform Ecosystems: The Case of Internet of Things Ecosystem. Software Business, DOI= http://dx.doi.org/10.1007/978-3-319-19593-3_3.

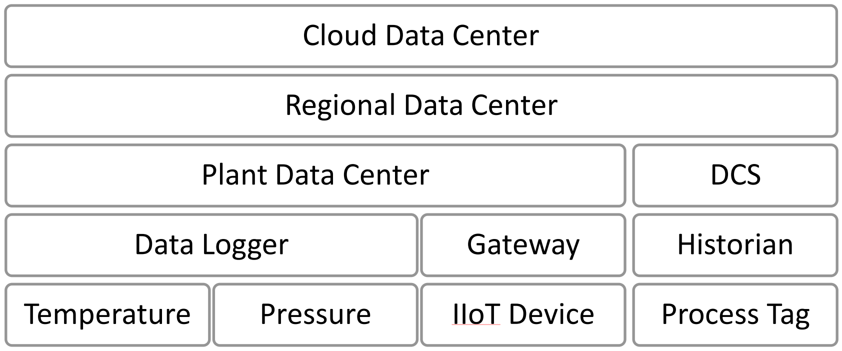

Figure 1: Industrial IoT Tiers

Communication between layers are interactions between architectural components, where some if not all the elements are digital twins. Digital twin interoperability standards could be used instead of proprietary protocols to reduce the complexity and cost of integrating different vendor solutions together.

There is limited scope of data in the lower layers and the co-located services have shortened latencies when interacting with industrial processes. In the supporting layers there can be multiple data centers, one for each vendor, and regional tiers may be required due to country-specific regulations for data sharing cloud-to-cloud. Plant tiers occur naturally from legacy operational technology deployments, and device tiers arise as embedded computers expand their storage capacity and processing power.

INDUSTRIAL CHALLENGES

The Industrial IoT market is targeted to grow by trillions of US dollars by the year 2030[3], driven by adoption, deployment and integration of billions of intelligent devices and their associated data. The devices can talk directly to one another when possible and handle much of their own computational tasks. [4] Edge computing provides elastic resources and services, while cloud computing supports workflows distributed in the production network.[5] This digital expansion faces several significant challenges, including reliable data management, security and privacy.

Aggregating all the raw data to a single data center before performing analysis increases response times, raising performance concerns in traditional industrial markets and requiring architectural tradeoffs. Low cost sensors and ubiquitous networking are

[3] Purdy, M. Davarzani, L. 2015. The Growth Game-Changer: How the Industrial Internet of Things can drive progress and prosperity. White Paper. Accenture Strategy.

[4] Froehlich, A. 2014. IoT: Out of the Cloud & Into the Fog. Blog Post. Information Week / Network Computing.

[5] Yi, S., Li, C. and Li, Q. 2015. A Survey of Fog Computing: Concepts, Applications and Issues. In Proceedings of the 2015 Workshop on Mobile Big Data (Hangzhou, China, June 22 – 25, 2015). Mobidata ’15. ACM, New York, NY, 37-42. DOI= http://dx.doi.org/10.1145/2757384.2757397.

enabling the next generation of industrial processing and service. Industrial raw measurements are created independent of hosted services, making it challenging to collect and process the inputs. Initial raw process data ownership is controlled by organizations, not individuals.

This increases the complexity of negotiations for who benefits from monetizing the data, especially when industrial activity and intellectual property can be revealed simply by the characteristics and timing of the measurements. Industrial installations can have multiple vendors each with their own data representations and legacy technology stacks. Many of these concerns can be addressed by using digital twins in the ways we propose.

BROWNFIELD PERSPECTIVE

Traditional industry is characterized by plants where the equipment is installed, configured and operated for years, even decades. These legacies cannot be forgotten or discarded but instead need to be integrated with new technologies. Industrial IoT market growth will accelerate only if there is business value for both the consumers and suppliers of products and services. Legacy devices may encounter system security challenges because they are usually deployed in places without rigorous surveillance and protection.[6] Industrial IoT providers must convince existing stakeholders that their intellectual property is safe. This requires a holistic cybersecurity solution that addresses the various security and privacy risks at all abstraction levels,[7]

An industrial process may be orchestrated by a single control system, but the assets performing the work are selected with a best of breed strategy. Process plant design is guided in part by requirements for manufacturing precision and the cost of the individual workflow elements, bringing many different vendors into the solution space. Each asset vendor has unique subject matter expertise for their equipment, making them the best analyst of the related data. Traditionally, analysis is performed only when there is a process issue where temporary service access to the data is allowed close to the site.

Industrial IoT promises to increase scalability for process plant services by reducing the need to be on site. This is made possible by data collection using access from a remote location, potentially transferring the relevant measurements to the cloud. The dominant approach of aggregating all the data to a single datacenter can significantly

---

[6] Stojmenovic, I., Wen, S., Huang, X., and Luan, H. 2015. An overview of Fog computing and its security issues. Concurrency Computat.: Pract. Exper., DOI= http://dx.doi.org/10.1002/cpe.3485.

[7] Sadeghi, A.R., Wachsmann, C. and Waidner, M. 2015. Security and Privacy Challenges in Industrial Internet of Things. In Proceedings of the 52nd Annual Design Automation Conference (San Francisco, June 07 – 11, 2015). DAC '15. ACM, New York, NY, Article 54. DOI= http://dx.doi.org/10.1145/2744769.2747942.

increase the timeliness of analytics.[8] One approach is to establish a compromise between data duplication and the performance cost of update and select queries.[9]

Digital Twin Standards

The International Organization for Standardization (ISO) covers industrial data in TC 184 SC 4.[10] The standard for a Digital Twin Manufacturing Network is currently under development.[11] For the Joint Technical Committee (JTC 1) that includes ISO and the International Electrotechnical Commission (IEC), an established Joint Advisory Group (JAG) on Emerging Technology and Innovation (JETI) published their Technology Trend Report[12] and identified Digital Twin as one of four top emerging technologies out of fifteen. The report has led to formation of the Digital Twin Advisory Group (AG 11) who provide recommendations to JTC 1.

We encourage industry to support these formal activities and to develop inputs for standardization. Our work identifies six architectural interactions for digital twins to support the common operations proposed in the Introduction above, as conceptualized by the central diagram in Figure 2 below.

For example, standards could define Application Programming Interfaces (APIs) for digital twin data access to securely and reliably store, manage and retrieve records. The digital twin architecture context delineates a security domain to control access and restrict operations to authorized clients. Clients must authenticate using security best practices, perhaps facilitated by federated identity. Authorized clients exchange key-value pairs with a digital twin using CRUD (Create, Read, Update and Delete). Values can be simple or structured (object) types. Some implementations may restrict updates and deletes to support data consistency goals.

---

[8] Pu, Q., Ananthanarayanan, G., Bodik, P., Kandula, S., Akella, A., Bahl, P. and Stoica, 2015. Low Latency Geo-Distributed Data Analytics. In Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (London, August 17 – 21, 2015). SIGCOMM '15. ACM, New York, NY, 421-434. DOI= http://dx.doi.org/10.1145/2785956.2787505.

[9] Hassen, F. and Touzi, A.G. 2015. Towards a New Architecture for the Description and Manipulation of Large Distributed Data. Big Data in Complex Systems, DOI= http://dx.doi.org/10.1007/978-3-319-11056-1_17.

[10] Industrial Data: https://www.iso.org/committee/54158.html

[11] Digital Twin Manufacturing Framework: https://www.iso.org/standard/75066.html

[12] JETI: https://jtc1info.org/technology/jeti/

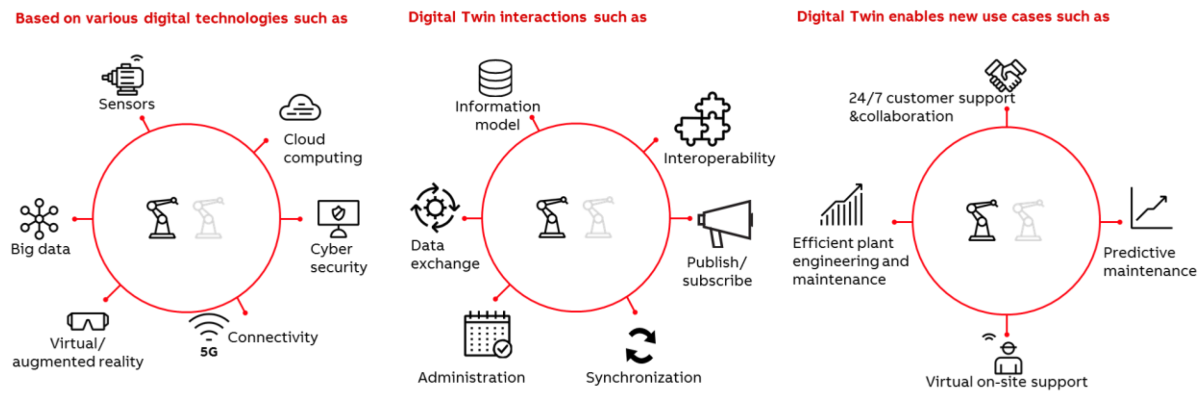

Figure 2: Digital Twin Context Diagrams[13]

The interaction APIs are realized with appropriate technology available in the tier. Digital twin implementations might be deployed using an app store like those for mobile computing. The app store content is replicated in each Industrial IoT tier and enables direct access for third party participation in the common ecosystem. App store transactions in a disconnected tier are journaled and replicated to other tiers when communication is re-established.

Each digital twin deployment can have a different information model allowing for diversity in data representation and relationships. This parallels the trend in microservices where every service has a unique set of programming interfaces, and applications must know how to use them. In a similar way, the digital twin information model API enables discovery and classification of types, properties and instances.

Digital twins connect to applications and to each other. To address conditions where connectivity is not reliable, configured digital twin duplicates are deployed in adjacent Industrial IoT tiers and bi-directionally synchronized according to filter criteria defined by the digital twin owner. A Digital twin instance takes on the same policies configured in the app store regardless of which tier the replica resides in. The data content in a replica may become stale over time until the next synchronization but still provide reliable, error-free data access for local applications.

Each digital twin serves as a publish and subscribe hub in its tier, enabling event driven application development using the Model-View-Controller (MVC) pattern for services. Any data exchange operation on the digital twin generates a corresponding notification published to all subscribed clients for that event. Subscribers can use these events to exercise digital twin CRUD operations based on the metadata content of an event.

[13] Malakuti, S., Ganz, C., Schlake, J., Harper, K.E., Digital Twin: An Enabler of New Business Models, Automation (2019).

The digital twin owner controls the data contents and access to functions using administrative operations. The properties and configuration are declaratively specified. Column stores can be created and deleted. The digital twin contents can be encrypted with the owner's certificate. Each column store in a digital twin can be connected to an ingest data source, subscribing and automatically creating records as new readings are published by the data source or by polling the ingest source periodic basis.

Digital twin clients are provisioned and assigned to roles associated with the different interfaces, column stores, ranges of data and policies for access. Programmatic callbacks are registered for fine grained filtering of ingested, exchanged and synchronized values. Finally, the interoperability API makes it possible for a digital twin to register with the ecosystem and expose its characteristics for access to the other APIs.

ARCHITECTURAL EVALUATION CRITERIA

Our vision is that digital twins can be deployed in any Industrial IoT tier, realized with the available technology choices, and synchronization between digital twins is the only communication between tiers. Data replicated into a digital twin looks like ingest and triggers the associated published notification events. The following expectations summarize the digital twin architectural capabilities and their motivations.

C1. App store deployment of configuration. Digital twin information model and policy definitions are deployed independent of services as first class participants for Industrial IoT. This provides a separation of concerns between data and service ownership and enables declarative integration of applications, services and digital twins.

C2. Integrated information model. Asset types and instances are crucial aspects of the ecosystem: discoverable, navigable and organized independent of naming conventions. Classification of types apply to related instances and property values. Multiple information models can be federated within a tier to provide a broad view of the available storage.

C3. Flexible classification of types, properties and instances. Every digital twin can invent its own type system, imposing the constraint on clients to configure and program accordingly. No different than the complexity introduced by microservice APIs, it is unrealistic that all Industrial IoT applications will agree on a common information model taxonomy and attributes.

C4. Encrypted data at rest and in transfer. Digital twins can store encrypted data, i.e. only readable with guaranteed integrity by the provisioned users. Encryption is used for sensitive API parameters to protect privacy and reduce the possibility for malicious control.

C5. Role-based access control configured for authenticated users. A digital twin imposes a security domain to protect and manage access to data. Digital twin owners define (select) the EULA (End User License

Agreement) policies by which sharing is allowed, protecting intellectual property and sensitive information. Synchronized replicas in adjacent tiers are guarded by the same controls.

C6. Data ingest configuration for each column store. The Industrial IoT life blood is collections of data, including historical records that can be replayed as streams. Digital twins are populated by creating key-value pairs, and data source ingest is a ubiquitous scenario.

C7. CRUD data exchange with cascading side effects based on role. Writing and reading key-value pairs in digital twin column stores is the fundamental application programming model for persistence and analysis. These fine-grained transactions within a tier can be extended with programmatic callbacks configured by the repository owner, providing maximum control over the content.

C8. Publish and subscribe notification of CRUD transactions. Digital twins enable clusters of processing activity in a tier by generating events associated with repository access. The events are not intended to directly share content. Instead, applications use the notifications to drive key-value pair access, like the MVC pattern.

C9. Filtered synchronization between tiers. Industrial IoT data is created at the network edge yet delivers the best business value when aggregated in the cloud. Communications between the edge and cloud may not be reliable or intentionally air-gapped for security protection. Selected column store synchronization uses the network bandwidth for transferring data in bulk and reduces the cybersecurity attack surface of a tier.

Table 3 shows the cross references between the six proposed digital twin interactions and the nine evaluation criteria.

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**
Interoperability	■	■	■	■	■	■	■	■	■
Information Model		■	■						
Data Exchange				■	■	■	■		
Administration	■			■	■				
Synchronization									■
Publish / Subscribe							■	■	

Table 3: Evaluation Criteria Applied to Digital Twin Interactions

Lessons Learned

The number one challenge facing architects who are designing and building Industrial IoT is the interchange of data between the mechanical, digital and human components of an industrial process.[14] There are ongoing standards efforts at each of these levels, but competitive pressures and market demands are more powerful factors in determining the components and their interactions.

There are many choices for Industrial IoT communication protocols, middleware, infrastructure, services and application frameworks. Successful interoperability will come not from waiting for all the ecosystem participants to agree on a set of standards, but by providing adapters that facilitate data exchange between different systems. Digital twin abstractions can serve as these adapters.

The first lesson learned is to embrace the overall complexity rather than hope to avoid it. Commonality occurs naturally in clusters, not from imposed governance. Within those clusters, built to purpose, the APIs and information models evolve to meet specific objectives. Over time the applications and markets will determine which standards bring business value.

The second lesson learned, also related to complexity, is that requirements are best developed incrementally and iteratively starting with business objectives and concerns. Presenting a detailed list of candidate concerns and scenarios to stakeholders creates confusion because there may be less interest in some of the topics. It is more effective to create topic headlines that are subject to interpretation and allow interviewees to shape their meaning and provide additional alternatives. Once the list is narrowed by the audience, it is the right time to drill deeper for clarification.

The third lesson is that stakeholders rarely have crisp ideas of what they want. Initial motivations come from their existing and potential customers and what is perceived to be provided by competitors. True innovation comes from presenting alternative visions and future scenarios independent of technology choices, inviting feedback and revision. This process brings out the key stakeholder concerns and helps communicate what is possible.

Conclusions and Future Work

The top three benefits from the architecture recommendations for our digital twin approach are to:

1. Reduce API (but not information model) complexity,
2. Enhance privacy and security, and
3. Manage connectivity.

Digital twin capabilities provide a common facility for data persistence and application notification in the Industrial IoT ecosystem, standardizing how data is managed and distributed. Applications connect to digital

[14] Boughton, P. 2015. Blog Post. The challenges of creating the Industrial Internet of Things. ENGINEERLIVE. URL= http://www.engineerlive.com/content/challenges-creating-industrial-internet-things.

twins only within a single tier, reducing the cybersecurity attack surface. Digital twins empower and delegate responsibility to data owners to protect, manage and monetize their intellectual property. Finally, given potentially unreliable communications between tiers, digital twin integration adapts to and maximizes opportunities for sharing when synchronization channels are open between tiers.

As ecosystem tiers are increasingly and continuously connected, the synchronization-only communication constraint between tiers could be relaxed to allow CRUD access to digital twins located in federated namespaces across the tiers. Digital twins could be used to distribute content between tiers—for example, application configurations, algorithm specifications or even the manifests and content for synchronized app stores. Finally, submitting work requests to digital twins in the local tier could synchronize with other tiers, triggering activities in adjacent tiers whose results are collected and synchronized back to the requesting tier.

➢ Return to IIC Journal of Innovation landing page for more articles and past editions