

¿Qué es un Gemelo Digital? Una Propuesta de Arquitectura para su Implementación^{*}

Paula Muñoz, Antonio Vallecillo

ITIS Software. Universidad de Málaga
{paulam,av}@uma.es

Resumen En los últimos años, los gemelos digitales se han presentado como un paradigma innovador para mejorar el funcionamiento de sistemas ciberfísicos, software o procesos. El gemelo digital es una réplica virtual del sistema cuyo funcionamiento pretende optimizarse, ya sea mejorando su rendimiento, o permitiendo su mantenimiento preventivo y la detección de anomalías. Aunque existe cierto consenso en la comunidad sobre lo que son los gemelos digitales, aún hay cuestiones no resueltas sobre su definición, arquitectura y conceptos relacionados, como por ejemplo las sombras digitales. Este artículo propone una arquitectura para la implementación de gemelos digitales, y la compara con otras existentes en la literatura con el objetivo de alcanzar un entendimiento común.

Keywords: Gemelos Digitales · Arquitectura · Data Lake

1. Introducción

Los Gemelos Digitales, en inglés *Digital Twins* (DT), han irrumpido como un mecanismo muy efectivo para la optimización del funcionamiento de cualquier tipo de sistema ciberfísico, proceso o servicio software [4,1]. El concepto se basa en la *replicación* de un sistema existente. Sus primeras aplicaciones se remontan a la misión Apolo de la NASA [5], donde los ingenieros crearon réplicas de los sistemas enviados al espacio con el objetivo de poder ayudar a los astronautas desde tierra, reproduciendo su comportamiento. Así, los ingenieros en tierra podrían evaluar las acciones más óptimas a partir de las réplicas, sin poner en riesgo a los astronautas y manteniendo la integridad del sistema en el espacio.

En las primeras aplicaciones, las réplicas eran físicas y la sincronización era manual. Sin embargo, una vez que los científicos de la NASA comprendieron el valor de estas técnicas, las réplicas se digitalizaron convirtiéndose en simulaciones y cálculos diferenciales, y la sincronización se automatizó. En nuestro trabajo, nos referiremos a este último tipo de sistema como DT. Según el Digital Twin Consortium [4], un DT es “una réplica digital de un sistema que se sincroniza con cierta frecuencia y con un determinado nivel de fidelidad”. En la literatura, el sistema replicado se conoce como *Gemelo Físico* (*Physical Twin*, PT) [11,8], u Objeto Observable según la norma ISO [7].

En este artículo se propone una arquitectura desacoplada que introduce el concepto de *Sistema de Gemelo Digital*, o *Digital Twin System* (DTS). Este

^{*} Trabajo subvencionado por los proyectos PID2021-125527NB-I00 (IPSCA) y TED2021-130523B-I00 (SoCUS)

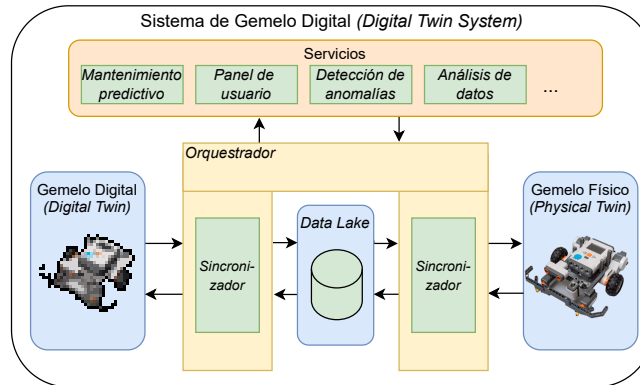


Figura 1. Arquitectura del sistema de gemelo digital.

sistema incluye no sólo la réplica o los servicios que esta presta, sino también los elementos de conexión entre los dos gemelos. Las réplicas digitales (DT) pueden ser de muy distintas naturalezas: algoritmos, modelos de simulación, sistemas de ecuaciones diferenciales, etc. Obsérvese que una simulación o un algoritmo por sí solos no pueden considerarse un DT. Estos elementos sólo serán considerados así dentro del contexto de un DTS, donde exista una sincronización automática y bidireccional con el elemento al que replica y donde se intercambian datos y órdenes para permitir la optimización del funcionamiento del sistema.

La arquitectura propuesta tiene como objetivo optimizar la escalabilidad, interpretación y composición de los DT. Para lograrlo, la arquitectura está centralizada en un *Data Lake* (DL) [6] con un sistema de drivers que orquestan los intercambios entre elementos.

2. Arquitectura

La figura 1 muestra nuestra arquitectura propuesta para la implementación de DTSs. Antes de describirla con detalle, definiremos algunos términos claves:

- **Gemelo Físico (*Physical Twin*) o Sistema Real (*Actual System*).** Es el sistema, servicio o producto cuyo comportamiento deseamos optimizar mediante una réplica digital. Se monitoriza su estado en todo momento y se obtienen periódicamente los valores de sus propiedades de interés. Este sistema puede existir en el momento en el que se crea el DT o no. Esto último puede ocurrir si deseamos emplear el DT para la fase de diseño.
- **Gemelo Digital (*Digital Twin*).** Es la réplica del PT sincronizada con una frecuencia específica y modelada con la fidelidad adecuada para capturar las propiedades de interés del sistema con el nivel de detalle preciso para su propósito. El DT recoge las propiedades de interés del sistema y su comportamiento a partir de datos y modelos, que se actualizan continuamente durante toda la vida del sistema y se almacenan en el *Data Lake*.

A partir de las dos definiciones anteriores, definimos el **Sistema de Gemelo Digital (*Digital Twin System*)** como un sistema construido con un propósito concreto que se compone de:

- El *Gemelo Físico* (PT).
- La réplica virtual, es decir, el *Gemelo Digital* (DT).
- Los mecanismos de *sincronización* entre ambos gemelos, mediante los que se intercambian comandos e información.
- El conjunto de *Servicios* que permiten la explotación de los datos que se transmiten entre los gemelos y que permiten cumplir el objetivo del DTS (por ejemplo, la detección de anomalías).

En nuestra arquitectura, para conseguir el máximo desacople entre los elementos y permitir una mayor escalabilidad y flexibilidad, la conexión entre los elementos se hace a través de un *Data Lake* (DL) que implementa el patrón de diseño Pizarra. Esto permite a todos los elementos intercambiar información de forma asíncrona, escribiendo y leyendo del DL. En caso de que existiera información crítica que requiriese una mínima latencia, los elementos también podrían conectarse directamente a partir del *orquestador*. Esta entidad se usa para intercambiar información directamente entre elementos o para conectarse al data lake. Además, dicho orquestador cuenta con los servicios de sincronización necesarios para garantizar el correcto funcionamiento del sistema.

Una de las principales ventajas de esta arquitectura es que permite la conexión de elementos de forma sencilla, puesto que sólo habría que adaptar la entrada de información al orquestador, sin afectar a ninguno de los demás elementos que ya estuvieran conectados al sistema. Igualmente, si quisiéramos tener más de un gemelo digital interconectado entre sí y que sincronizasen su comportamiento, podría hacerse a través del orquestador. También podríamos tener gemelos digitales a más de un nivel de fidelidad organizados para elegir de cuál usamos sus datos según el estado del gemelo físico o incluso podríamos sustituir el gemelo físico por otro modelo virtual para hacer pruebas sobre el sistema completo. Todo ello se conseguiría únicamente alterando la implementación de la interfaz concreta del elemento para el orquestador.

Esta arquitectura ha sido aplicada con éxito en el desarrollo de dos sistemas concretos: un coche Lego Mindstorms NXT [9] y un brazo robótico Arduino [10].

3. Trabajos y conceptos relacionados

La primera conceptualización de una arquitectura para Gemelos Digitales fue la de tres elementos que originalmente propuso Grieves [5]. En esta arquitectura conceptual aún no se contemplaba el concepto de servicios como un elemento externo al Gemelo Digital o la necesidad de un almacén de datos para el sistema.

Posteriormente encontramos arquitecturas que coinciden con nuestra propuesta en cuanto a los elementos que incluyen, aunque con distinto énfasis o terminología. Por ejemplo, las basadas en cinco elementos [12,11,2] coinciden con nuestra propuesta en sus elementos: una entidad física, una virtual, una base de datos, un conjunto de servicios y las conexiones entre ellos. En [12,11], sin embargo, definen el DT como el conjunto de todos estos elementos y no únicamente la réplica como define el Digital Twin Consortium [4] o nuestro enfoque. En [2], separan al DT del *Data Lake* y de las conexiones, pero incluyen en el interior del DT los servicios para la optimización del comportamiento, sin permitir

la adición de servicios adicionales. Tampoco define el concepto de Gemelo Físico. Finalmente, la arquitectura que define ISO para gemelos digitales [7] es similar a estas últimas propuestas y a la nuestra también, aunque ellos introducen el concepto de *Digital Twin Framework* que se corresponde a nuestro Digital Twin System (DTS) pero sin incluir el sistema físico.

Otro concepto utilizado en la literatura es el de *Sombra Digital (Digital Shadow, DS)*. Este término se define como el conjunto de trazas de datos del sistema y su agregación, recopiladas con un propósito concreto. Estas trazas en nuestro sistema son almacenadas en el *Data Lake* y explotadas por los distintos servicios. La diferencia principal entre un modelo de simulación, el DS y el DT es su grado de sincronización e interacción con el PT. El primero no está sincronizado con el PT ni interactúa con él. El segundo está permanentemente actualizado con los datos que se recogen del PT, pero no envía información al PT. El DT se mantiene sincronizado con el PT e interactúa con él durante su funcionamiento [3].

Para concluir, mencionar que nuestra propuesta hace énfasis en la natural simetría que existe al hablar de gemelos, dos entidades hermanas que son réplica una de otra y que se sincronizan continuamente e intercambian tanto datos como operaciones de control. Sin esa sincronización no se puede hablar de Gemelo Digital. Además, hacemos gran énfasis en el desacople entre los elementos de la arquitectura para permitir la máxima libertad a la hora de componerlos.

Referencias

1. Dalibor, M., Jansen, N., Rumpe, B., Schmalzing, D., Wachtmeister, L., Wimmer, M., Wortmann, A.: A cross-domain systematic mapping study on software engineering for digital twins. *J. Syst. Softw.* **193**, 111361 (2022)
2. Dalibor, M., Michael, J., Rumpe, B., Varga, S., Wortmann, A.: Towards a model-driven architecture for interactive digital twin cockpits. In: Proc. of ER'20. LNCS, vol. 12400, pp. 377–387. Springer (2020)
3. Dalibor, M., et al.: Generating customized low-code development platforms for digital twins. *J. Comput. Lang.* **70**, 101117 (2022)
4. Digital Twin Consortium: Glossary of digital twins. <https://www.digitaltwinconsortium.org/glossary/index.htm> (2021)
5. Grieves, M., Vickers, J.: Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems, pp. 85–113. Springer (2017)
6. Hai, R., Quix, C., Jarke, M.: Data lake concept and systems: a survey. *CoRR abs/2106.09592* (2021)
7. ISO 23247:2021: Automation systems and integration – Digital Twin Framework for manufacturing. International Organization for Standardization (ISO) (Oct 2021)
8. Madni, A.M., Madni, C.C., Lucero, S.D.: Leveraging digital twin technology in model-based systems engineering. *Systems* **7**(1), 7 (2019)
9. Muñoz, P., Troya, J., Vallecillo, A.: Using UML and OCL Models to Realize High-Level Digital Twins. In: Proc. of ModDiT@MODELS'21. pp. 212–220. IEEE (2021)
10. Pérez-Porrás, D., Muñoz, P., Troya, J., Vallecillo, A.: Key-Value vs Graph-based data lakes for realizing Digital Twin systems. In: Proc. of MeSS@STAF'22 (2022)
11. Tao, F., Zhang, H., Liu, A., Nee, A.Y.C.: Digital twin in industry: State-of-the-art. *IEEE Trans. Ind. Informatics* **15**(4), 2405–2415 (2019)
12. Tao, F., Zhang, M., Liu, Y., Nee, A.: Digital twin driven prognostics and health management for complex equipment. *CIRP Annals* **67**(1), 169–172 (2018)