Design and Implementation of Digital Twin Data Engine

Lei Wang[1], Ning Lei[1], Tao Peng[1], Xiaona Zhang[1], Dehui Zhang[2], Pin Li[2*]
[1] Beijing Capital International Airport Co., Ltd., Information Technology Department, Beijing 100621, China
[2] Technology Center, Beijing Brilliant and Technology Co., Ltd., Beijing 100621, China
DOI: 10.32629/aes.v3i4.1069

Abstract: With the advent of the digital twin era, the construction of digital twin engine based on basic twin data plays a very important role in future application scenarios such as intelligent monitoring, dynamic perception, scientific evaluation, intelligent prediction and intelligent decision-making. This paper mainly analyzes the definition of digital twin, the definition of digital twin model, data characteristics and the application scenario of digital twin, and puts forward a design framework and implementation method of digital twin data engine.
***Keywords*:** digital twin model, spatial entity identity code, dynamic updating technology of spatio-temporal data, holographic three-dimensional data, spatio-temporal variation

Optimizing a process, product or service with digital twin technology is achieved based on more complex visual interaction, mechanism simulation, and AI modeling capabilities. Twin big data technology is the key factor to support the implementation of digital twin technology. At present, digital twin data is stored and managed in the form of multi-source independence. Wang Hongtao, Xu Huayan [1][2] et al. applied the temporal database technology in industrial equipment monitoring. Tang Xinming, Wei Haiping [3][4] et al. constructed spatio-temporal databases based on spatial and temporal models to manage spatio-temporal data and changes. Jiang Shaohua, Wang Baohui [5][6] et al. studied BIM-based data storage methods and retrieval mechanisms. Li Jingxin, Liu Changhong [7][8] et al. studied how to solve object and relational graph data storage based on graph data engine. In conclusion, due to the different characteristics of data storage of each data engine, great difficulty in technical crossover and other problems leads to the obstacle in the data storage layer management, and thus data inconsistency and other problems can be aroused. Therefore, it will be an inevitable trend to build a unified multi-source twin big data engine and realize digital twin.

1. Classification of related concepts and models

1.1 Definition of digital twins

A digital twin is a formal digital representation of an asset, process, or system that captures the properties and behavior of the corresponding entity, which can be communicated, stored, interpreted, or processed in a specific sense.

There are three main types of relationships between digital twins: hierarchical, associative, and point-to-point relationship. In order to achieve the deep integration of virtual information space and the actual physical world, the relationship between the twin in the information space and its corresponding physical entity in the physical world remains identical. Hierarchical relationship means the relationship between a single part twin and a composite twin composed of multiple part twins. Associative relationship means the relationship between the digital twin of natural gas pipeline and the twin of natural gas production and equipment[9]. Point-to-point relationship signifies the relationship between multiple wind turbine twins in a power grid. As shown in Figure 1 below:

1.2 Definition of digital twin model

Digital threads is an extensible, configurable, and modular communication framework. This framework can promote the controlled interoperation and fusion of authoritative data, information or knowledge among various data-information-knowledge systems in the digital space, we can construct an integrated view of digital models across time scales or spatial scales covering several stages of the life stage of the entity of interest or several links of the value chain. Furthermore, the unified model is used to drive the life cycle activities of the concerned entities to provide support for decision makers and stakeholders.

Digital twin model is a digital thread to provide time and location in the life cycle, and state information data driven to realize 3D model, digital model, behavioral model, physical model. As shown in Figure 2 below:

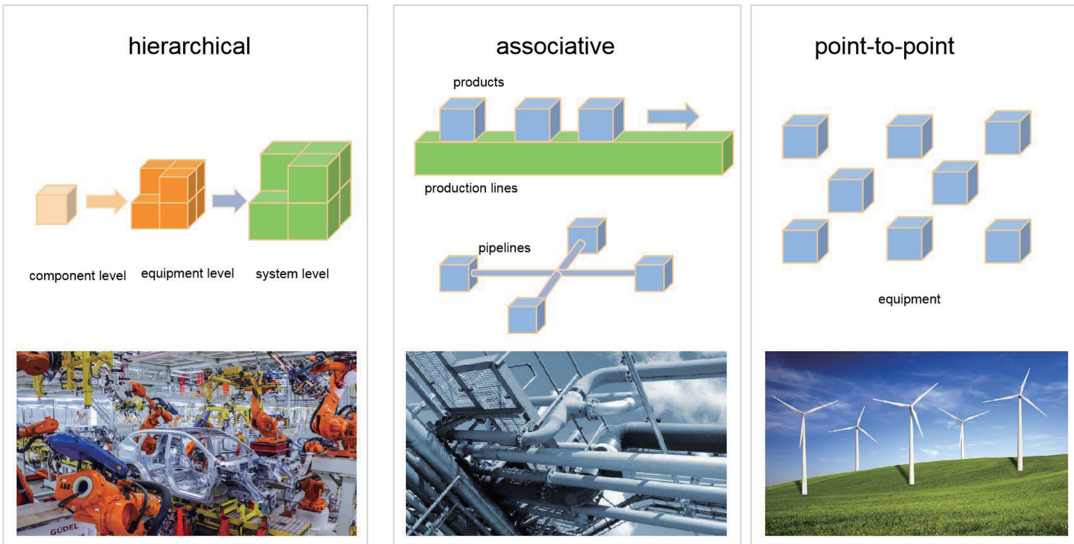

Figure 1. Digital twin relationship

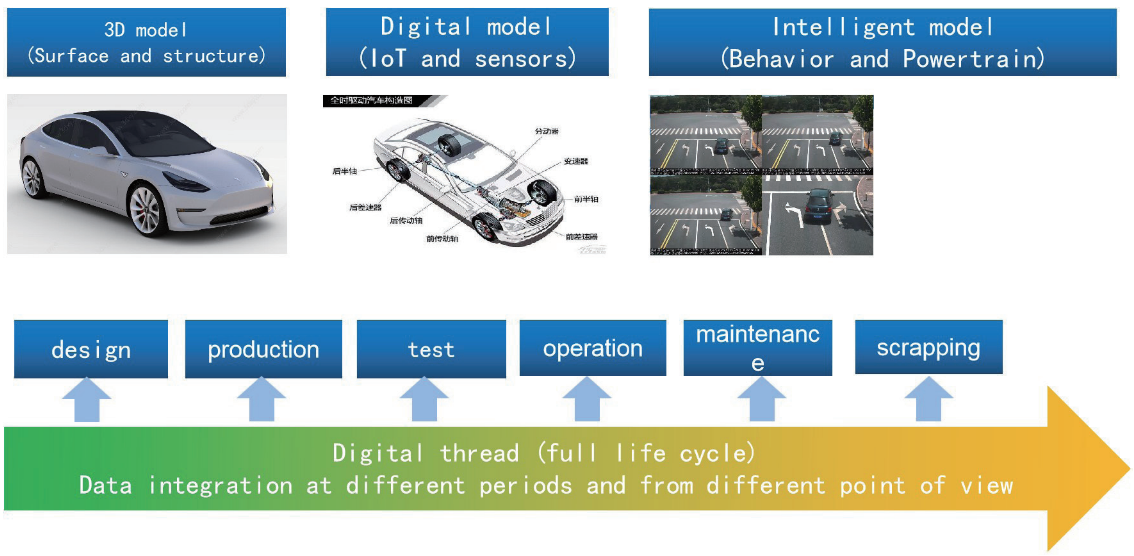

Figure 2. Life-cycle digital threads

1.3 Definition of twin data

Twin data include geometric model related data such as geometric size, assembly relationship and location, physical model related data such as material properties, loads and features, behavioral model related data such as driving factors, environmental disturbance and operation mechanism, and rule model related data such as constraints, rules and association relations and simulation data of process simulation, behavior simulation, process verification, evaluation, analysis and prediction based on the above model.

2. Engine design

2.1 Twin data analysis

2.1.1 Holographic 3D data

Holographic 3D data includes geographical scenes and geographical entities. The geographic scene includes digital elevation model (DEM), digital surface model (DSM), digital orthophoto (DOM), real projective image (TDOM), oblique photography 3D model, laser point cloud, etc. Geographic entity includes basic geographic entity, 3D model of components and other entities. Basic geographic entities include ground object entities and geographic units, which can be expressed in two-dimensional and three-dimensional forms. The 3D model of components includes building structure components, building interior components, road facilities components, underground space components, etc. Other entities include professional type entities produced by other industry sectors. The characteristics of data are low data frequency, strong temporal correlation, strong spatial correlation, large single raster data file, and easy redundancy of data volume update. Among them, BIM 3D data has complex geometric structure, deep layers and rich construction semantic information, including model parameters, materials and dimensions.

2.1.2 Iot data

Iot data refers to the sensing data of various sensor devices, including temperature, humidity, vision, and location sensing devices. The characteristics of the data are: high frequency of data, strong time correlation, and continuous increase of data.

2.1.3 Event data

Event data is automatically generated based on timing sensing data (Iot data, location data) and business rules. This data is characterized by high data frequency, strong temporal correlation and continuous increase of data.

2.1.4 Positioning data

The positioning data is based on different positioning methods to obtain X, Y, Z, speed and direction data. This data is characterized by high data frequency, strong temporal correlation, strong spatial correlation, and continuous increase of data.

2.2 Twin data based scenario analysis

Build a unified spatial entity identity code (spatial + semantic). After importing spatial entities the twin data engine will automatically generate identity codes to produce unified records at the bottom of the multi-source data engine. Provide a unified data source portal, which can query information of different time, different state and different changes according to the ID of twin entities. Support spatio-temporal change capability of geographic entities, and store updates to geographic entities in incremental form. Provide spatio-temporal query interface, including point query (given a point object, find out the method of all spatial objects containing the point); window query (given a query range, find out the spatial objects intersecting with the window or within the range); nearest neighbor query (find out a spatial object with the smallest distance from the given object); inverse nearest neighbor query (find out the spatial object with the given object as the nearest neighbor). A query that is restricted by a given time condition on the basis of a spatial query is a spatio-temporal query.

3. Key technologies

3.1 Spatial entity identity coding technology

The spatial information coding and attribute information coding of the 3D model, and the spatial information coding and attribute information coding of the real-time 3D model are spliced together to get the identity coding information of the holographic 3D model. The specific method is to encode the longitude information and dimension information of the center point of the three-dimensional model recorded in the center coordinate and the distance information between the center point of the three-dimensional model and the center of the earth by octree coding. The specific coding process of the spatial information coding of the real-time three-dimensional model is as follows.

Compared with the two-dimensional planar graph, the three-dimensional model has three coordinate information. In order to represent the three-dimensional information of any point on the earth, the GeoSOT-3D global grid is used to divide the range from the earth's bottom to the center of the earth and up to more than 60 million kilometers. Octree coding is used to encode the center coordinates of the model. Since the resolution of level 32 of GeoSOT-3D global grid is 1.5cm, which can meet most object models in real life, in order to express a sophisticated model, this patent uses 32 layers to encode the center of the model.

In order to unify the units, the grid is first divided according to the GeoSOT-3D global segmentation, and 8° represents a distance of 1024km relative to the equator, which converts the unit m of the distance of the object from the center of the earth into °. Then we expand the sphere of the earth into a flat surface, with the center of the earth equal to 0° point, so as to create a right-handed cartesian coordinate system with due east as the X-axis, due north as the Y-axis, and the axis perpendicular to the XY plane as the Z-axis.

As the global latitude and longitude coordinate range is [-180°,180°], [-90°,90°]. The distance between the surface of the earth and the center of the earth is then less than 50° after taking ° as the unit. Therefore, the range from the center of the earth to the surface of the earth is [0°,50°), and the global coordinates are extended to the three-dimensional space of [256°, 256°,512°] with the center of the earth as the original point. Then, the expanded area is divided into eight equal parts, as shown in Figure 3 below.

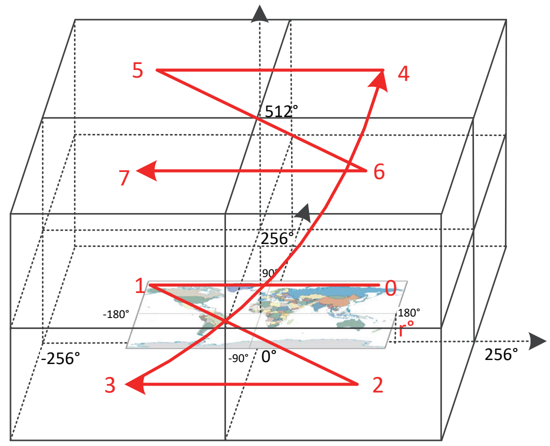

Figure 3. Geographic segmentation

During coding, four possible codes are confirmed by the interval where the X-axis coordinates are located, and then impossible values are excluded by the interval where the Z-axis coordinates are located, and another value is excluded by the interval where the Y-axis coordinates are located. The rest is the final code value.

3.2 Spatio-temporal data dynamic update technology

The operations used for dynamic update of spatio-temporal data mainly include: create, reincarnate, modify, eliminate, etc. Time is 1 d line time, time interval [start, end] expresses space object life cycle, and {(Tstart,Tend)|Tstart<=Tend} is proved. The space object whose life cycle is [start, *] is the current object. Spatial objects whose life cycle is [start, end] are inactive objects (i.e. historical objects). Seven kinds of dynamic operators in object-oriented spatio-temporal database are defined as follows.

Create: To create or generate a new spatio-temporal database record. This operation defines the hour kill of the newly created object as [start, *].

Destroy: To permanently remove a current spatio-temporal object from the database, meaning it will no longer exist in the spatio-temporal database. For spatio-temporal databases, in principle, data should only be added and not deleted, so this operation should be performed with special caution. But we know that in the process of data building and updating, it is hard to avoid mistakes, that is, in the current object of the database, there may be wrong or should not be stored in the object, then we should completely delete the record.

Modify: To modify the properties (including spatial and semantic properties) of the current or non-current object. In a spatio-temporal database, modifying the properties means adding a new version to the object being manipulated.

Eliminate: To alter current record to a non-current record, which defines the lifetime of an object as [start, end].

Recall: To add a record to the database that is not the current object whose lifetime is [start, end].

Forget: To forget historical objects, that is, a record is completely cleared from the historical data. The record no longer exists in the database.

Reincarnate: To create a new space-time database record with the same space, attributes, and semantic information as the original history object and with the timestamp [start, *] on top of the original history object. The operations used to dynamically update spatio-temporal data include: create, reincarnate, modify, eliminate, etc.

3.3 Data engine technology

3.3.1 Sequential data engine

The sequential database is based on the LSM data structure and supports the sequential data to be appended to the database according to the time series. The physical storage sequence is consistent with the logical storage sequence and consistent with the query direction of the user. It is highly efficient for sequential disk reading. The index efficiency of time series database is high. Instead of using the form of B+ tree, it adopts the form of inverted index. Compared with tree index, inverted index has high time efficiency and space efficiency. The sequential database is stored in a column physical structure. For data with the same data type, the compression ratio is higher, which reduces the storage capacity and saves the storage cost.

3.3.2 Holographic 3D spatial data engine

Efficient 3D spatial index is a key technology for fast retrieval of large-scale 3D spatial data. Due to various project such as construction, geological body, pipeline model in geometrical shape and texture characteristics and spatial distribution characteristics of obvious difference, it is often difficult to adopt a common index of three dimensional space to meet all types of 3 d model data and efficient organization. Based on the principle of "efficient unification and reasonable differentiation", this paper develops an efficient multi-level and multi-type hybrid 3D spatial indexing method under the unified 3D spatial indexing mechanism, which takes into account all kinds of 3D spatial entities with multi-layer index and multi-layer filtering from coarse to fine, the 3D spatial data can be queried quickly and accurately.

The multi-level and multi-type hybrid 3D spatial index method firstly uses grid index as the first level index to achieve fast localization. Then, the improved R-tree index and X-List index are used as the secondary index to realize the accurate search of the upper and lower 3D objects, and the LOD characteristics of the upper and lower 3D objects are effectively taken into account. For the upper and lower ground building models with discrete distribution, different target shapes and multiple detail levels, a structurally balanced 3D R-tree index extension method is adopted [10]. The level of detail model is introduced into the intermediate nodes to realize the dual tasks of target query and level of detail query. The node selection algorithm based on bottom-up and then top-down global search and node splitting algorithm based on k-medoids clustering algorithm are used to ensure uniform node size, regular shape and reduced overlap [11].

3.3.3 BIM data engine

The BIM hierarchical objects are stored and managed based on the graph database. Since the spatial relationship data model of IFC entities can directly correspond to Neo4j graph data model, Neo4j graph data model is selected in this paper. There are two important data types in Neo4j: Node and Relationship. Nodes are connected through relationships defined by relationship. At the same time, Node and Relationship are endowed with attributes in the form of key/value keys to represent the detailed information in IFC, which is convenient for Node and Relationship to perform various retrieval operations. In this paper, Neo4j is used to retrieve the spatial relationship between IFC entities for the operation of graph data.

4. Architecture design

The core part of the digital twin data engine is the fusion engine, as shown in Figure 4 below:

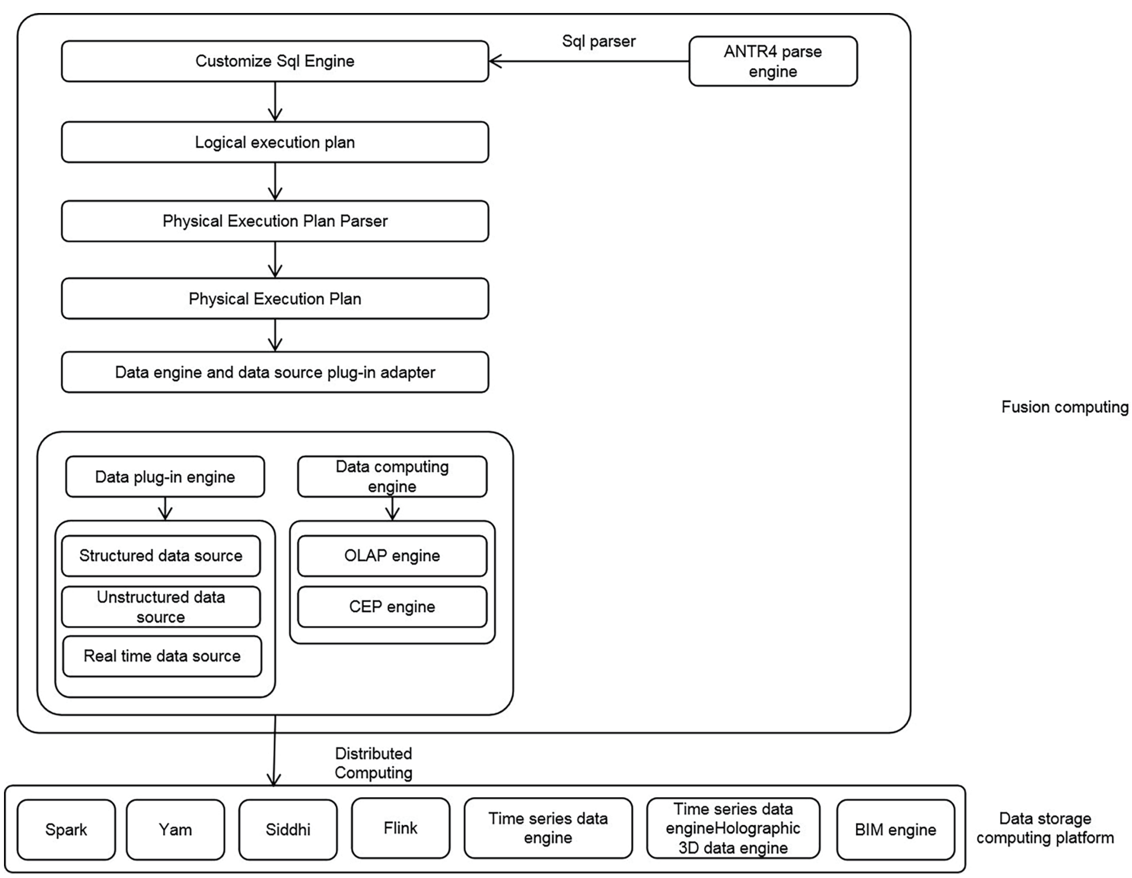

Figure 4. Fusion Data Engine

Data analysis based on multiple sources includes structured data, unstructured data and real-time data.

Based on SQLAccess (custom SQL engine), the data source in the data platform is transparent (users can directly use sql to operate multi-source data fusion analysis).

The overall research and development is carried out through micro-services, and service orchestration and resource management are carried out through K8S+docker, with the capability of automatic service expansion and high service availability.

The distributed architecture is used for storage and computing, which improves the storage and computing performance of the data fusion platform for OLAP.

The platform supports CEP calculation, and the real-time data processing is processed separately, which can support the user's independent algorithm calculation. At the same time, SQLAccess (custom sql engine) can be used for real-time data through SQL to achieve multi-stream fusion, stream batch calculation.

Data storage computing platforms include Spark, Yarn, Siddhi, Flink, temporal database, holographic 3D data engine, and BIM data engine.

5. Application test

Digital twin engine provides spatio-temporal change data model, which can realize the dynamic display of geographic space over time. The changes of different buildings in the two periods are shown in Figure 5 below:

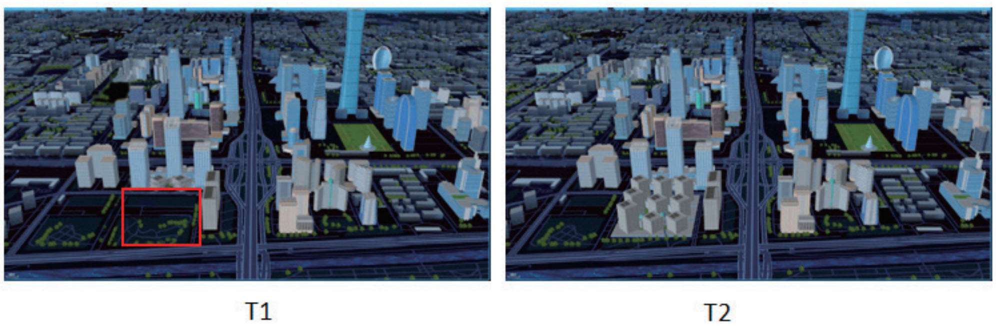

Figure 5. Visualization of spatio-temporal changes

The data twin engine provides SQLIDE page, which can quickly query the digital twin in space and time, as shown in Figure 6 below:

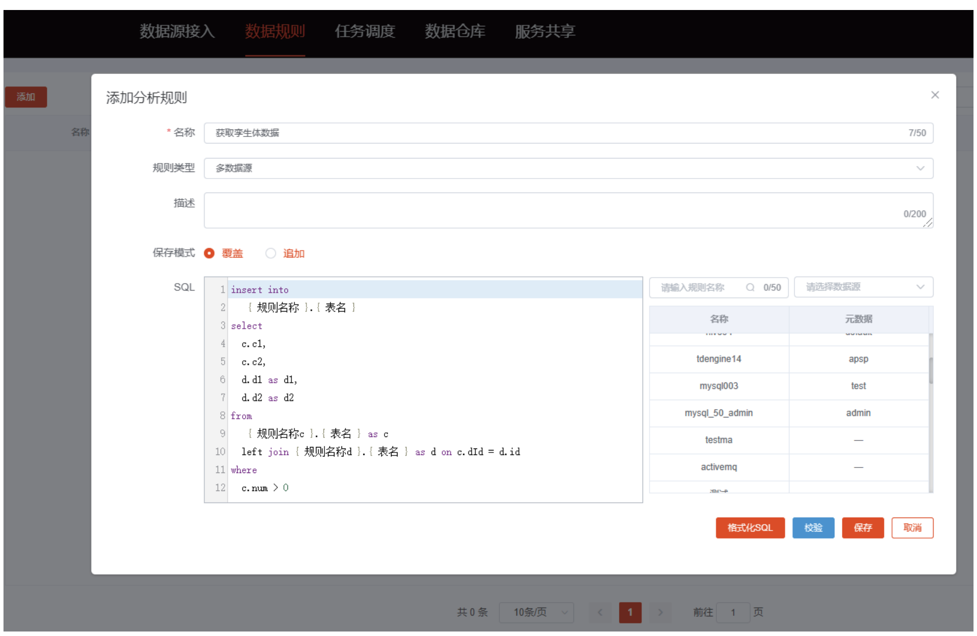

Figure 6. Data fusion SQLIDE

Through the cabinet BIM model data interface provided by the digital twin engine and combined with the digital twin visualization rendering technology, the machine room server linkage alarm is realized, as shown in Figure 7 below:

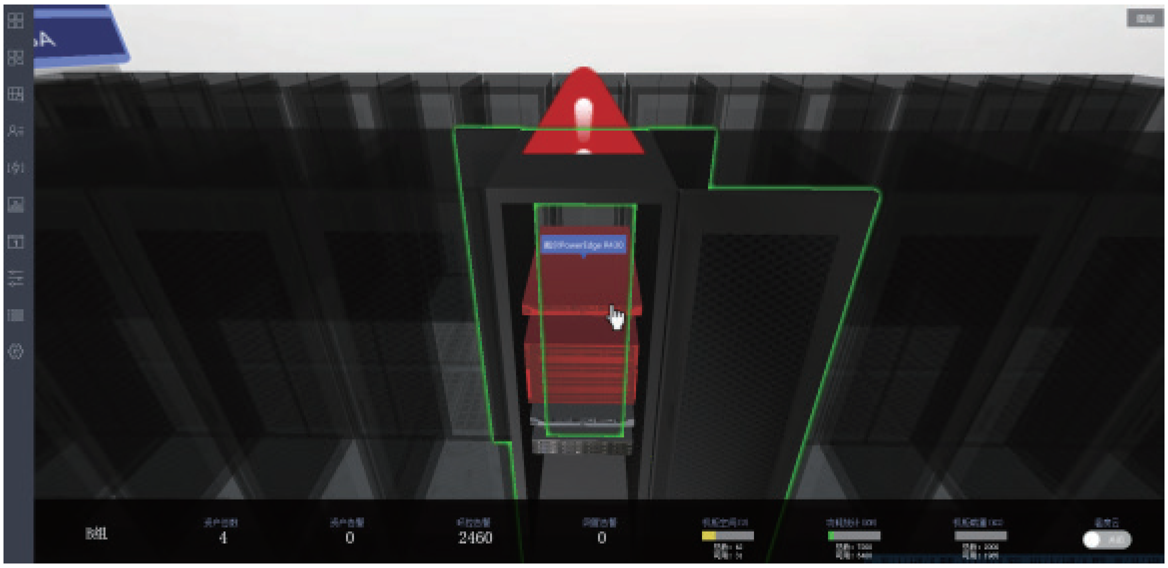

Figure 7. Machine room warning

Through the digital twin engine, the spatial and temporal location, 3D scene data and Iot sensing data are provided to realize the digital twin visual scene restoration, as shown in Figure 8 below:

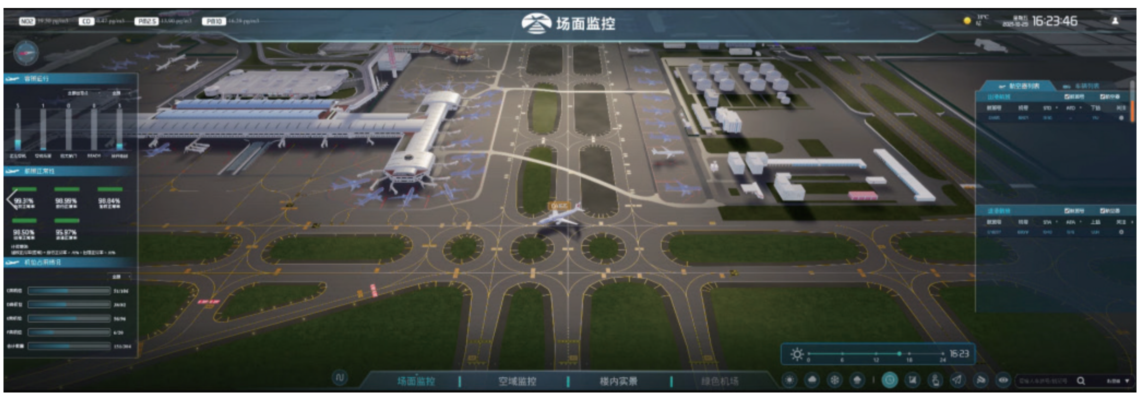

Figure 8. Flight areas based on digital twin

6. Conclusion and outlook

In this paper, spatial entity identity coding technology, spatio-temporal dynamic update technology and fusion engine architecture are used to realize distributed computing, integrated storage, and spatio-temporal change and traceability of digital twin data. At the same time, the engine design is verified according to the application scenario of digital twin core. At present, the efficiency of BIM data retrieval needs to be improved, and the research on BIM data index and scheduling strategy should be continued.

Acknowledgments

This paper was supported by National Key Research and Development Program (2020YFB1600101).

References

[1] WANG Hong-tao, WANG Zhi-chao, CHEN Feng, FENG Lian-qiang, SUN Si-hao. Research on industrial big data application based on time series databas [J]. *Heavy Machinery*, 2020, (04): 17-21.

[2] XU Hua-yan, CHU Yan-long. Design of industrial time series database engine based on influxDB [J]. *Computer Appli-*

cations and Software, 2019, 36(09): 33-36.
[3] TANG Xin-ming, WU Lan. spatio-temporal database model and Temporal Geographic Information System Framework [J]. *Remote sensing information*, 1999(01): 4-8.
[4] WEI Hai-ping. Research and practice of spatio-temporal GIS modeling [D]. PLA Information Engineering University, 2007.
[5] JIANG Shao-hua, WU Zheng. Research on cloud storage and retrieval method of BIM spatial relational data [J]. *Journal of graphics*, 2018, 39(05): 835-842.
[6] WANG Bao-hui, XING Jing-xuan, GAO Yuan. Using FastDFs and drill to build a massive BIM family data storage and query platform [J]. *Civil engineering information technology*, 2016, 8(06): 23-28.
[7] LI Jing-xin. CMDB system based on graph database storage engine [J]. *Software and integrated circuits*, 2021, (09): 24-27.
[8] LIU Chang-hong. Graph query and graph calculation hybrid engine for distributed graph database [D]. University of Electronic Science and technology, 2021.
[9] Grieves M,Vickers J. D*igital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems* [M]. Springer International Publishing, 2017.
[10] ZHU Qing, LI Xiao-ming, ZHANG Ye-ting, LIU Gang. Design and implementation of an efficient 3D GIS database engine [J]. *Journal of Wuhan University* (Information Science Edition), 2011, 36(02): 127-132+139.