The Fuse Platform: Integrating data from IoT and other Sensors into an Industrial Spatial Digital Twin

Gregory Biegel¹, Nicholas Bower², Will Castelnau²

¹Amazon Web Services, 240 St Georges Terrace, Perth, Western Australia 6000, Australia - gbiegel@amazon.com ²Woodside Energy, 11 Mount St, Perth, Western Australia, 6000, Australia - (nick.bower, will.castelnau)@woodside.com

Keywords: Digital Twin, IoT, Knowledge Graph, Graph Database, Cloud Computing

Abstract

Digital Twins as virtual representations of industrial assets are being used to assimilate varied sources of data for improved awareness and decision making in operations and process optimisation. This paper explores the integration of IoT sensors into a spatial digital twin called Fuse that Woodside Energy has been building for the assets it operates. We describe the Fuse platform and its knowledge graph data core that is used to organise and inter-relate data for presentation within 3D visualisations, domain-specific contexts and immersive augmented reality presentations. The key contribution here is the use of a knowledge graph to link diverse data sources so as to contextualise sensor data for actionable insights. One area of significant innovation has been the development and use of new Internet of Things (IoT) devices which have been enabled by advances in sensor technology, connectivity, and cloud computing. These new tailored data sources are complimenting existing plant and resource planning data for improved asset monitoring, predictive maintenance and process automation.

1. Introduction

Woodside Energy, an international energy company established in Australia, produces oil and liquefied natural gas (LNG) and plays a significant role in the global energy industry. It owns and operates assets internationally in many locations around the world that are often remote and with harsh environmental factors compounding already hazardous work environments. A key feature of Woodside Energy's vision for future asset optimisation is the Intelligent Asset equipped with abundant sensors, sophisticated digital integration patterns and advanced analytics capabilities, made accessible to users through a digital twin that seeks to reduce distraction by repetitive tasks, reduce travel time through improved accessibility to asset data, enable re-use of data for accelerating digital transformation and enriching decision making through better awareness of asset health.

Woodside Energy's approach to devising a digital twin called Fuse is documented in (Burgin and Wallace, 2023) and can be summarised by having developed competencies in industrial sensing, in determining and conferring insights, and in enabling systems for action, a so-called sense-insight-action strategy. Fuse was first deployed for the Woodside Energy operated Pluto facility in north-western Australia where natural gas is extracted, liquefied, stored and transported to customers. Here Fuse combines data from over 200,000 Industrial IoT (IIoT) sensors embedded within the plant (Chanthadavong, A., 2016), data from custom built wireless battery-operated IoT sensors described in greater detail here, and data from a variety of enterprise systems. This data is presented in a variety of contexts from rich three dimensional (3D) visualisations derived from CAD models, LiDAR scans and photogrammetry to conventional map orientations, mobile device and augmented reality experiences. Far more than just a data exploration tool, the Fuse digital twin platform is today involved with the execution and therefore optimisation of actionable work such as scheduled inspections, active condition monitoring, maintenance task and procedures and proactive surveillance.

This paper describes the authors' experience of building a large scale industrial spatial digital twin over a period of five years for an operations workforce. The contribution of this work is to provide insights gained from the adoption of a large scale spatial digital twin in the real world with a focus on the incorporation of wireless IoT sensors using a knowledge graph.

Continued advances in sensor miniaturisation, low-volume part availability, electronics cost reduction, ubiquitous connectivity and cloud computing have formed the right conditions to enable in-house development of IoT devices for equipment condition monitoring and visual surveillance. These multi-sensor devices allow us to fill gaps in physical world plant data so that we may pursue more ambitious opportunities in visualisation and analysis that may not have been possible with existing plant data only. We have found that integrating IoT data into spatial digital twins increases situational awareness, aids in predictive analytics and enhances decision support. This paper describes the challenges inherent in integrating IoT data into a spatial digital twin and provides a semantic framework for organizing and interconnecting data related to the physical system being mirrored by the digital twin.

At the core of Fuse is a digital representation of the plant environment called the Reality Engine. Its purpose is to maintain knowledge about how existing data is inter-related and therefore how to combine and contextualise it. This virtual representation needs to be able to model a complex physical industrial environment from multiple perspectives and possess efficient storage and retrieval for structured processing of information. In addition, the virtual model needs to be flexible enough to represent different types of relationships, dependencies and interactions found in the physical environment including between equipment, data sources, and metadata. The model has to accommodate updates and changes in the physical reality it represents in order to continue to provide an accurate reflection of its current state.

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4-2024 ISPRS TC IV Mid-term Symposium "Spatial Information to Empower the Metaverse", 22–25 October 2024, Fremantle, Perth, Australia

2. Modern Industrial IoT

The continuing decrease in the cost of sensor hardware (Microsoft, 2019) (average sensor cost falling from \$1.30 in 2004 to \$0.38 in 2020) has led to measurements that were cost prohibitive only a decade ago now becoming commoditised. Ongoing innovation in sensor technologies, fabrication techniques, and algorithms has led to new classes of sensors that can detect and quantify a broad range of physical and chemical phenomena reliably and with efficient use of energy.

When combining modern connectivity pathways into existing plants, and at-scale computing as offered by the cloud, the data generated by these sensors is increasingly being used by organisations like Woodside Energy to increase the efficiency, flexibility, and dynamism of their operations.

2.1 Fixed sensors

Industrial plants, such as the Woodside Energy operated Pluto LNG plant in north-western Australia are routinely equipped with hundreds of thousands of fixed physical industrial IoT sensors that measure all aspects of the industrial process, and are integrated into the plants Operational Technology (OT) Distributed Control System (DCS) (Chanthadavong, A., 2016). The data from these sensors is made available within the Information Technology (IT) network as a timeseries or historised database with each sensor reading made up of a control system tag identifier, a timestamp, and a value. The equipment under measurement by IoT sensors is represented in many other enterprise information systems (e.g. maintenance, engineering and permit systems), but often the same piece of equipment is identified differently in each system and it can be difficult and time consuming to link data from different systems together for analysis.

2.2 Wireless sensors

Wireless sensors are physical sensors that are not built into the plant infrastructure and can be moved around to gather focused measurements of different equipment at different points in time. Wireless sensors are battery powered and use wireless communication protocols such as Wi-Fi and LoRa, making them easily portable. Due to the portability of this type of sensor and the fact that they may be measuring different entities at different points of time, the resulting data stream requires additional contextualisation before it can be consumed, which is in contrast to fixed sensors that constantly measure the same entity at a fixed position.

2.2.1 Wireless IoT sensors within Fuse Within the Fuse platform, Woodside Energy has developed a custom in-house IoT hardware platform (known as Intellisense Pulse) that leverages ultra low power and Long Range (LoRa) wireless communication technology, enabling sensor measurements to be transmitted over a range of several kilometers, with battery life measured in years. Given the possibility of free hydrocarbons in and around the process chains where these are deployed, the sensors are designed for safe use in explosive atmospheres, whereby they are Intrinsically Safe and certified to offer insufficient energy as a source of ignition, even at elevated temperatures.

The IoT hardware devices form an extensible platform that can carry different external sensor implements - an example device

Figure 1. An IoT device with four accelerometer probes attached

with magnetic accelerometer sensor probes attached (to measure vibration for purposes of condition monitoring) is shown in Figure 1.

The device's compact form factor, combined with its low power consumption and long range wireless communication capabilities allows for easy deployment to hard to reach sensing points in the plant, and can also easily be moved to measure different equipment as needed. An example device is shown deployed in the field in Figure 2 (base transmitter and probes are easily mounted to metallic surfaces thanks to built-in strong rare earth magnets). The attachment of the accelerometer to plant equipment is shown in greater detail in Figure 3.

Figure 2. An IoT device and sensor probes deployed in the field

Woodside Energy also leverages mobile sensors including mobile phone cameras and sensors deployed via platforms including ground and aerial robotics to capture image, acoustic, and thermal data, with new sensor technologies currently being trialled.

2.3 Virtual sensors

A virtual sensor, also known as a soft sensor, is a function or prediction model (machine learning model, thermodynamic model, mathematical equation, etc.) that derives a measurement from other available measurements and sensor data. The output of the soft sensor is assigned an identifier and timestamp and recorded in the timeseries database in much the same way as

Figure 3. An accelerometer sensor attached to equipment

other sensor measurements are. Virtual sensors are often adopted when measuring a property with a physical sensor is difficult, costly, or impractical. Woodside Energy applies a number of data science models to physical sensor data and records output as virtual sensors (Coyne, A., 2016).

3. Contextualisation of IoT data in a spatial digital twin

There are several challenges inherent in integrating data from IoT sensors into an industrial spatial digital twin such as that developed by Woodside Energy through the Fuse platform.

Challenges exist in effectively managing a multitude of disparate sensor data streams arriving at high volumes and velocities, in diverse formats, and using different communication protocols, however the greatest challenge lies in contextualising the plethora of data available, or in other words relating sensor data with the entities that it measures.

3.1 Data contextualisation

Data contextualisation refers to the process of enhancing raw data with additional information to make it more meaningful and relevant to consumers of the data. In many modern enterprises data contextualisation remains a costly manual process, often depending on individual domain knowledge.

A common example of manual data contextualisation within Woodside Energy is the need for users to query multiple different applications to gather information about a single piece of plant equipment. These include querying the maintenance system to understand what work may have recently occurred on the equipment, the engineering system to understand the characteristics of the equipment, and the historian to review recent sensor data related to the equipment. Data for the same equipment may be identified differently in each system, and the user is required to understand not only which applications to query to find the data, but also how the data is identified in each of these applications.

With the ongoing increase in both the quantity and diversity of data sources (as a result of the proliferation of IoT devices), manual or even point-to-point data contextualisation becomes increasingly complex and consequently expensive. Fuse's approach to this has been the invention of the Reality Engine that consists of a data assembly layer and a knowledge graph, subject to further discussion in the following sections.

3.2 Knowledge graphs for contextualisation

Within Fuse we have built a knowledge graph implemented using a labelled property graph running within the Amazon Neptune managed graph database service (Amazon Web Services (AWS), 2024a) to model Woodside Energy's complex industrial plant assets and the relationship between physical equipment, logical identifiers, spatial locations, and data sources including images and documents.

A graph is a simple data structure composed of nodes (vertices) connected by relationships (edges), to create high fidelity models of a domain. Formally, a Directed Acyclical Graph (DAG) is an ordered pair $G = (V, E)$, comprised of a set of vertices, V , and a set of directed edges, E , each of which is an ordered pair (v_i, v_j) where $v_i, v_j \in V$, representing a connection from v_i to v_j as shown in Figure 4.

Figure 4. A directed acyclical graph (DAG)

A knowledge graph enriches the simple graph, extending it with semantic metadata to add explicit representation of knowledge within the model (Barrasa and Webber, 2023). Knowledge graphs are critical to many enterprises today and are employed by many of the largest global technology companies in their digital products (Noy et al., 2019). The Fuse knowledge graph provides an ontology by defining categories of objects with similar behaviour and attributes (e.g. types of equipment), the types of relationships between these objects (e.g. a vertex representing piece of equipment can be related to a vertex representing a geographical location), and finally a set of constraints over objects and relationships (e.g. a vertex representing a sensor may only be related to a vertex representing an IoT device).

An example of the ontology provided by the semantic graph model in Fuse is shown in Figure 5. If we know that vertex V_1 represents a **sensor**, and vertex V_2 represents a piece of equipment (as defined by the labels on the vertices), and that the edge E_1 represents a measurement that the sensor makes of the equipment, a software program will be able to answer a question such as 'what equipment does this sensor measure?', as well as the inverse question of 'which sensor measures this equipment?'. In addition, the set of properties (mandatory and optional) that are defined on both sensor and equipment vertices in the graph is defined within the model.

Figure 5. A simple knowledge graph where V_1 represents a sensor, V_2 represents equipment, and E_1 represents a measurement the sensor makes of the equipment

3.3 The Fuse knowledge graph

The Fuse knowledge graph models the physical engineering structure of the industrial plant (as a hierarchy of plant equipment), as well as the representation of the plant found in the maintenance and historian systems in the form of a canonical

data model. The canonical data model provides a standardised representation of the plant equipment that is independent of other specific systems where the data is represented. The canonical data model links specific identifiers from other systems together to the canonical representation of an entity, allowing data that exists, for example, in a maintenance system, to be linked with that in an engineering system.

The Fuse knowledge graph enables the contextualisation of data within the digital twin, crucially allowing data streams arriving from static, mobile, and soft sensors at high volumes and velocities to be related to the entities and equipment being measured at a specific point in time.

This is illustrated in Figure 6 showing a fragment of the Fuse knowledge graph. Each vertex in the graph has a label, represented by colours in the figure, with vertices being related to each other by labelled, directed edges. The edge representing the relationship between the mobile IoT temperature sensor and the equipment it measures has an additional temporal component, represented by the properties on the measures edge named *from* and *to*. This is due to the IoT sensor being a mobile device that can be moved to measure different equipment at different points in time. In Figure 6 we can see that between January 1^{st} and March 25^{th} 2024 (t_1), the IoT sensor measured the temperature of the Fan, but from March 26^{th} onwards (t_2), the sensor has been moved and is now measuring the temperature of the Pump.

Figure 6. Sample knowledge graph fragment

Table 1 illustrates the data stream from the mobile IoT temperature sensor between March 1^{st} and March 26^{th} , 2024. As shown in the table, without any additional semantic metadata to add context to the data stream, there is no way to determine what the measurements relate to at t_1 versus t_2 , and therefore the data has limited value in decision making. As we see in Figure 6, the equipment being measured by the sensor changed on March 25^{th} , meaning that the last three readings in Table 1 relate to a completely different entity being measured by the sensor than the first three readings in the table.

By representing the relationship between the mobile IoT sensor and the physical entity it is measuring at a specific point in time

	Sensor Id	Timestamp	Reading
	temperature_sensor_1	1709254084	34.76
t_1	temperature_sensor_1	1709254184	35.68
	temperature_sensor_1	1709254284	36.01
	temperature_sensor_1	1711414084	31.46
t_2	temperature_sensor_1	1711414184	31.34
	temperature_sensor_1	1711414284	31.78

Table 1. IoT Temperature Sensor data stream

as illustrated in the knowledge graph in Figure 6 we can programatically determine what a set of sensor measurements relate to for a given time period by simply performing a graph traversal as shown in the Gremlin query in Listing 1 showing that, between January 1^{st} and February 1^{st} 2024, the temperature being measured was that of the Fan. Changing the dates within our traversal to cover a time period after March $25th$ would yield the Pump.

```
gremlin> g.V("iot_temperature_sensor")
.outE("measures")
.union(
    has(
        "from", lte("02/01/2024")
    ).hasNot("to"),
    has(
        "from", lte("02/01/2024")
    ).has("to", gte("01/01/2024"))
)
==>e[1][temperature_sensor-measures->fan]
```
Listing 1. Graph traversal to determine what the temperature sensor is measuring between January 1^{st} and February 1^{st} , 2024

The Fuse knowledge graph also represents relationships between physical equipment and its spatial location, represented by co-ordinates stored as properties on a vertex with label GeoLocation. Equipment is related to its location by an edge with label of **locatedAt** as shown in Figure 6. Relationships between static IoT sensors and the equipment they measure are also represented by vertices with the label ControlSystemTag. By contextualising the data produced by static and mobile IoT sensors, with the entities that they measure and their spatial location, Fuse is able to combine data with three dimensional visualisations to provide immersive, augmented reality experiences that bring data closer to users as shown in Figure 8. The lifelike representation of the physical environment allows users to gain a deeper spatial understanding of the system and related data which is especially valuable for complex, large-scale systems where spatial relationships are critical, as they are in an industrial plant such as that operated by Woodside Energy.

3.4 Graph query and data retrieval

Using a knowledge graph to model the physical world enables the use of mature and well understood graph retrieval algorithms to gather insights from the graph database. The depth first search algorithm forms the basis of most of the transactional queries within Fuse that follow a path to retrieve information. A good example of this is the traversal shown in Listing 1 where an informed depth first search (Robinson et al., 2015) is used to retrieve the equipment measured by a sensor during a given time range.

Today both Apache Gremlin (Apache Software Foundation, 2024) and the recently open-sourced openCypher (openCypher, 2024) are used within Fuse. Despite beginning with Gremlin, over time we expect to see greater use of openCypher due to its

succinct path matching capabilities and alignment to ISO/EIC incorporated Graph Query Language (GQL) (ISO, 2024).

Whilst most of the data contextualisation queries within Fuse are short lived queries within the timeframe of a user interaction, we can also use the knowledge graph for analytical queries via Neptune Analytics, a memory optimized graph database engine for analytics that provides a range of analytical algorithms (Amazon Web Services (AWS), 2024b) including path-finding algorithms that can be applied for route finding in combination with the spatial location data stored in the graph.

3.5 Spatialisation

Relating streams of sensor data to the physical equipment that it measures at a point in time contextualises the data and enables it to be used to make data driven decisions. Lifelike 3D visualisations provide additional visual context that is easier for humans to understand than tabular dashboards or 2D representations by providing a realistic and immersive representation of the physical system being mirrored, with data displayed spatially located alongside the equipment it relates to. They offer a lifelike representation of the real world, allowing stakeholders to gain a deeper spatial understanding of the system which is especially valuable for complex and large scale systems where spatial relationships are critical, for example in a liquefied natural gas plant.

In order to overlay data in the correct location in a 3D scene, we need to know the relationship between the data and the equipment, as well as the relationship between the equipment and its physical location. As shown in Figure 6 Fuse stores the physical location of equipment in the real world, as GeoLocation vertices within the knowledge graph, as well as storing the relationships between equipment, and sensors measuring it.

Figure 7. Spatialisation of equipment enables proximity based search and contectualisation

3.5.1 Proximity based data retrieval A key feature of the data contextualisation capabilities of the Fuse platform, is the ability to perform data contextualisation based on physical proximity. This feature is leveraged extensively within the 3D visualisation component of the digital twin, allowing users to be presented with relevant data as they move through the 3D scene, but the spatial query capabilities of the platform are exposed via API to allow other systems to perform data contextualisation based on location.

An example of proximity based data retrieval is illustrated in Figure 7 where we see the user performing a query for all equipment within 5 meters of their location. Pump 1 and Tank 1 are located within 5 meters of the provided user co-ordinates, and are thus returned by the query, whilst Tank 2 and Pump 2 fall outside the queried proximity and therefore are not. Once the equipment of interest has been identified via a proximity search, the knowledge graph can be traversed to return associated data.

Figure 8. Spatially located IoT data integrated into a 3D model

It is the relationship between equipment, spatial location, sensors, and control system tags that enables the embedding of sensor data into 3D visualisations as shown in Figure 8 that provides users with enhanced perception, aiding in better understanding of the layout, structure, and inter-connectedness of equipment and data within Fuse.

3.6 Retrieval Augmented Generation

The recent popularisation of the generative capabilities of Large Language Models (LLMs) has demonstrated how they can be used to present data in a way that is even easier for users to understand and work with. Most LLMs in common use today are trained on vast amounts of publically available data and are able to create new content based on patterns they have observed in this training data. Whilst LLMs can produce extremely realistic looking outputs, without a grounding in actual facts about a domain, these outputs may be inaccurate (a phenomenon known as a hallucination).

We leverage the data stored in the knowledge graph within Fuse to provide context, in the form of known facts about entities modelled by the digital twin, to an LLM in a process known as Retrieval Augmented Generation (RAG). Following this approach we can use an LLM to summarise and collate a range of data to derive higher level insight, in a natural conversational style.

4. Capabilities

The Fuse platform provides a wide range of capabilities to Woodside Energy, including the following:

4.1 Equipment 360 Application Programming Interface (API)

The relationships stored by Fuse in its knowledge graph connect the many different sources of equipment data available within

Woodside Energy, including structured, unstructured, and timeseries data. Equipment 360 is an equipment-centric API within Fuse that uses the context captured in these relationships to query and collate data from source systems for a given piece of equipment within a provided time range, in effect returning a 360◦ view of it. As a GraphQL based API, Equipment 360 allows callers to specify exactly what data they require, and returns only the data fields and thus source systems requested. This API has found use within the organisation both as an enabler for Fuse and for several other digital products.

4.2 Autonomous robotic data capture

Spector (Woodside Energy, 2023) is a robotic data capture service developed by Woodside Energy using Boston Dynamics quadruped robots to autonomously navigate the Pluto LNG plant, capturing data to be used in electrical inspections. The robots carry a variety of sensor payloads including a 360-degree camera, 30x optical zoom camera, and thermal camera. The data captured by these robots is sent to Fuse where it is contextualised, including through the use of computer vision techniques to extract equipment identifiers from images, allowing relationships to be created between equipment and images. The contextualisation of the captured data allows inspectors to access it via a web application and perform their inspections without having to locate and travel to the equipment in the field as before.

4.3 Condition-Based Monitoring

Woodside Energy has leveraged the Fuse mobile IoT sensors and data contextualisation capabilities, together with diagnostic machine learning models running in the cloud, to continuously monitor rotating mechanical equipment and detect anomalies before they cause faults that may impact plant operation. Using Fuse's knowledge graph to combine sensor data with maintenance history is enabling Woodside Energy to shift from reactive maintenance to more predictive maintenance of specific classes of equipment.

4.4 Knowledge graph System Development Kit (SDK)

The Fuse knowledge graph provides a data resource that is valuable outside the spatial digital twin itself, but accessing and querying the graph requires specialist skills and knowledge of a property graph query language (either Gremlin or openCypher). To enable data contextualisation without the need to understand graph primitives, Fuse provides a TypeScript SDK that provides other applications a higher level interface into the data stored in the graph.

4.5 Mobility and Site Based Data Collection

With its enabling knowledge graph and API components, it has been an obvious step for Woodside Energy to project its digital twin capability out to site based users to gain improved data quality on entry and efficiency in scheduled task/procedure execution. Mobile based applications are better able to incorporate richer contexts than previously for the given task at hand.

5. Conclusions

The approach of integrating data streams from static and wireless IoT sensors, as well as virtual sensors into a spatial digital twin as adopted by the Fuse platform has been proven in the

real world deployment of the platform to multiple Woodside Energy assets (Woodside Energy, 2021). The contextualised data provided by Fuse continues to find new uses within Woodside Energy as the company strives to improve asset awareness and digitise repetitive and unsafe work. Outside conventional digital techniques in software delivery, specific focus has been placed on improving our ability to sense new information and in recognising the importance of using a knowledge graph to store information that permits the combination of the many integrated data sources of Fuse.

References

Amazon Web Services (AWS), 2024a. Amazon Neptune. https://aws.amazon.com/neptune/ (26 March 2024).

Amazon Web Services (AWS), 2024b. Neptune Analytics algorithms. https://docs.aws.amazon.com/neptuneanalytics/latest/userguide/algorithms.html (26 March 2024).

Apache Software Foundation, 2024. Apache gremlin. https://tinkerpop.apache.org/gremlin.html (26 March 2024).

Barrasa, J., Webber, J., 2023. *Building Knowledge Graphs A Practitioner's Guide*. O'Reilly.

Burgin, L., Wallace, L., 2023. Sense, Insight, Action - Woodside's Intelligent Asset Journey. *The APPEA Journal*, 63, s159– s163.

Chanthadavong, A., 2016. Woodside takes the controls of its gas plants with connected sensors. https://www.zdnet.com/article/woodside-takes-the-controls-ofits-gas-plants-with-connected-sensors/ (26 March 2024).

Coyne, A., 2016. Woodside deploys IoT sensors to prevent plant disasters. https://www.itnews.com.au/news/woodsidedeploys-iot-sensors-to-prevent-plant-disasters-418730 (26 March 2024).

ISO, 2024. Graph query language. https://www.iso.org/standard/76120.html (26 March 2024).

Microsoft, 2019. 2019 Manufacturing Trends Report. https://info.microsoft.com/rs/157-GQE-382/images/EN-US-CNTNT-Report-2019-Manufacturing-Trends.pdf (26 March 2024).

Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., Taylor, J., 2019. Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue*, 17(2), 48–75. https://doi.org/10.1145/3329781.3332266.

openCypher, 2024. https://opencypher.org/ (26 March 2024).

Robinson, I., Webber, J., Eifrem, E., 2015. *Graph Databases*. O'Reilly.

Woodside Energy, 2021. Dollars and Sensors. https://www.woodside.com/media-centre/newsstories/story/dollars-and-sensors (26 March 2024).

Woodside Energy, 2023. Introducing Spector: A woodside Robotics First. https://www.woodside.com/media-centre/newsstories/story/introducing-spector-a-woodside-robotics-first (26 March 2024).