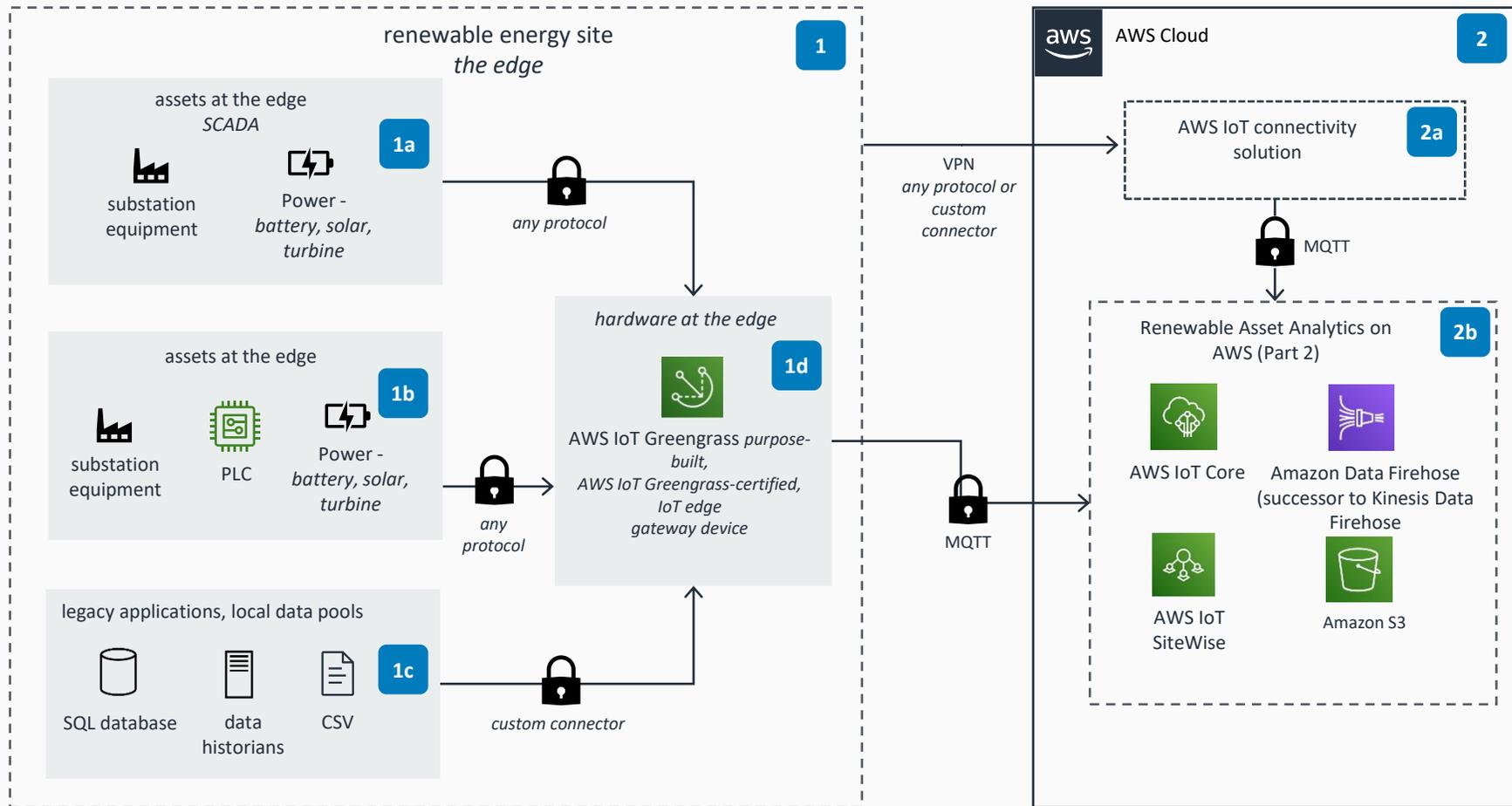


Guidance for Renewables Data Lake and Analytics on AWS

Part 1

This diagram shows different scenarios for renewable energy operators to consider when ingesting data from different edge configurations, such as supervisory control and data acquisitions (SCADAs) and legacy applications. The configurations or applications communicate data over a secure virtual private network (VPN) connection to an AWS internet of things (IoT) connectivity solution.

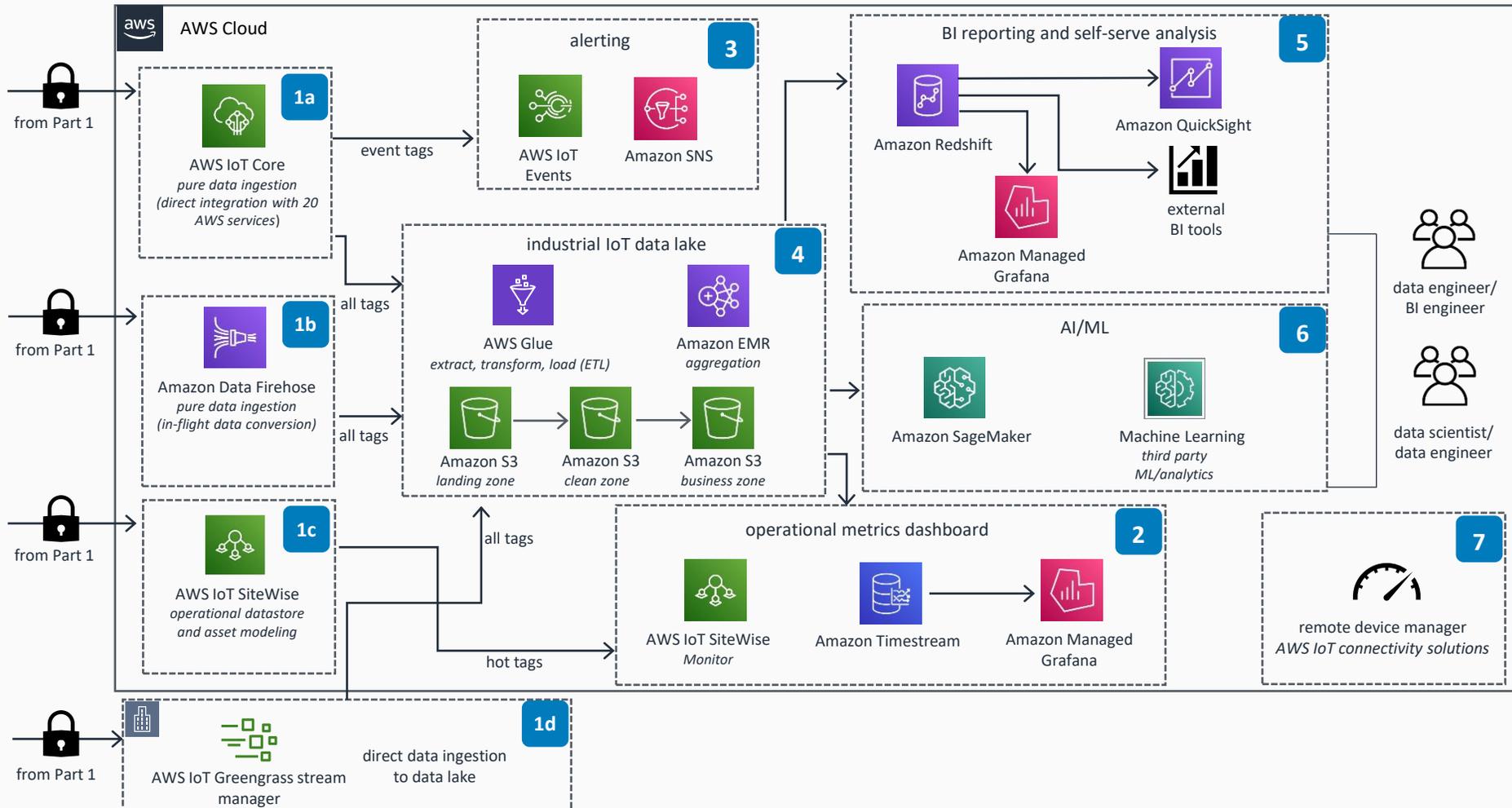


- 1** The renewable energy site represents the edge and includes multiple assets, topology, and device configurations.
- 1a** **Configuration scenario:** Supervisory control and data acquisition (SCADA) is available for assets **substation equipment** and **power - battery, solar, turbine**.
- 1b** **Configuration scenario:** When SCADA is not available (for example, when the energy output of a site doesn't justify the investment in a SCADA system). In this case, you communicate directly with the programmable logic controller (PLC) of the assets (**substation equipment** and **power - battery, solar, turbine**).
- 1c** **Configuration scenario:** SCADA isn't available and there is no access to PLCs. This might be due to regulatory, compliance, or security reasons. In this case, the renewable energy operators deposit the asset data in an external application. This can be a **Structured Query Language (SQL) database, data historian, or CSV files**.
- 1d** **Hardware at the edge option:** The device can ingest Internet of Things (IoT) data from SCADA or PLCs in any protocol (for example, Open Platform Communications Unified Architecture (OPC-UA), Distributed Network Protocol 3 (DNP3), Modbus, and Sunspec). For legacy applications, the device connects through prebuilt or custom connectors. It communicates the IoT data over the internet and through Message Queuing Telemetry Transport (MQTT) protocol to the Part 2 solution hosted in the Cloud. All traffic is encrypted through x509 certificates.
- 2** The AWS IoT connectivity solution and the Part 2 solution are hosted within the AWS Cloud.
- 2a** No hardware at the edge option - Sometimes it is too costly to install an IoT edge gateway device at each renewable site. In this case, we use an AWS IoT connectivity solution, hosted entirely in the Cloud. The SCADA, PLCs, or legacy applications communicate the IoT data through the native device protocol, over a secure virtual private network (VPN) connection, to the AWS IoT connectivity solution. The connectivity solution forwards the IoT traffic over MQTT for performance analytics, as shown in part 2. All traffic is encrypted through x509 certificates.
- 2b** Continue to Part 2.

Guidance for Renewables Data Lake and Analytics on AWS

Part 2

The AWS IoT connectivity solution can help operators build a modern, complete, edge-to-cloud solution to ingest near real-time data from renewable assets such as wind turbines, solar farms, and hydro dams. An industrial IoT data lake is created where advanced analytics can be performed. Operators can derive insights from their asset data by using machine learning, near real-time dashboards, alert management, business intelligence (BI) reporting, and comprehensive device management.



1a

Data is ingested to **AWS IoT Core**, for non-asset modeled data, including native integration with 20 AWS services.

1b

Data is ingested through **Amazon Data Firehose** to **Amazon Simple Storage Service (Amazon S3)** with optional in-flight data conversion (for example, conversion from JSON to Parquet).

1c

Data is ingested at scale with detailed asset modeling in **AWS IoT SiteWise**.

1d

AWS IoT Greengrass stream manager transfers high volume data directly to the AWS Cloud, with low latency.

2

AWS IoT SiteWise, **Amazon Timestream**, and **Amazon Managed Grafana** make up the near real-time operational dashboard of "hot tags" (critical tags for health monitoring of assets).

3

Build detector models in **AWS IoT Events** to continuously monitor the state of assets and issue immediate alerts in **Amazon Simple Notification Service (Amazon SNS)**. This done through email and short message service (SMS) to operational staff.

4

The industrial data lake is hydrated by different sources at different velocities. The data lake serves as a single version of truth for all consumers. Data lands "as-is" from sources, in a landing zone **Amazon S3** bucket. From here, it is cleansed and normalized through **AWS Glue** ELT into a curated state and placed in a clean zone **Amazon S3** bucket. **Amazon EMR** consumes this curated data to calculate 10-minute averages. **Amazon EMR** also converts the clean data into the IEC-61400-25-2 standard for wind and IEC 61850-7-420 standard for solar. **Amazon EMR** then deposits the aggregated and standardized data in an **Amazon S3** bucket called *business zone*.

5

Data from the **Amazon S3** bucket *business zone* is loaded into **Amazon Redshift**. Detailed business intelligence (BI) reporting can be done using **Amazon Managed Grafana** or **Amazon QuickSight** which uses Super-fast, Parallel, In-memory Calculation Engine (SPICE). It is also possible to connect with external BI tools like Tableau.

6

Artificial intelligence and machine learning (AI/ML) services, like **Amazon SageMaker**, use curated data from the data lake for predictive health analysis and assessment.

7

AWS IoT connectivity solutions have the full range of remote device management capabilities.



Reviewed for technical accuracy September 25, 2024

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Reference Architecture