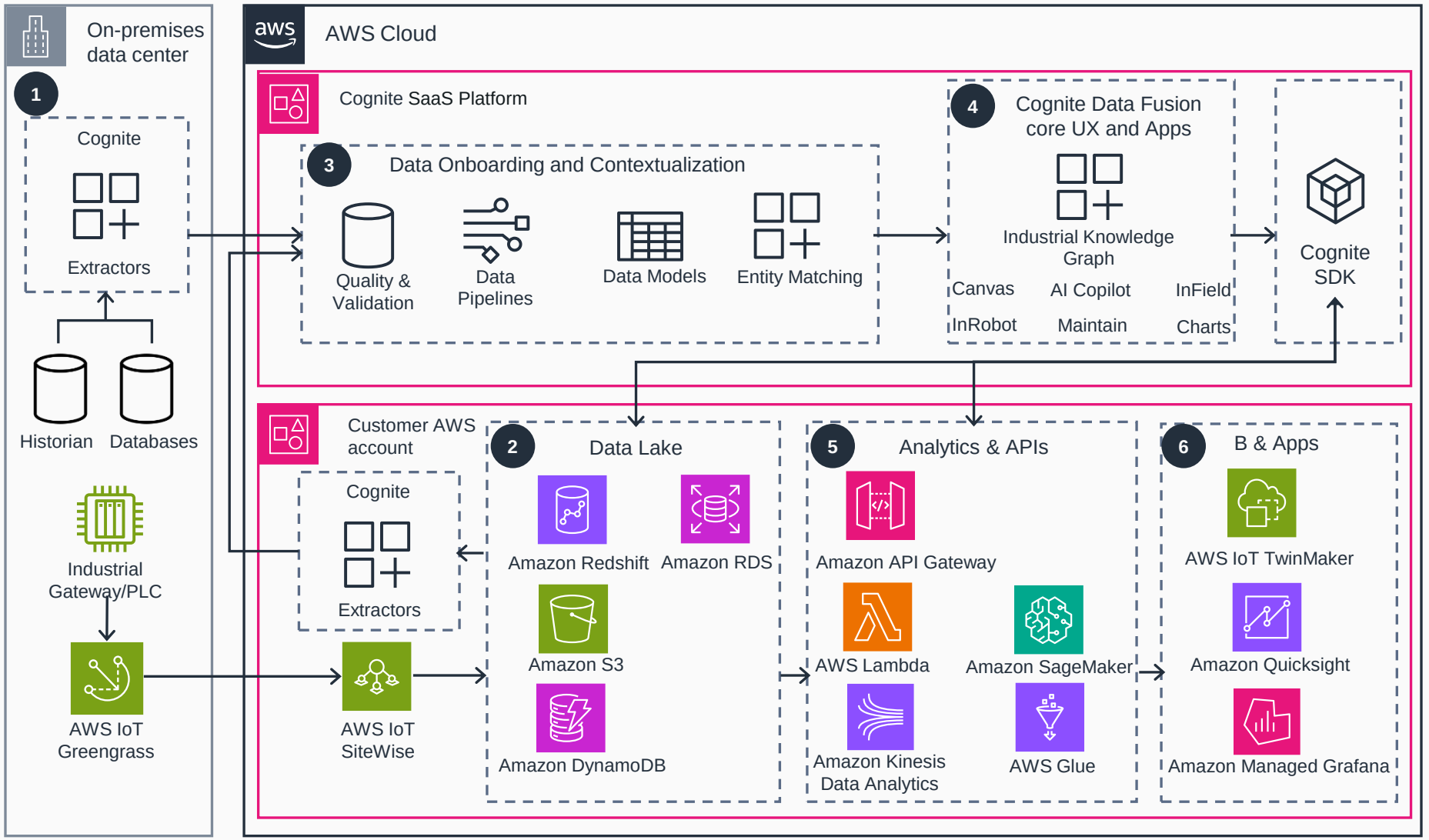
Guidance for Integrating an Industrial Data Fabric Using Cognite Data Fusion® on AWS

This architecture diagram demonstrates how to implement an industrial data fabric with Cognite Data Fusion on AWS. Cognite Data Fusion integrates, connects, and unifies the disparate industrial data sources into a cohesive, accessible data environment to enable data-driven insights and applications for industrial customers.



1. Purpose-built extractors ingest data from various industrial sources. Information technology (IT), operational technology (OT), and engineering data such as industrial historians and Programmable Logic Controllers (PLCs) are included. Additionally, this Guidance ingests data from smart sensors or gateways into the AWS customer's data lake through the integration of **AWS IoT Greengrass** and **AWS IoT SiteWise**.

2. In situations where customers already have their data in their AWS data lake, then Cognite Native Extractors for **Amazon Simple Storage Service (Amazon S3), Amazon Redshift**, **Amazon DynamoDB**, and **Amazon Relational Database Service (Amazon RDS)** ingest pre-aggregated data into Cognite Data Fusion.

3. Data onboarded into Cognite Data Fusion first undergoes a quality and validation check. Customers build a comprehensive and dynamic Industrial Knowledge Graph to deliver near real-time insights using generative AI-powered data pipelines and entity matching to create relationships between siloed industrial data at scale.

4. Cognite Data Fusion offers a collection of core user experiences (UX) and apps that use the Industrial Knowledge Graph built using contextualized industrial data. These apps include Canvas, AI Copilot, InField, InRobot, Maintain, and Charts designed to maximize production efficiency, ensure safe and sustainable operations, and enable high-quality, AI-powered business decisions.

5. Customers use **AWS Lambda** and **Amazon API Gateway** to ingest and transform data. **AWS Glue** is used to write data back into Cognite Data Fusion using the Cognite software development kit (SDK). Additional analytics and machine learning capabilities are provided by **Amazon Kinesis** Data Analytics and **Amazon SageMaker**, respectively.

6. **AWS IoT TwinMaker**, **Amazon Managed Grafana**, and **Amazon QuickSight** are used by end users to create applications. Data is consumed from Cognite Data Fusion using **Lambda** functions.

Reviewed for technical accuracy August 21, 2024

AWS Reference Architecture