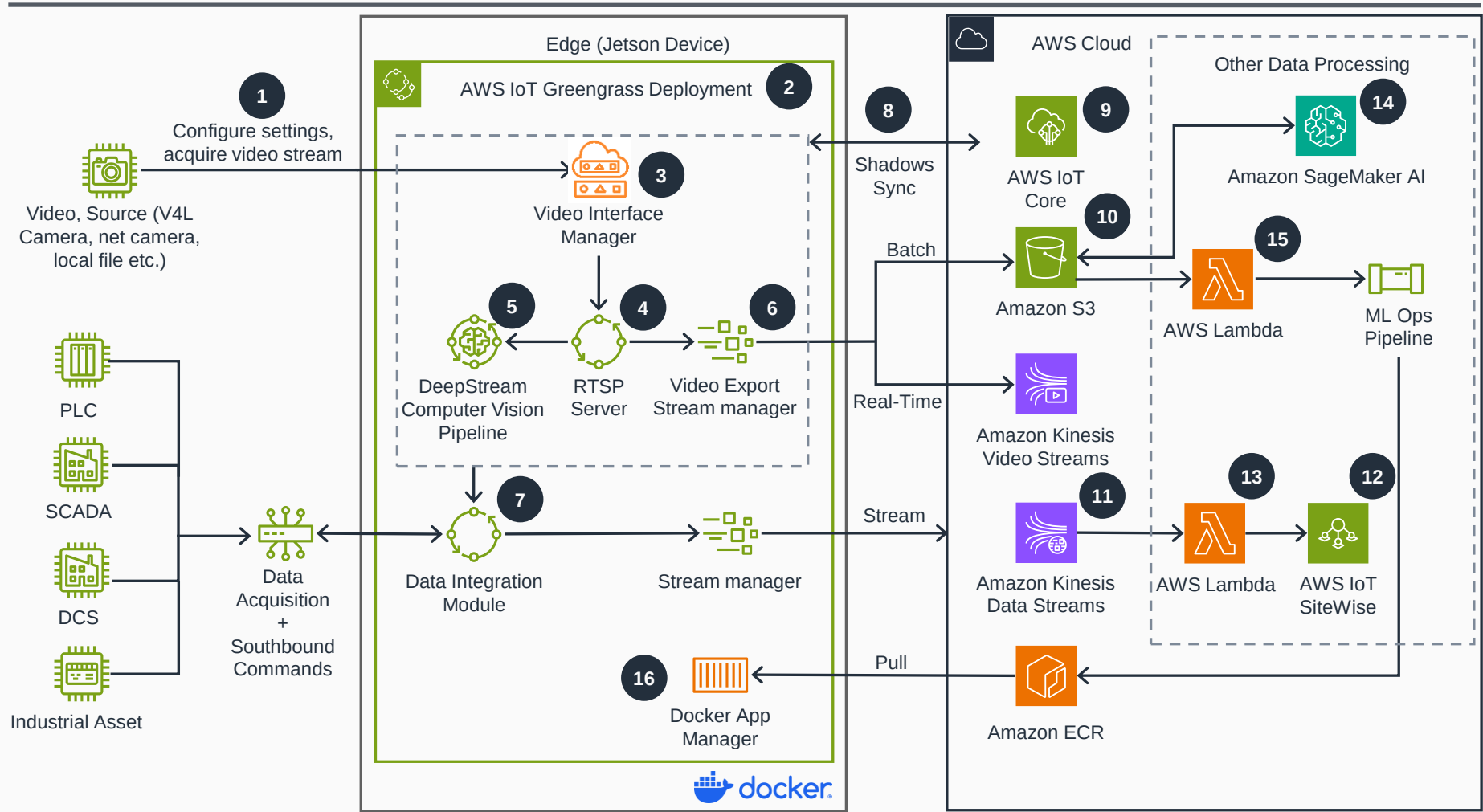


Guidance for Implementing AI-Powered Visual Quality Management with SoftServe EdgeInsight on AWS

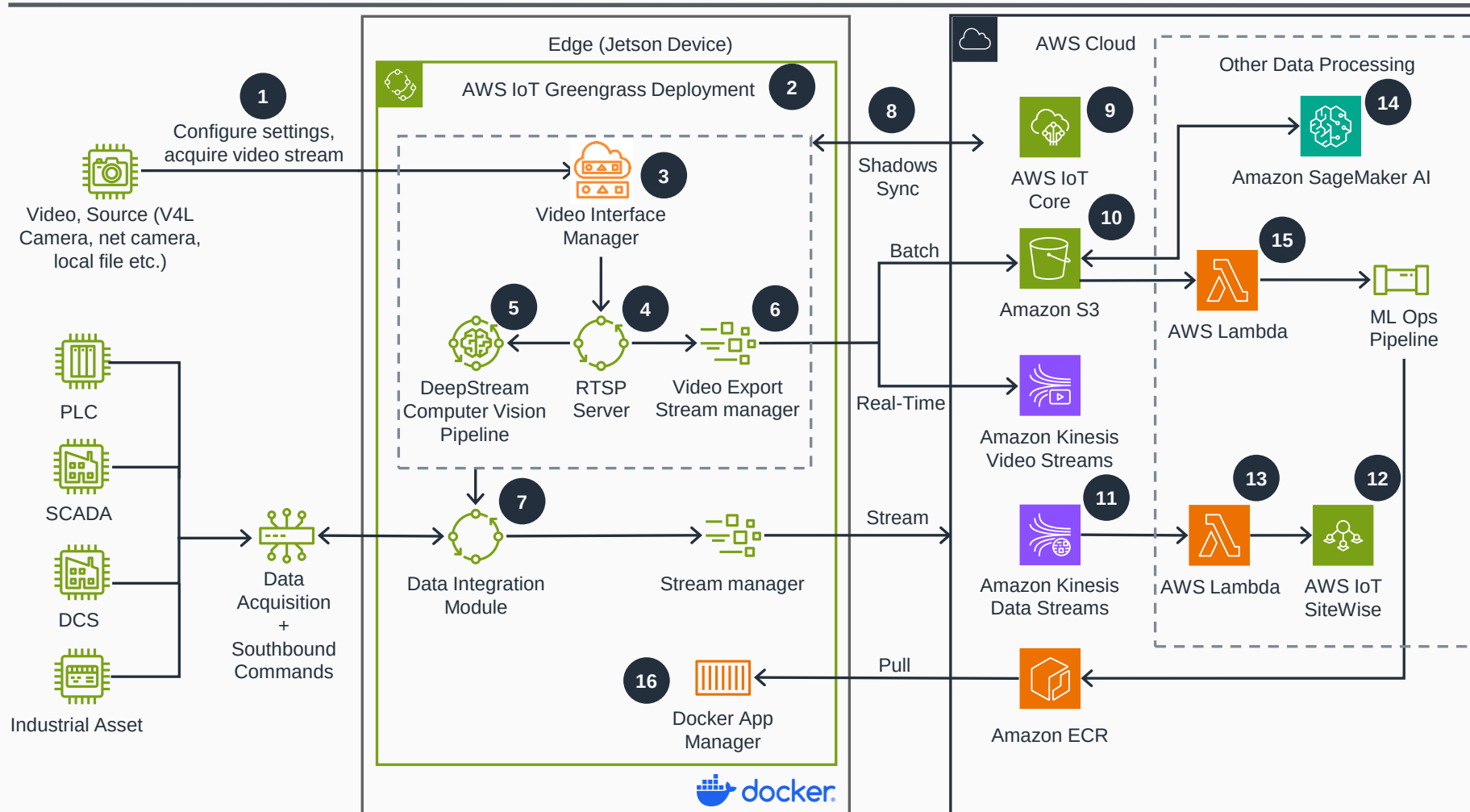
This architecture diagram shows how the EdgeInsight solution accelerator enables rapid deployment of AI-on-the-edge computer vision applications for near real-time industrial monitoring, analysis, and decision-making. This slide shows steps 1-5.



- 1 Connect to video feeds from factory-mounted cameras or Video Management Systems (VMS) for primary data input. Supplement training with synthetic data to expand test datasets and simulate rare defect patterns beyond typical factory scenarios.
- 2 Deploy **AWS IoT Greengrass** on an NVIDIA Jetson device or Industrial PC (IPC) with discrete GPU to run the edge data processing pipeline.
- 3 Create a Data Processing pipeline with Video Interface Manager to standardize interactions with Open Network Video Interface Forum (ONVIF) cameras, Real-Time Streaming Protocol (RTSP) streams, and GigE Vision systems. This abstraction layer eliminates the need to handle proprietary hardware protocols or vendor-specific specifications.
- 4 Implement an RTSP server as a video stream broker to manage complex routing of video streams. This enables multiple components to access and process video data simultaneously while maintaining system performance.
- 5 Deploy the NVIDIA DeepStream Computer Vision (DSCV) pipeline to analyze video in near real-time. This high-throughput, low-latency platform uses hardware acceleration to perform tasks from object detection to quality inspection.

Guidance for Implementing AI-Powered Visual Quality Management with SoftServe EdgeInsight on AWS

This architecture diagram shows how the EdgeInsight solution accelerator enables rapid deployment of AI-on-the-edge computer vision applications for near real-time industrial monitoring, analysis, and decision-making. This slide shows steps 6-10.

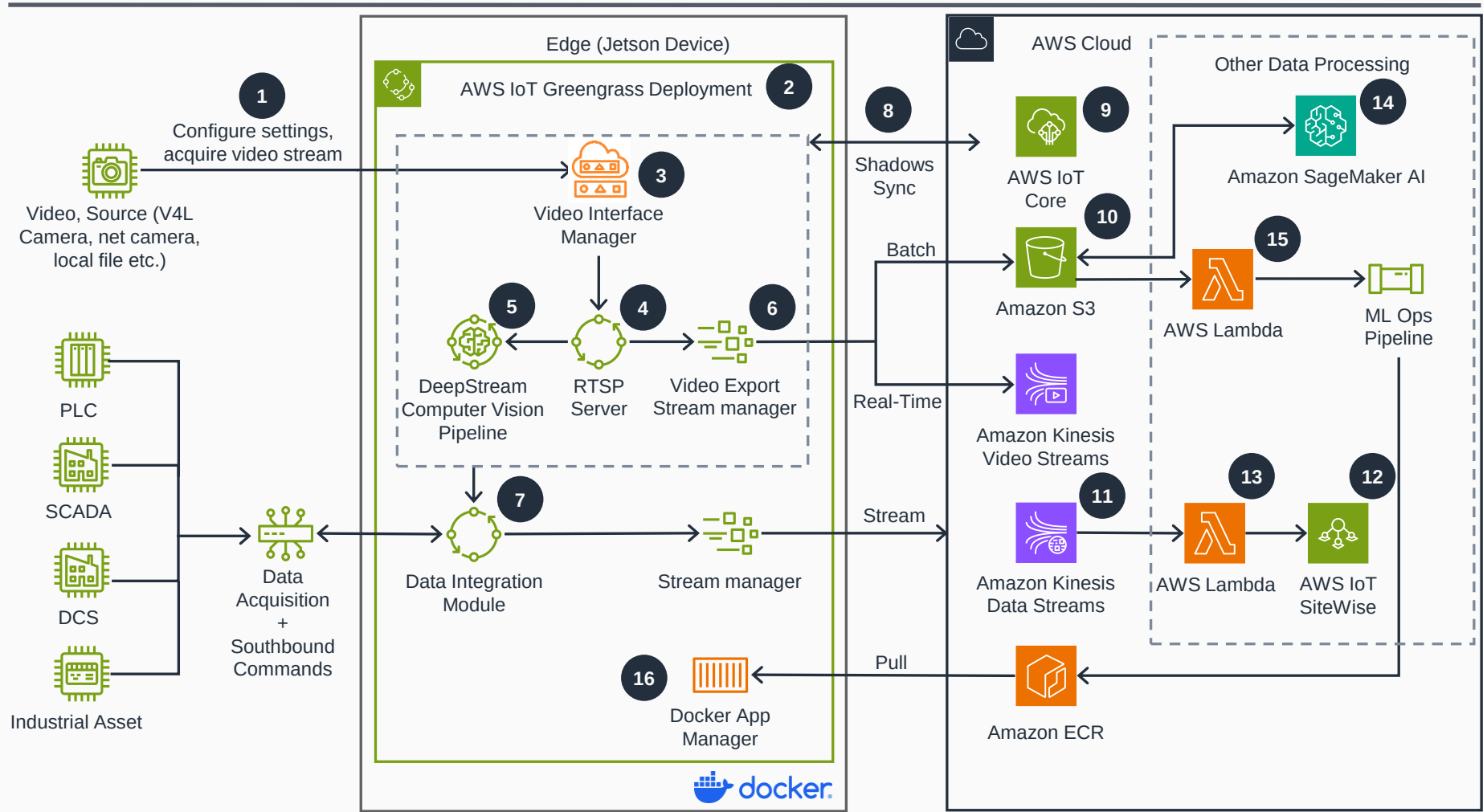


- 6 Use the Video Stream Export Manager (VSEM) to transfer processed or raw video to AWS Cloud storage and services through batch or near real-time operations.
- 7 Configure the Data Integration Module to process inference data and execute business logic at the edge. Transform raw DeepStream CV data through filtering, aggregation, and enrichment. Combine video analytics with OT process data from Programmable Logic Controllers (PLC), Distributed Control System (DCS), and Supervisory Control and Data Acquisition (SCADA) systems for comprehensive context, such as correlating conveyor state data with object detection.
- 8 Implement near real-time and batch video/data stream processing through interconnected services for seamless edge-to-cloud operations using **AWS IoT Core**, **Amazon Simple Storage Service (Amazon S3)**, **Amazon Kinesis Video Streams**, **Amazon Kinesis Data Streams**, and **AWS IoT SiteWise**.
- 9 Set up **AWS IoT Core** for secure bidirectional communication with edge compute devices. Use device shadows to maintain configuration synchronization during intermittent connectivity or partially air-gapped factory operations.
- 10 Use **Amazon S3** to store video artifacts and ML models for edge deployment, providing scalable object storage.



Guidance for Implementing AI-Powered Visual Quality Management with SoftServe EdgeInsight on AWS

This architecture diagram shows how the EdgeInsight solution accelerator enables rapid deployment of AI-on-the-edge computer vision applications for near real-time industrial monitoring, analysis, and decision-making. This slide shows steps 11-15.

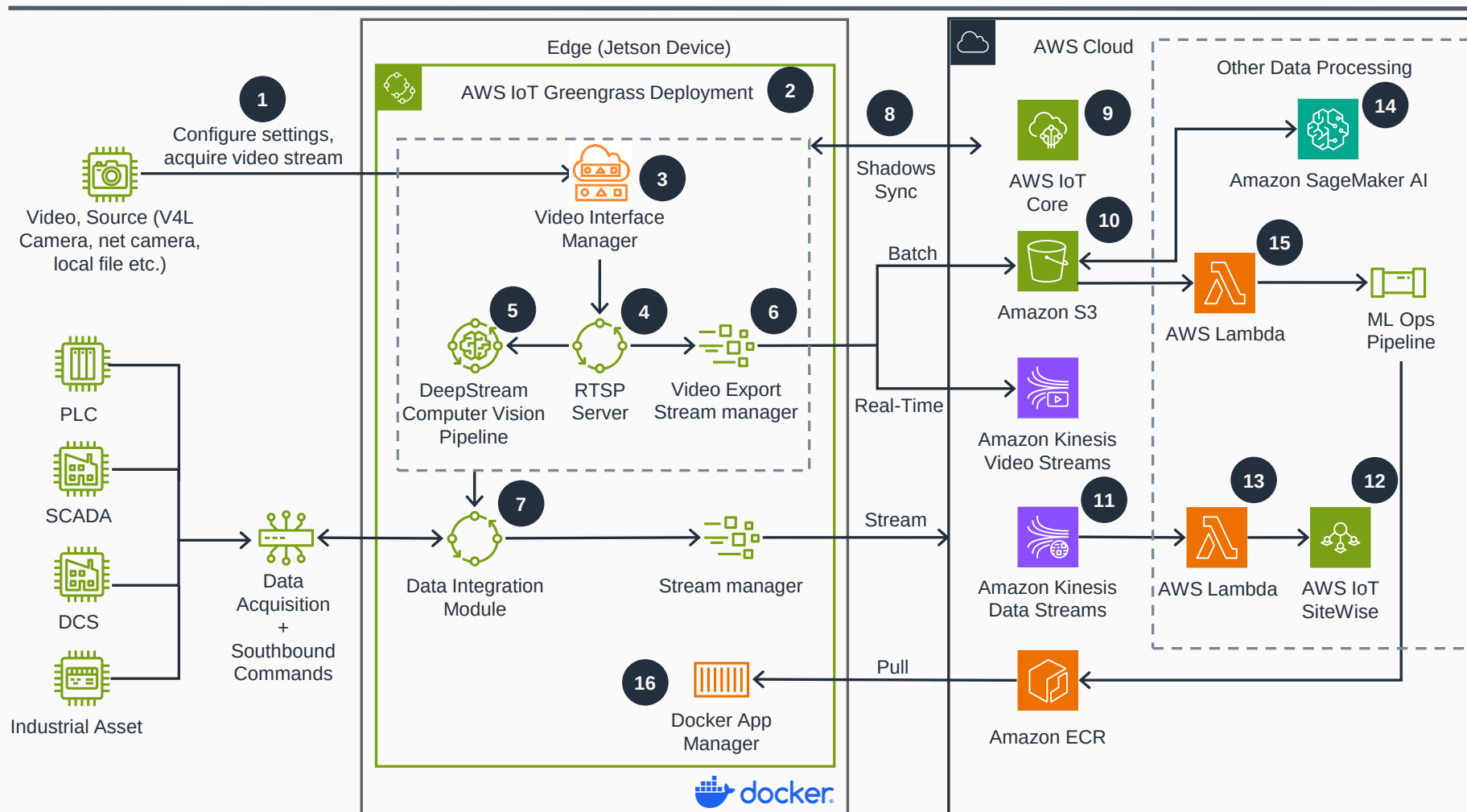


- 11 Use **Amazon Kinesis Data Streams** and **Amazon Kinesis Video Streams** to ingest real-time video and inference metadata.
- 12 Configure **AWS IoT SiteWise** to create digital representations of industrial data hierarchies. Enable production managers to correlate defect rates with machine parameters, linking visual defects to specific temperature, vibration, or other operational fluctuations.
- 13 Use **AWS Lambda** as a connective layer between services to process edge-generated data and cloud operations. Transform data between services, extract video metadata, process inference outputs, trigger container deployments, and manage over-the-air update notifications and verifications.
- 14 Deploy **Amazon SageMaker AI** as the ML foundation for EdgeInsight's AI-on-the-Edge computer vision continuous improvement cycle. Data scientists use **Amazon SageMaker AI** to train models using real or NVIDIA Omniverse-generated synthetic data, then store trained models in **Amazon S3**.
- 15 Trigger ML Ops pipeline using **AWS Lambda** to package ML models as Docker containers and store them in **Amazon Elastic Container Registry (Amazon ECR)**.



Guidance for Implementing AI-Powered Visual Quality Management with SoftServe EdgeInsight on AWS

This architecture diagram shows how the EdgeInsight solution accelerator enables rapid deployment of AI-on-the-edge computer vision applications for near real-time industrial monitoring, analysis, and decision-making. This slide shows step 16.



16 Configure Docker Application Manager to pull images from **Amazon ECR** and deploy software updates and optimized **Amazon SageMaker AI** MLOps models to edge devices, ensuring secure and efficient delivery of improvements.



Reviewed for technical accuracy October 9, 2025
© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Reference Architecture