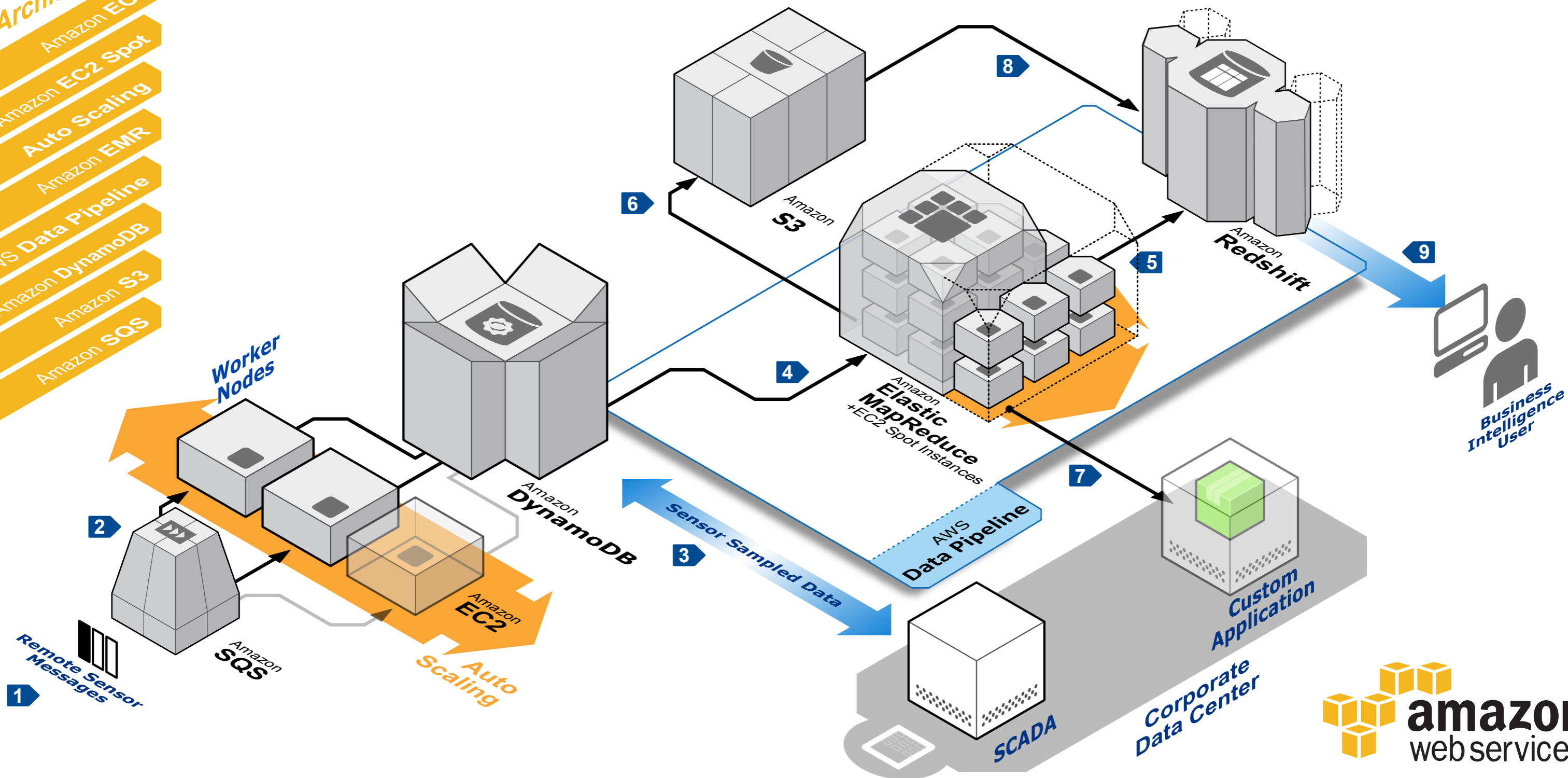# TIME SERIES PROCESSING

When data arrives as a succession of regular measurements, it is known as time series information. Processing of time series information poses systems scaling challenges that the elasticity of AWS services is uniquely positioned to address.

This elasticity is achieved by using Auto Scaling groups for ingest processing, AWS Data Pipeline for scheduled Amazon Elastic MapReduce jobs, AWS Data Pipeline for intersystem data orchestration, and Amazon Redshift for potentially massive-scale analysis. Key architectural throttle points involving Amazon SQS for sensor message buffering and less frequent AWS Data Pipeline scheduling keep the overall solution costs predictable and controlled.

Worker Nodes

Amazon DynamoDB

Amazon S3

Amazon Elastic MapReduce +EC2 Spot Instances

Amazon Redshift

Business Intelligence User

Remote Sensor Messages

Amazon SQS

Amazon EC2

Auto Scaling

Sensor Sampled Data

AWS Data Pipeline

SCADA

Custom Application

Corporate Data Center

amazon web services™

System Overview

1 Remote devices such as power meters, mobile clients, ad-network clients, industrial meters, satellites, and environmental meters measure the world around them and send sampled sensor data as messages via HTTP(S) for processing.

2 Send messages to an **Amazon Simple Queue Service** queue for processing into Amazon DynamoDB using autoscaled **Amazon EC2** workers. Or, if the sensor source can do so, post sensor samples directly to **Amazon DynamoDB**. Try starting with a DynamoDB table that is a week-oriented, time-based table structure.

3 If a Supervisory Control and Data Acquisition (SCADA) system exists, create a flow of samples to or from **Amazon DynamoDB** to support additional cloud processing or other existing systems, respectively.

4 Using **AWS Data Pipeline**, create a pipeline with a regular **Amazon Elastic MapReduce** job that both calculates expensive sample processing and delivers samples and results.

5 The pipeline places results into **Amazon Redshift** for additional analysis.

6 The pipeline exports historical week-oriented sample tables, from **Amazon DynamoDB** to **Amazon Simple Storage Service** (Amazon S3)

7 The pipeline also optionally exports results in a format custom applications can accept.

8 **Amazon Redshift** optionally imports historic samples to reside with calculated results.

9 Using in-house or Amazon partner business intelligence solutions, **Amazon Redshift** supports additional analysis on a potentially massive scale.