

IoT and data analytics in the construction industry

08/29/2018 • 5 minutes to read • Contributors 

In this article

[Relevant use cases](#)

[Architecture](#)

[Considerations](#)

[Pricing](#)

[Related resources](#)

This example scenario is relevant to organizations building solutions that integrate data from many IoT devices into a comprehensive data analysis architecture to improve and automate decision making. Potential applications include construction, mining, manufacturing, or other industry solutions involving large volumes of data from many IoT-based data inputs.

In this scenario, a construction equipment manufacturer builds vehicles, meters, and drones that use IoT and GPS technologies to emit telemetry data. The company wants to modernize their data architecture to better monitor operating conditions and equipment health. Replacing the company's legacy solution using on-premises infrastructure would be both time and labor intensive, and would not be able to scale sufficiently to handle the anticipated data volume.

The company wants to build a cloud-based "smart construction" solution. It should gather a comprehensive set of data for a construction site and automate the operation and maintenance of the various elements of the site. The company's goals include:

- Integrating and analyzing all construction site equipment and data to minimize equipment downtime and reduce theft.
- Remotely and automatically controlling construction equipment to mitigate the effects of a labor shortage, ultimately requiring fewer workers and enabling lower-skilled workers to succeed.
- Minimizing the operating costs and labor requirements for the supporting infrastructure, while increasing productivity and safety.
- Easily scaling the infrastructure to support increases in telemetry data.
- Complying with all relevant legal requirements by provisioning resources in-country without compromising system availability.
- Using open-source software to maximize the investment in workers' current skills.

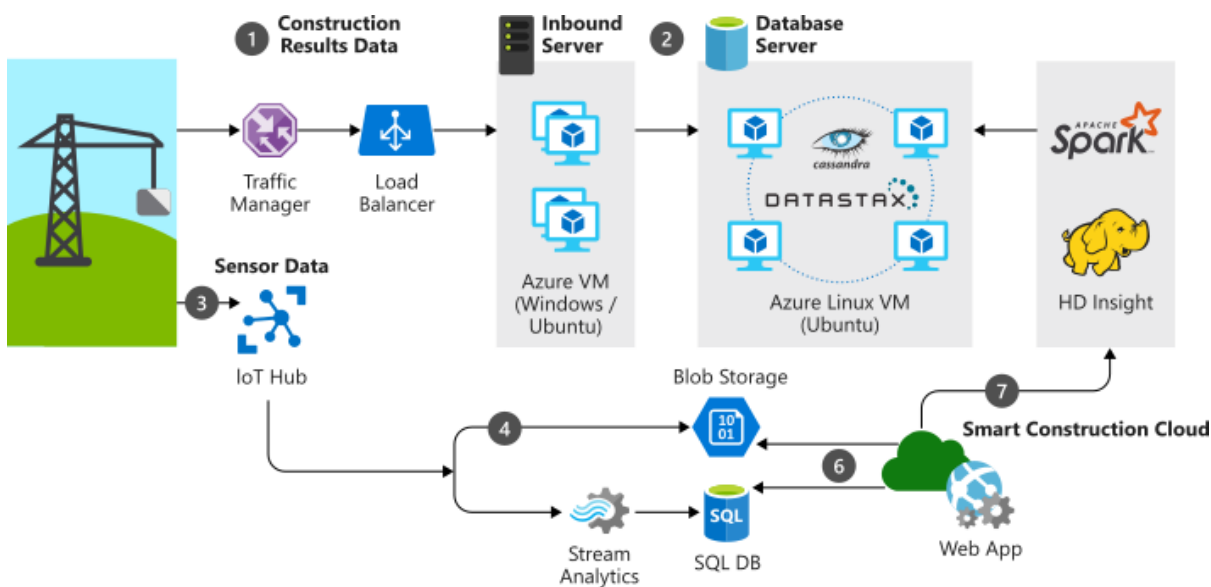
Using managed Azure services such as IoT Hub and HDInsight will allow the customer to rapidly build and deploy a comprehensive solution with a lower operating cost. If you have additional data analytics needs, you should review the list of available [fully managed data analytics services in Azure](#).

Relevant use cases

Other relevant use cases include:

- Construction, mining, or equipment manufacturing scenarios
- Large-scale collection of device data for storage and analysis
- Ingestion and analysis of large datasets

Architecture



The data flows through the solution as follows:

1. Construction equipment collects sensor data and sends the construction results data at regular intervals to load balanced web services hosted on a cluster of Azure virtual machines.
2. The custom web services ingest the construction results data and store it in an Apache Cassandra cluster also running on Azure virtual machines.
3. Another dataset is gathered by IoT sensors on various construction equipment and sent to IoT Hub.
4. Raw data collected is sent directly from IoT Hub to Azure blob storage and is immediately available for viewing and analysis.
5. Data collected via IoT Hub is processed in near real time by an Azure Stream Analytics job and stored in an Azure SQL database.
6. The Smart Construction Cloud web application is available to analysts and end users to view and analyze sensor data and imagery.
7. Batch jobs are initiated on demand by users of the web application. The batch job runs in Apache Spark on HDInsight and analyzes new data stored in the Cassandra cluster.

Components

- **IoT Hub** acts as a central message hub for secure bi-directional communication with per-device identity between the cloud platform and the construction equipment and other site elements. IoT Hub can rapidly collect data for each device for ingestion into the data analytics pipeline.
- **Azure Stream Analytics** is an event-processing engine that can analyze high volumes of data streaming from devices and other data sources. It also supports extracting information from data streams to identify patterns and relationships. In this scenario, Stream Analytics ingests and analyzes data from IoT devices and stores the results in Azure SQL Database.
- **Azure SQL Database** contains the results of analyzed data from IoT devices and meters, which can be viewed by analysts and users via an Azure-based Web application.
- **Blob storage** stores image data gathered from the IoT hub devices. The image data can be viewed via the web application.
- **Traffic Manager** controls the distribution of user traffic for service endpoints in different Azure regions.
- **Load Balancer** distributes data submissions from construction equipment devices across the VM-based web services to provide high availability.
- **Azure Virtual Machines** host the web services that receive and ingest the construction results data into the Apache Cassandra database.
- **Apache Cassandra** is a distributed NoSQL database used to store construction data for later processing by Apache Spark.

- [Web Apps](#) hosts the end-user web application, which can be used to query and view source data and images. Users can also initiate batch jobs in Apache Spark via the application.
- [Apache Spark on HDInsight](#) supports in-memory processing to boost the performance of big-data analytic applications. In this scenario, Spark is used to run complex algorithms over the data stored in Apache Cassandra.

Alternatives

- [Cosmos DB](#) is an alternative NoSQL database technology. Cosmos DB provides [multi-master support at global scale](#) with [multiple well-defined consistency levels](#) to meet various customer requirements. It also supports the [Cassandra API](#).
- [Azure Databricks](#) is an Apache Spark-based analytics platform optimized for Azure. It is integrated with Azure to provide one-click setup, streamlined workflows, and an interactive collaborative workspace.
- [Data Lake Storage](#) is an alternative to Blob storage. For this scenario, Data Lake Storage was not available in the targeted region.
- [Web Apps](#) could also be used to host the web services for ingesting construction results data.
- Many technology options are available for real-time message ingestion, data storage, stream processing, storage of analytical data, and analytics and reporting. For an overview of these options, their capabilities, and key selection criteria, see [Big data architectures: Real-time processing](#) in the [Azure Data Architecture Guide](#).

Considerations

The broad availability of Azure regions is an important factor for this scenario. Having more than one region in a single country can provide disaster recovery while also enabling compliance with contractual obligations and law enforcement requirements. Azure's high-speed communication between regions is also an important factor in this scenario.

Azure support for open-source technologies allowed the customer to take advantage of their existing workforce skills. The customer can also accelerate the adoption of new technologies with lower costs and operating workloads compared to an on-premises solution.

Pricing

The following considerations will drive a substantial portion of the costs for this solution.

- Azure virtual machine costs will increase linearly as additional instances are provisioned. Virtual machines that are deallocated will only incur storage costs, and not compute costs. These deallocated machines can then be reallocated when demand is high.
- [IoT Hub](#) costs are driven by the number of IoT units provisioned as well as the service tier chosen, which determines the number of messages per day per unit allowed.
- [Stream Analytics](#) is priced by the number of streaming units required to process the data into the service.

Related resources

To see an implementation of a similar architecture, read the [Komatsu customer story](#).

Guidance for big data architectures is available in the [Azure Data Architecture Guide](#).