Microsoft Azure - Starter Kits for Partners

Azure Assessment

Intelligent Apps & Analytics

Last Update: August 2016





**MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.**

The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The descriptions of other companies’ products in this document, if any, are provided only as a convenience to you. Any such references should not be considered an endorsement or support by Microsoft. Microsoft cannot guarantee their accuracy, and the products may change over time. Also, the descriptions are intended as brief highlights to aid understanding, rather than as thorough coverage. For authoritative descriptions of these products, please consult their respective manufacturers.

© 2014 Microsoft Corporation. All rights reserved. Any use or distribution of these materials without express authorization of Microsoft Corp. is strictly prohibited.

Microsoft and Windows are either registered trademarks of Microsoft Corporation in the United States and/or other countries.

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

Contents

[Overview 5](#_Toc460594360)

[Azure Scenario Assessment 6](#_Toc460594361)

[Microsoft’s Data Platform: The Big Picture 6](#_Toc460594362)

[Business intelligence 7](#_Toc460594363)

[Technology Snapshot: Power BI 8](#_Toc460594364)

[Technology Snapshot: SQL Server Analysis Services 8](#_Toc460594365)

[Understanding Your Options 8](#_Toc460594366)

[Big data and analytics 9](#_Toc460594367)

[Technology Snapshot: Analytics Platform System 10](#_Toc460594368)

[Technology Snapshot: Azure SQL Data Warehouse 10](#_Toc460594369)

[Technology Snapshot: SQL Server Integration Services 10](#_Toc460594370)

[Understanding Your Options 11](#_Toc460594371)

[Remote monitoring with IoT 11](#_Toc460594372)

[Technology Snapshot: Azure Stream Analytics 12](#_Toc460594373)

[Technology Snapshot: HDInsight Storm 12](#_Toc460594374)

[Technology Snapshot: HDInsight Spark Streaming 12](#_Toc460594375)

[Technology Snapshot: Azure IoT Hub 13](#_Toc460594376)

[Understanding Your Options 13](#_Toc460594377)

[Predictive maintenance with IoT 14](#_Toc460594378)

[Technology Snapshot: Azure Blobs 14](#_Toc460594379)

[Understanding Your Options 14](#_Toc460594380)

[Cognitive Services 15](#_Toc460594381)

[Questions 16](#_Toc460594382)

[Business Needs 16](#_Toc460594383)

[Information Architecture 16](#_Toc460594384)

[Technical Architecture 17](#_Toc460594385)

[Organization 17](#_Toc460594386)

[Big Data Maturity Level 18](#_Toc460594387)

[IoT Service 21](#_Toc460594388)

[Resources and Tools 28](#_Toc460594389)

[Big data and analytics 28](#_Toc460594390)

[Business intelligence 28](#_Toc460594391)

[Data warehouse 28](#_Toc460594392)

[Remote monitoring with IoT 28](#_Toc460594393)

[Predictive maintenance with IoT 28](#_Toc460594394)

[Cognitive Services 28](#_Toc460594395)

[Microsoft Azure IoT Starter Kits 28](#_Toc460594396)

# Overview

The purpose of this document is to provide Microsoft Partners with an assessment to identify key components in the customer scenario and serve as a guide to the available resources. This will help the partner to build an efficient architecture for the customer scenario and have an accurate cost proposal based in the customer needs.

In this document, we will cover the following topics:

* Common Scenarios
* Questionnaire
* Resources and Tools
* FAQ

# Azure Scenario Assessment

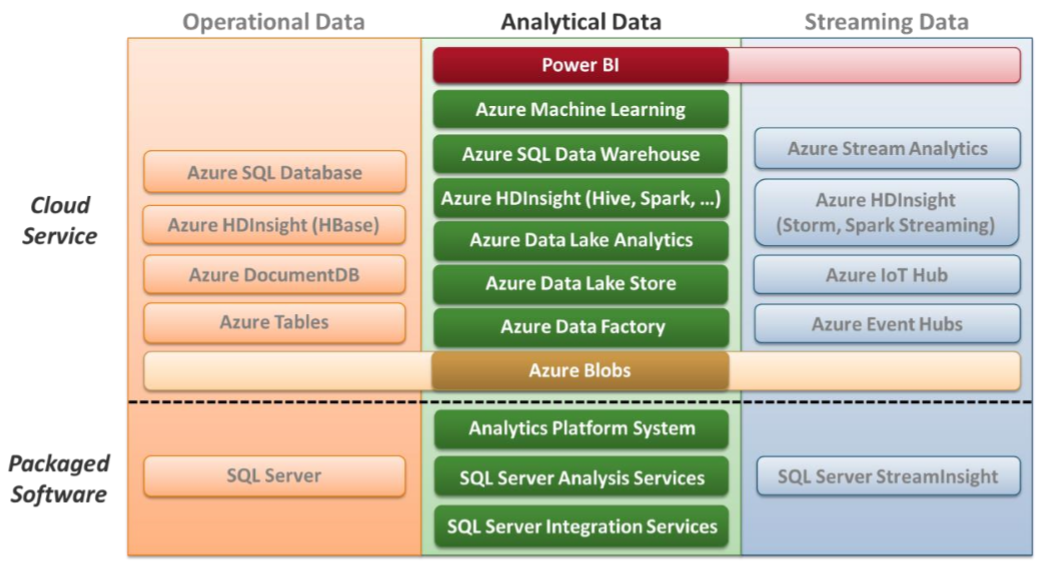
## Microsoft’s Data Platform: The Big Picture

We use data in many different ways. The volume, variety, and velocity of that data increases every day. Due to this, organizations rely on a handful of different data technologies. Taken as a group, these technologies make up a data platform.

One way to think about the technologies in a data platform is to divide them into three categories based on the kind of data they work with. Those categories are:

* Operational data, such as transactional data used by a banking system, an online retailer, or an ERP application. This data is typically both read and written by applications, commonly in response to user requests. A banking application might read your account balance, for instance, then write a new value to reflect a deposit you make. While operational data was once almost entirely relational, the increasing volume and variety of data have changed this. Today, working with unstructured, operational data can be just as important.
* Analytical data, such as the information kept in a data warehouse. This data is typically read-only and usually includes historical information extracted over time from other data sources, such as operational databases. Analytical data is commonly used for things such as business intelligence and machine learning, and like operational data, it can be either relational or unstructured.
* Streaming data, such as data produced by sensors. The defining characteristic of streaming data is velocity; if the data isn’t processed quickly, it can lose a large share of its value. Many streaming scenarios today relate to the Internet of Things (IoT), where the focus is on interacting with data provided by a lot of devices. Streaming data is also used in other situations, such as analyzing financial transactions as they happen. In both cases, the challenge is to work effectively with large amounts of data being produced in real time.

With the latest Microsoft Intelligent Apps & Analytics technology, you gain greater business insights without dramatically increasing IT costs and with a very positive user experience. It can give you the ability to control cost, find new opportunities, and carry out your vision. The following figure summarizes the Microsoft’s offerings in each area.



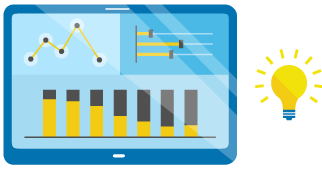
Microsoft Intelligent Apps & Analytics allows you to effectively drive your business processes with familiar tools utilizing extended capabilities for data analytics and reporting. These tools can bring you many benefits such as:

* Empower users to discover, analyze, and visualize data with powerful BI tools.
* Easily deliver dashboards that aggregate data from multiple sources and track success metrics aligned to business strategies.
* Provide integrated ad-hoc data exploration and visualization to explore root cause.
* Address complex reporting needs with professional reporting tools.
* Quickly deploy a cloud-based business intelligence solution that provides an environment for rich analysis, collaboration, and governance.
* Improved user productivity through familiar tools in Excel and SQL Server.
* Take advantage of enterprise-class capabilities like high availability, superior performance, and scalability, without the up-front cost of infrastructure.

Along the way, we’ll take a brief look at each of the analytical data technologies shown in the above figure. The goal is to provide a big-picture view of how the Microsoft data platform addresses the challenges of working with analytical data.

## Business intelligence

Transform your company’s data to support informed decision-making. Gain deeper insight into your data to stay in the know and spot trends as they happen. With modern business intelligence, your entire organization can understand and quickly act on data.



**Get the right insights into the right hands**

Offer business analysts, and everyone in your organization, powerful, self-service analytical tools to drive better, faster decision-making. Combine data from multiple sources to build tailored reports and create rich analytics that bring your data to life.

**Integrate and manage business intelligence securely**

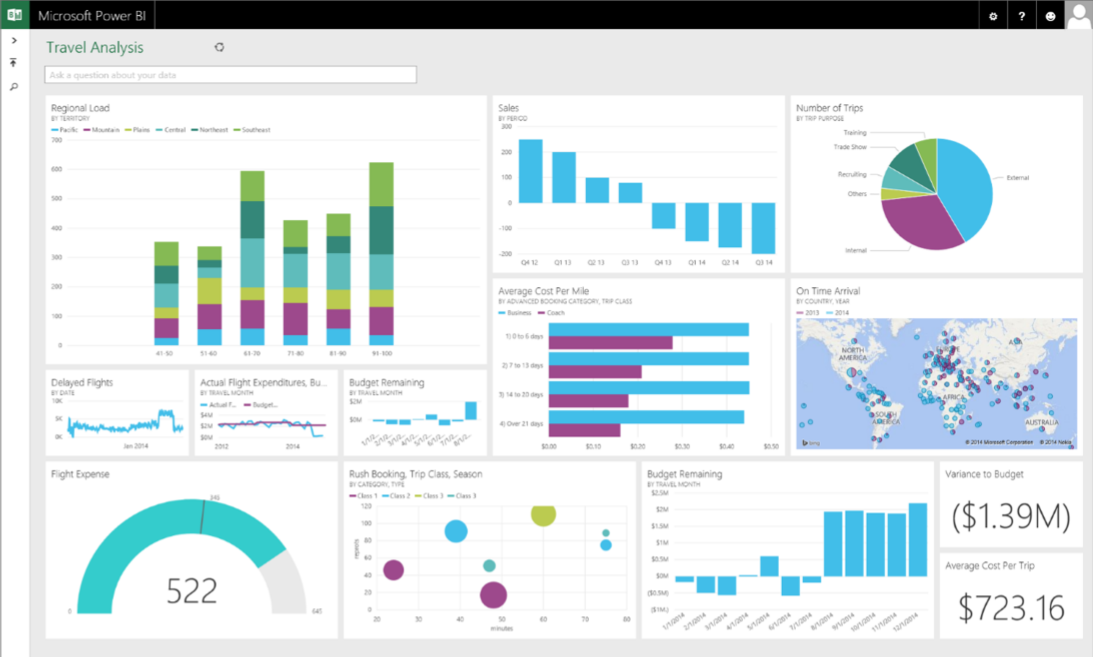
Better integrate with existing systems by choosing a feature-rich, enterprise-ready BI platform. Maximize resources, monitor access to data and assets, help ensure security and compliance, and deliver a business intelligence solution designed for the needs of your organization.

**Bring data to life in your apps**

Help your customers easily access valuable data on any device at any time. Embed fully interactive and up-to-date visual analytics directly in your apps with ease—without the time and expense of writing code.

### Technology Snapshot: Power BI

Power BI is a cloud-based service that lets users access diverse data from anywhere. It can present up-to-the-minute views of data from many different sources and then make those views accessible on desktops and mobile devices, including iOS and Android phones. The sources of data can include on-premises analysis technologies, analysis services that run in the cloud, and cloud applications from Microsoft and other vendors. The figure below shows an example of a Power BI interface.



As this example suggests, Power BI can display information from many different sources in a unified way. Just as important, business users can use Power BI to define these interfaces and reports themselves—they don’t need to rely on developers. Power BI also provides pre-built dashboards and reports for Office 365, Salesforce.com CRM, and other cloud applications. The tool supports natural language query as well, letting you ask questions such as “What are the total sales by the hour for diapers as a line chart?”, then get back a graphical answer. All of these things have a common goal: providing a modern UI for accessing diverse data from anywhere.

### Technology Snapshot: SQL Server Analysis Services

If you’re like most IT leaders, you know that cloud computing will play a bigger role in your organization’s future, but you also know that on-premises technologies will be important for many years to come.

Data analysis technologies provide a good example of this. Today, many organizations store periodic snapshots of operational data in on-premises data warehouses, and then create business intelligence (BI) applications to analyze this data. In the Microsoft data platform, the fundamental technology for doing this is SQL Server. This relational database lets its users create data warehouses, and then analyze the data they contain using SQL Server Analysis Services (SSAS). SSAS is a mature offering—it was first released in 1998—and it supports online analytical processing (OLAP), data mining, and more.

### Understanding Your Options

IT leaders around the world face a common challenge: How should their organization adopt cloud technology? The cloud certainly has benefits, but it also has some clear risks.

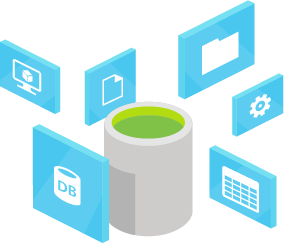
Using Power BI can be a good place to start. The value of a common UI has obvious appeal—it can make your users happier. Starting here also lets you provide a widely accessible interface in the cloud while leaving critical data on premises, an approach that can minimize both regulatory concerns and your sense of risk.

Another way to get started with data analysis in the cloud is to do a new SSAS project on Microsoft Azure. Azure provides a technology called infrastructure as a service (IaaS) that lets you create virtual machines (VMs) on demand in Microsoft datacenters. It’s possible to install SSAS and other software on these VMs, then run the environment much like your on-premises system. You might do this to save money, to get faster access to computing resources, or both.

Whatever approach you choose, one thing is clear: Microsoft’s focus, in data analysis and other areas, is moving to the cloud. If you’re a Microsoft customer, finding a way to adapt to this change should be a high priority for your IT organization.

## Big data and analytics

Deliver better experiences and make better decisions by analyzing massive amounts of data in real time. Get the insight you need to deliver intelligent actions that improve customer engagement, increase revenue, and lower costs.

**Bring together all of the data you need**

Data volumes are exploding – from traditional point-of-sale systems and e-commerce websites, to new customer sentiment sources such as Twitter and IoT sensors that stream data in real time using Apache, Hadoop, and Spark. By analyzing a diverse dataset from the start, you’ll make more informed decisions that are predictive and holistic, rather than reactive and disconnected.

**Hold onto your most valuable asset – data**

Keep your organization’s data indefinitely, no matter the size. Instead of making cost trade-offs on what data to hold onto, retain your data to meet regulatory and company standards at affordable prices – now possible with Hadoop, Spark technologies, and the cloud.

**Deliver a personalized experience to customers**

Different people want different experiences. Delight your customers with a personalized experience that changes based on their behavior, even offering recommended products that include dynamic discounts for a personalized shopping experience. Give suppliers a predictive list of things to purchase based on current order information and historic customer data.

**Create a more cost-effective supply chain**

Integrate data from across the enterprise value chain and analyze it in real time to optimize supply-side performance and save money. Embrace proactive measures with a live view into your supply chain – assess inventory levels, predict product fulfilment needs, and identify potential backlog issues.

**Be more efficient in everything you do**

Uncover insights buried in your data to optimize the way you do business. Whether it is organizing human resources, managing supply chains, or forecasting staff and customer needs, understanding the factors that affect operational efficiency is essential to streamlining your business.

### Technology Snapshot: Analytics Platform System

Many organizations today successfully use on-premises data warehouses created with SQL Server, but what if your situation requires handling many terabytes or even a few petabytes of relational data? APS is designed for scenarios like this.

APS is a dedicated hardware appliance that runs in your own datacenter and it can handle petabytes of data. The appliance contains multiple physical servers, with the hardware supplied by Dell, HP, or another vendor. Applications running on APS use MPP, which lets them exploit the processing power of the appliance’s multiple servers.

Yet in many organizations, the lion’s share of their new data isn’t relational—it’s unstructured. For analyzing large amounts of unstructured data, the industry standard has become the Hadoop technology family. To let you work with both relational and unstructured data, APS also allows creating a Hadoop partition within the appliance.

Combining relational and unstructured data raises another question: How can an application issue a query against both? With APS, the answer is a technology called PolyBase. Using this technology, an application can issue standard T-SQL queries against relational data in APS, non-relational data in APS, or both, then let PolyBase handle the details of getting the result. Among other things, this lets users work with APS data from common tools such as Excel.

### Technology Snapshot: Azure SQL Data Warehouse

APS lets you analyze large amounts of data in an on-premises appliance, but more and more of the data that you want to work with lives in the cloud. For instance, maybe that data is created by a customer-facing web application running on Azure or perhaps it’s coming from devices in an IoT scenario that uses Azure as a back end. Whatever the source, the problem is to store and analyze very large amounts of data in the cloud. To help you do this, the Microsoft data platform provides Azure SQL Data Warehouse.

To a great degree, SQL Data Warehouse replicates the functionality of APS in the cloud. Like APS, it can store large amounts of relational data, then let applications use MPP to execute high-performance queries across that data. It also supports PolyBase, letting you issue T-SQL queries across both relational and unstructured Hadoop data.

SQL Data Warehouse has an important difference from APS, however. APS is a physical appliance, which implies that you must choose the size you need when you buy the hardware. SQL Data Warehouse is a cloud service, so you can increase or decrease the processing resources you use as your needs change and since it’s a cloud service, you only pay for the resources you actually use.

### Technology Snapshot: SQL Server Integration Services

To create and maintain a data warehouse, organizations regularly pull data into the warehouse from operational databases. The warehouse can be built using SQL Server, APS, SQL Data Warehouse, or many other technologies. The operational databases that provide the source data might use SQL Server, Oracle, a NoSQL technology, or something else. Whatever the specifics, the process is commonly called extract, transform, and load (ETL), and it’s usually automated.

SQL Server Integration Services (SSIS) is a technology for executing ETL and more. It can be used with many different data technologies, including those just listed, and it provides a drag-and-drop interface for defining data workflows. Like SSAS, SSIS is included with SQL Server and it’s become a widely used tool for data integration.

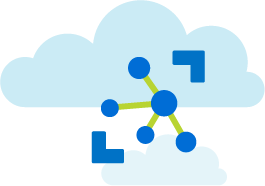
### Understanding Your Options

Using either APS or Azure SQL Data Warehouse can make sense whenever using MPP makes sense. There are also cases where using the two together can be helpful, such as these:

* Aging data from on-premises storage to cloud storage. Suppose you have an on-premises application that needs the scale and MPP power of APS. Over time, the amount of data you need to store might outstrip even the capacity of this appliance. In a situation like this, you might choose to move older data that’s accessed less frequently to Azure SQL Data Warehouse. This would likely make storage cheaper while still letting you get at this aged data from your existing MPP applications.
* Application development and testing in the cloud. While there are some differences between APS and Azure SQL Data Warehouse, the two provide similar services. Because of this, it’s possible to create new MPP applications in the cloud, then run them on premises. This avoids the risk of development projects interfering with a production APS environment. It can also give development groups more control over the environment they work in, since the team can create and use its own instance of Azure SQL Data Warehouse.
* Disaster recovery in the cloud. Suppose your organization has created one or more mission-critical applications using APS. In cases like this, having a disaster recovery solution is essential. What if an on-premises datacenter goes down because of a flood, earthquake, or human error? Azure SQL Data Warehouse can help solve this problem. Because this cloud technology is so much like APS, your on-premises applications can potentially run in the cloud when they need to, such as when the on-premises application is unavailable.

## Remote monitoring with IoT

Bring the Internet of your things to life. Connect and monitor all your devices, assets, and sensors. Increase visibility into performance, efficiency, and enable innovation and improve business outcomes through previously untapped data.

**Scale to millions of devices for a quick time to market**

Create a broad-scale Internet of Things (IoT) solution by connecting devices, assets, and sensors to the cloud. Scale with ease and confidence – from just a few sensors to millions of simultaneously connected devices – with the reliable, global availability you need to keep a competitive edge.

**Uncover insights with previously untapped data**

With your company’s devices and sensors connected, you can collect previously untapped data, then use built-in capabilities to improve operational efficiencies. Set up real-time analytics in a scalable, high-performance and resilient way, without having to manage complex infrastructure and software.

**Build IoT solutions with confidence**

Enhance the security of your IoT solution across physical devices, connections, and data. Use per-device authentication by setting up individual identities and credentials for each of your connected devices. And retain the confidentiality of both cloud-to-device and device-to-cloud messages.

**Monitor assets anywhere in real time**

Whether assets are down the street or across the globe, automate the tracking of their status and health, more effectively, efficiently, and reliably than your existing systems. No matter how remote your assets are, get live data and insights to make real-time decisions that move your business forwards.

### Technology Snapshot: Azure Stream Analytics

Software that analyzes streaming data, must be able to handle fast-moving information with low latency. It must also help its users work with this data in useful ways, hiding as much complexity as possible. These are the exact goals of Azure Stream Analytics.

One of the most common things we want to do with streaming data is understand what’s happening in that stream within a specific period of time. For example, maybe we want to know how many cars have passed through an automated toll booth within the last three minutes or how much electricity was used by houses in a particular neighborhood within the last hour. Stream Analytics is designed to make this easy to do. A developer can use the Stream Analytics Query Language, a subset of T-SQL, to issues queries on an incoming stream of data. Each query can specify a window of time to which the query applies, returning a result for just the data that arrives within that window. Once it’s started, the query keeps running, sending back results for each window. Rather than querying tables, as in a relational database, Stream Analytics instead allows querying slices of an incoming stream.

### Technology Snapshot: HDInsight Storm

HDInsight is Microsoft’s cloud implementation of Hadoop and several other technologies. Among those technologies are MapReduce, Hive, and Pig (all of which are commonly used for analyzing large amounts of analytical data on disk) and HBase, a store for operational data. HDInsight also provides Storm, a technology for working with streaming data.

Storm is similar in some ways to Stream Analytics. Both run in the cloud and both support applications that process streaming data. In Storm, those applications are created using spouts and bolts. A spout accepts incoming data streams, while a bolt processes streaming data in some way. An application, called a topology, is made up of spouts and bolts. Storm takes quite a general approach to working with data streams making it useful in a broad range of streaming scenarios.

### Technology Snapshot: HDInsight Spark Streaming

Along with traditional Hadoop technologies, HDInsight also provides Spark as a cloud service. Spark is an integrated set of open source technologies that can run on a Hadoop cluster. The Spark family includes options for analyzing large amounts of operational data, machine learning, and more. It also includes Spark Streaming, a technology for working with streaming data.

Spark Streaming is similar to Storm in some ways. Like Storm, it’s a general-purpose technology for processing streaming data. Unlike Storm, Spark Streaming is implemented as an extension to the basic Spark engine—it’s not an add-on technology. This tight connection can make Spark applications faster and easier to create, since there’s less need to move data between components and everything uses the same core Spark technology. Due to this, Spark Streaming (and Spark in general) is getting more popular by the day.

Given the similarities between Spark Streaming, Storm, and Stream Analytics, which one should you choose? The answer depends on your situation - there’s no single right answer. For guidance on making this choice, see Understanding Your Options later in this section.

### Technology Snapshot: Azure IoT Hub

Stream Analytics, HDInsight Storm, and Spark Streaming lets your organization create software that processes streaming data. However, none of them are designed to take in and buffer massive amounts of streaming data, something that’s commonly required in IoT scenarios. Without some kind of buffering placed in front of these stream processing services, valuable data may be lost.

Azure IoT Hub addresses this problem. This cloud service is commonly used in front of Stream Analytics, Storm, or Spark Streaming by providing a place to store incoming data until it’s processed. IoT Hub can handle large amounts of incoming data from many devices and it also provides a way to communicate back to those devices.

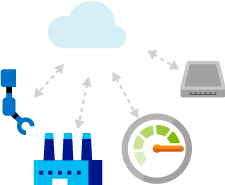
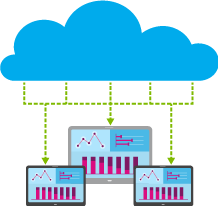
### Understanding Your Options

Since a basic IoT scenario can use Stream Analytics, Storm, or Spark Streaming, which one should you choose? The choice commonly depends on these factors:

* If your application uses time-based queries, Stream Analytics is probably a better choice. This cloud service is designed to answer questions like this and its SQL-based query language will likely be easier for your developers to understand. It’s possible to do time-based queries with Storm and Spark Streaming, but since neither one is specifically designed to make this easy, your development team will probably need to write more code.
* If your application uses event-based queries or other kinds of stream processing that go beyond what Stream Analytics is designed to do, either Storm or Spark Streaming is likely to be the better option. They’re more customizable and they let your developers work in more general programming languages, rather than just Stream Analytics Query Language. Both bring a bit more complexity, but Microsoft and the open source community provide a range of software (such as existing Storm spouts and bolts) to make developers’ lives easier.
* HDInsight is an ecosystem of related technologies and using any of them requires you to create an HDInsight cluster on Azure. If you’re already using an HDInsight cluster for let’s say, data analysis with Hive, using Storm is a natural extension. If you’re using an HDInsight cluster with Spark to do data analysis, using Spark Streaming probably makes sense. If you’re not using HDInsight for anything else, choosing either Storm or Spark Streaming will require you to spin up and pay for an HDInsight cluster. Stream Analytics, by contrast, is a managed service. You don’t need to create your own cluster to use it, which simplifies getting started with streaming applications.

## Predictive maintenance with IoT

Predict equipment failures before they happen and systematically prevent them for millions of machines across the globe. Use streaming data from sensors and devices to recognize warning signs, predict equipment maintenance needs, and pre-emptively repair equipment, saving you time and money.

**Increase asset reliability with rich insights**

Focus on what matters most to your customers: reliability. Improve asset availability by gathering and transforming data from sensors and systems to vastly improve operations, offering predictive (and even pre-emptive) maintenance.

**Predict equipment failures before they happen**

Monitor an asset’s health to prevent potential problems, whilst promoting equipment efficiency. Collect and analyze the untapped data from your connected assets to proactively plan maintenance, decrease downtime, and improve retention of the asset’s value.

**Monitor millions of devices globally**

Use a cloud solution to connect devices and assets, then collect untapped data and create predictive models in any location worldwide. Improve access to production and supply chain data worldwide, reducing costly downtime and maintenance, while increasing productivity.

### Technology Snapshot: Azure Blobs

The term “blob” is an acronym for Binary Large Object, and that’s exactly what data Azure Blobs store. Raw binary Blob data storage is quite scalable—a single blob can hold hundreds of gigabytes of data and relatively inexpensive at just a few cents per gigabyte per month. If you need to store large amounts of unstructured data as cheaply as possible, Azure Blobs are hard to beat.

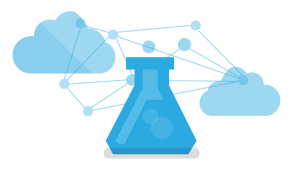
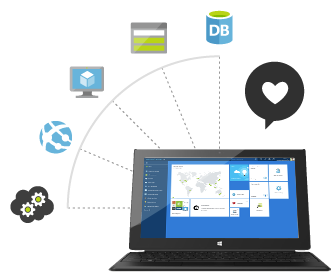
### Understanding Your Options

In fact, here are some things to think about:

* Historical data stored in Blobs might be read by Azure ML, as just described. It might also be examined by other analytical technologies in the Microsoft data platform, including Hive, Spark, and a Microsoft-created approach called Azure Data Lake Analytics. All of these are capable of analyzing large amounts of unstructured data in parallel.
* Along with Azure Blobs, the Microsoft data platform includes other options that might be used to store streaming data. For example, Azure Data Lake Store is a cloud offering that implements the Hadoop Distributed File System (HDFS) as a service. Especially for data that will be used for later analysis, Azure Data Lake Store can offer more scale and better performance than Blobs.
* As the figure shows, a streaming technology can send streaming data to several outputs simultaneously. It can even send that data to other Azure services, letting the same stream of data be processed in multiple ways.

## Cognitive Services

Enable natural and contextual interaction with tools that augment users' experiences using the power of machine-based intelligence. Tap into an ever-growing collection of powerful artificial intelligence algorithms for vision, speech, language, and knowledge.

**Language**

Allow your apps to process natural language, evaluate sentiment and topics, and learn how to recognize what users want.

**Speech**

Processing spoken language in your applications

**Vision**

State-of-the-art image processing algorithms to build more personalized apps by returning smart insights such as faces, images, and emotion recognition.

**Knowledge**

Map complex information and data in order to solve tasks such as intelligent recommendations and semantic search.

# Questions

Intelligent Apps & Analytics represents a large, complex undertaking with many interdependent parts. The first step of Microsoft Intelligent Apps & Analytics is determining where to begin, what to produce, and how to produce it. Complexity of the assessment is compounded by partial artifacts of previous projects, missing history, and multiple agendas. As with any complex undertaking, assessment is most successful when the large, complex problem is divided into smaller, more manageable pieces.

## Business Needs

Business Needs Assessment includes an analysis of the underlying business drivers and objectives and overall context of business needs that has been established for Intelligent Apps & Analytics. In an assessment the objective is not to perform the analysis. It is to determine the degree of analysis that has been done, and to identify any business analysis gaps and their impacts.

The following key questions are among those that a business needs assessment may address:

|  |  |
| --- | --- |
| Question | Response |
| Have business drivers and objectives been identified? |  |
| Have business requirements been documented? |  |
| Please describe your mains goals and objectives for this engagement. Do the requirements align with business drivers and objectives? Do they focus on strategies that respond to the drivers? |  |
| Do the information needs identify and target the enabling of specific business processes and tactics? |  |
| Do the information needs identify the key performance indicators (business metrics) and business perspectives (dimensions, descriptive attributes) needed to measure, analyze, and optimize the targeted business processes? |  |
| Do they identify the roles to be supported, their number and distribution? |  |
| Do they identify the frequency and volume of reporting and analytical needs? |  |

## Information Architecture

Information Architecture Assessment includes an analysis of logical data structures, their feasibility, completeness, documentation, and fit to business requirements. Information architecture assessment also includes analysis of data sourcing and transformation, the methods and assumptions applied, and validation of mappings to business requirements. Metadata, as part of the information architecture, is examined with respect completeness of metadata being tracked, user metadata requirements, and approaches to management of the metadata.

Key questions addressed by information architecture assessment include:

|  |  |
| --- | --- |
| Question | Response |
| Are the information requirements modeled? |  |
| Do they map to and support the identified business requirements? |  |
| Does data access and analysis clearly add business value and meet business needs? |  |
| Is the underlying data architecture flexible and extensible? Can it support multiple analytical needs? Does it readily allow for integration of new data? |  |
| Have data quality issues have been identified and addressed? Is the data being delivered is valid? |  |
| Can users get to, manipulate, and analyze the data when and in form needed? |  |
| Does data access and analysis clearly add business value and meet business needs |  |

## Technical Architecture

Technical Architecture Assessment looks at current Architecture designs and services adopted. Technical architecture assessment seeks to identify any technical risks or constraints with regard to performance, maintenance, scalability, data distribution, disaster recovery, and sizing. This assessment also seeks to identify opportunity to leverage the value of existing technical resources. Effective use of tools, and their overall fit to the business.

Some of the key questions of technical architecture assessment include:

|  |  |
| --- | --- |
| Question | Response |
| Does the technical architecture provide for the efficient delivery result? |  |
| Do they map to and support the identified business requirements? |  |
| Does data access and analysis clearly add business value and meet business needs? |  |
| Is the underlying data architecture flexible and extensible? Can it support multiple analytical needs? |  |
| Have data quality issues have been identified and addressed? Is the data being delivered is valid? |  |
| What kind of BI strategic platform is needed? Big Data platform, Analytics tools, or Visualization tools? |  |
| Is current access and performance adequate? What about performance of data acquisition and refresh processes? |  |

## Organization

Organizational Assessment includes an examination of the existing organizational structure and identification of the roles and responsibilities of both IT and the business community that need to be addressed.

Among the key questions that organizational assessment addresses are:

|  |  |
| --- | --- |
| Question | Response |
| Have roles and responsibilities for each service been identified and documented? |  |
| Has ownership of service used been addressed from the perspective of strategic business objectives and direction setting? From the perspective of tactical enhancement and ongoing business needs? From perspectives of information management and technical support and responsibilities? |  |
| Has the key issue of business/IT collaboration been directly articulated and addressed? |  |
| Has the need for ongoing operational support and tuning, in parallel with continuing development, been considered? |  |
| Has any structure been put in place for ongoing monitoring of the whole system, and for periodic assessments, as needed? |  |
| Are deliverables clearly identified? |  |

## Big Data Maturity Level

Big data is a journey. It involves building an ecosystem that includes technologies, data management, analytics, governance, and organizational components. We should think about the Big Data Maturity Model to describe the stages that most organizations follow when they embark on big data initiatives. The model provides the big picture of a big data program, where it needs to go, and how to get there. As organizations move through these stages, they gain more and more value from their investments.

|  |  |
| --- | --- |
| Question | Possible Answers (Choose One) |
| To what degree are your advanced analytics (natural language processing, sentiment analysis, predictive analytics and classification) paired with your Big Data efforts? | 0) Not applicable 1) They are completely separate 2) We are in the process of combining the two 3) They are somewhat integrated 4) They are fully integrated |
| To what degree do your advanced analytics projects (natural language processing, sentiment analysis, predictive analytics and classification) create consistent and repeatable response to patterns of change? | 0) Not applicable 1) The results are not consistent 2) The results are consistent about half the time 3) The results are consistent most of the time 4) The results are always consistent |
| Do you have dedicated resources exploring the possibilities of advanced analytics in Big Data for your business line? | 0) Not applicable 1) No 2) Planned 3) Implementing 4) Already implemented |
| Do you plan to employ machine learning technology while doing Advanced Analytics? | 0) Not applicable 1) No 2) Maybe 3) Yes  4) Already implemented |
| How is Social Media being monitored in your organization? | 0) Not applicable 1) It is not currently being monitored 2) An employee in the PR/Marketing department monitors social media manually 3) Outsourced to a Social Media Monitor (SMM) with quarterly reports on findings 4) Outsourced to SMM with internal access to monitor daily/weekly activity |
| What is your ability to scale in terms of storage and processing power? | 0) Not applicable 1) Currently unable to scale to meet demand 2) Plans are in place to scale 3) Currently able to scale using cloud-based infrastructure |
| Do you engage in parallel computing by dividing large problems into smaller ones that can be computed simultaneously? | 0) Not applicable 1) No 2) Currently planned/under development 3) Yes |
| Do you have a system in place to sort incoming data in near real time by potential value, data quality, and use frequency? | 0) Not applicable 1) No 2) We sort by one of these classifications 3) We sort by two of these classifications 4) Yes, we sort by all three |
| Do you use event-driven architecture to manage incoming data? | 0) Not applicable 1) No 2) Currently planned/under development 3) Yes |
| Do you have specialized data services that can accommodate different formats, security, and the management requirements of multiple data sources? | 0) Not applicable 1) No 2) Currently planned/under development 3) Yes |
| Is your organization ready to respond, in real time, to complex events that may affect your business? | 0) Not applicable  1) Not ready  2) Actively planning method for responding to events  3) Able to respond to internal events  4) Able to respond to internal/external events |
| Can you perform analysis on data as it is loaded at high velocity (1 gigabyte [GB] + per second)? | 0) Not applicable 1) No 2) Currently planned/under development 3) Yes |
| Do you use data models that support data-intensive distributed applications? | 0) Not applicable 1) No 2) Currently planned/under development 3) Yes |
| Is your organization currently using or considering in-memory analytics? | 0) Not applicable 1) Not currently using in-memory analytics 2) Considering in-memory analytics  3) Currently using in-memory analytics, but no formal process in place to manage consistency and updates 4) Currently using in-memory analytics and have a process in place for performance tuning |
| To what degree are you able to correlate data from your Big Data infrastructure with that from your enterprise data warehouse? | 0) Not applicable 1) Low - Big Data results are manually loaded into a database to be analyzed 2) Medium - Big Data results feed into a data warehouse, but are not integrated with structured data. 3) Medium-High - Some Big Data results feed into the EDW and can be selected alongside transactional data 4) High - All Big Data results feed into the EDW and can be selected alongside transactional data. |
| Have you extended the role of Data Stewards to include ownership of big data components? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Are your Big Data stewardship policies consistent with existing data policies, or do some of the policies require additional consideration? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Do the functional areas understand the data quality risks inherent in their respective data sources? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Do you prioritize data quality based on the source system (that is Facebook/Twitter data has lower quality thresholds than radio frequency identification (RFID) for a tracking system)? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Do your retention policies consider the different legal responsibilities for storing Big Data for a specific amount of time? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Do Data Scientists work in close collaboration with Data Stewards to ensure data quality? | 0) No  1) <50% of the time 2) <75% of the time 3) Always |
| How is access to attributes of Big Data being given out in the organization? | 0) Not applicable 1) Nothing is currently being done. Data is being collected and all levels have access to it 2) Filtered access to relevant stakeholders only 3) Multitier structure on what attributes will be shared and to whom 4) An encrypted database with personally identifiable information filtered out |
| Is Big Data included in the descriptions of key roles, such as Chief Data Officer and Information Governance officer? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Are roles related to Big Data (Advanced Analyst, Data Scientist) clearly defined? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| How involved is risk management in the Big Data governance process? Is it a key stakeholder? | 0) Not applicable 1) Not currently involved 2) Informed of plans after they are made 3) Actively involved in the planning |
| Is there a set of documented policies regarding Big Data governance? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Is there an enforcement mechanism or approach to ensure that policies are followed? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Is a GRC framework in place to assess the efficacy by which the company adheres to its data governance policies? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Who is the key sponsor for your Big Data governance program? (CIO is best) | 0) Not applicable 1) IT Manager 2) IT VP 3) CIO |
| Do you understand the terms and conditions of data procured from social media sites? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| Do you have defined policies surrounding the use of social media data for potential employees and customers, as well as the use of customer geo-location data? | 0) Not applicable 1) No 2) No, but plans are in place  3) Yes |
| How accessible are complex analytic routines to your user base? | 0) Not Applicable 1) Routines are manually coded by IT for the business 2) Routines are stored by IT for reuse and accessible by some users 3) Routines, custom and open source, are available to all users 4) Routines, custom and open source, are available through a self-service graphical UI |
| What is the level of involvement with outside vendors and third parties in regard to the planning and execution of Big Data projects? | 0) Not applicable 1) Exclusively in-house 2) In-house but considering hiring external consultants 3) Mainly in-house with some external |
| To what extent does your organization use sandboxes for experimentation in data analysis? | 0) No sandboxes are used  1) A sandbox is sometimes available  2) A sandbox is always available 3) A sandbox is always available and non-standard tools are available |
| What programming technologies are utilized by your data warehouse/BI staff when working with Big Data? | 0) Same technologies as with the Enterprise DW 1) Low-level programming languages (Ruby, Python, Java) 2) MapReduce-specific programming languages (Hive, Pig) 3) Combination of both high and low, depending on the task |
| What is the level of business/industry domain knowledge on the part of IT in the planning of Big Data projects and operations? | 0) Not applicable 1) Low - IT has little/ no domain knowledge.  2) Average - IT has average domain knowledge but lacks proactive abilities to meet the needs of the business 3) High - IT has deep domain knowledge and can easily integrate with business representatives |
| How many of the following skill sets do your BI staff possess?  1) Quantitative R&D: Creation of theory and development of algorithms for all forms of quantitative analysis. 2) Data Scientist or Quantitative Analyst: Incorporation of advanced analytical approaches derived from Quantitative R&D. 3) Operational Analytics: Application of the above applications into real work. 4) Business Intelligence (BI) and Discovery: Reporting, dashboard, online analytical processing (OLAP) and visualization use. Performing posterior analysis of results driven by quantitative methods. | 0) Not applicable 1) None 2) 1 Group 3) 2 Groups 4) 3 Groups 5) All groups |
| How many Data Scientists (individuals with a deep understanding of the business model and industry who are able to extrapolate underlying knowledge from data using mathematics, statistics and computer science) do you have on your BI staff? | 0) Not applicable 1) None 2) Planning to hire one or more 3) 1 4) 2 or more |

## IoT Service

The following question list can help gain more insights into the environment and IoT needs.

|  |  |
| --- | --- |
| Question | Response |
| Devices (General) | |
| Are the devices directly IP addressable? Or, is communication managed (or to be managed) via field gateway? |  |
| What are the physical characteristics of the device(s) and (if used) field gateways? |  |
| Well known or custom device? |  |
| Which Microcontroller / Microprocessor? Speed? |  |
| What OS (if any) do the devices utilize. |  |
| Memory capacity (RAM, ROM, Flash, durable storage (HD, SD, etc.) |  |
| Is power consumption an issue (e.g., battery powered, ultra-low powered, BLE Beacons)? If so, what are the specs - power consumption, battery size, expected/required battery duration? |  |
| Any other device-specific characteristics that are pertinent to the project? |  |
| Is real-time functionality within the device required (e.g., control operations/actuators that are critically sensitive to timing) |  |
| How many devices exist today? |  |
| Is there a plan or need to design/build additional devices? |  |
| Are the devices used for telemetry (send data such as sensor measurements only), command & control (receive data only to manipulate actuators), or both? |  |
| What is the projected plan for adding devices in future? How quickly with those devices be rolled out in the future? |  |
| Is managed deployment of firmware/software required? |  |
| Are there any restrictions to building/deploying custom code to the devices or the gateways? |  |
| Connectivity & Communication | |
| If devices AND field gateways will be used: how do the devices communicate to the field gateways? (e.g., what is the message structure, size, message protocol and/or the transport protocol?) |  |
| What is the protocol (HTTPS/MQTT/AMQP/CoAP/TCP/UDP etc.) required for messaging from the device/field gateway to the cloud gateway? |  |
| What are the protocols (HTTPS/MQTT/AMQP/CoAP/TCP/UDP/SMS etc.) required for messaging from the service platform to the device/field gateway? |  |
| Are wireless communications being used (Cellular, WiFi, Zigbee, Z-Wave, etc.)? |  |
| Is the communication one-way from device (telemetry data being sent out) or two-way (where device receives commands and/or receipt notifications) |  |
| Is there schema/data-model associated with what data will be sent & how it will be structured? Is there preferred serialization & packaging format (say JSON/BSON, GZIP etc.) |  |
| Will connections between the cloud and the field device/gateway be long lived (durable) or short (connect, communicate message(s), disconnect) - possibly mobile? |  |
| Is it likely that the device/gateway will have significant periods of time where cloud connections may not possible? (away from Wi-Fi, cellular connections) |  |
| If there is a gateway to be used (or in place) and command & control will be utilized, will C&C messages be addressed to individual devices or will it be up to the gateway to determine which device will receive a message? |  |
| Field Gateway | |
| How does the field gateway communicate with the end devices? (hard wired, wireless, CAN bus, serial, etc.? |  |
| If command and control is to be used, how are messages to be directed from the field gateway to the individual device? (e.g., field gateway acts as proxy, field gateway will have intelligence to determine where commands are to be directed) |  |
| Will the field gateway be used as a end device as well? Does the field gateway have directly connected sensors or actuators? |  |
| Is there a need for local intelligence within the gateway to manage or handle events before they are sent to the cloud intelligence? (e.g., closed-loop processing for safety/speed when the latency of a call to the cloud is unacceptable) |  |
| Messages | |
| What message structure(s) are used between devices and field gateways and/or cloud gateways? |  |
| What message encoding will be used? (JSON, CSV, etc.) |  |
| What are the sizes of the messages? Any messages over 64K? Over 256K? |  |
| How often will the messages be sent? (# of messages per timeframe) |  |
| Do messages (data) need to be compressed before sending/receiving? |  |
| Device Provisioning and Registration | |
| Is there existing (internal and/or external) device management, message management/brokering software or services currently in place? |  |
| How are device provisioned from the factory? How are they registered/linked to user (vehicle owner or fleet administrator etc.)? |  |
| How is a device authenticated with the platform? Are there some pre-shared access keys that are burnt into device (as part of provisioning)? |  |
| How is a device authorized to send and/or receive messages? |  |
| Will devices be registered/provisioned on-the-fly or will credentials and endpoints need to be pre-provisioned? |  |
| Is there common configuration or state data required across devices? |  |
| Geo Requirements | |
| How geographically dispersed are the devices? |  |
| Do the devices need to communicate with a "local" ingestion endpoint to reduce latency? |  |
| Firmware Updates | |
| How are firmware updates send to the device? |  |
| Telemetry | |
| what type of information are you planning to capture from devices? |  |
| how many messages will be ingested (provide expected rates / second)? |  |
| is there a message schema that is defined for the messages? (Provide example) |  |
| do all devices follow a similar message schema? (if different provide examples) |  |
| what is the max size of a message sent from device to backend? |  |
| what is the average size of a message sent from device to backend? |  |
| do you want to support sending large message over the wire (> 192 KB)? |  |
| what is the message payload format (binary, string, JSON etc.?) |  |
| do you also transmit metadata for the message? (these are generally in the form of custom configuration properties which are not part of the message payload etc.) |  |
| do you apply any transformation to the message payload before transmitting the message? |  |
| do you need to perform any validations to the message at the backed? |  |
| do you capture telemetry data in batches or on a per event basis? |  |
| does the message contains attributes to uniquely identify a device? (example Device Id etc.) |  |
| Inquiry | |
| does the device require to ask for information from the backend systems (example: weather updates etc.) |  |
| what types of inquiry can the device perform (please provide examples) |  |
| does the inquiry require need to result in an immediate response or the response can be delivered in near real time / batch? |  |
| is the device capable of sending inquiries or a device gateway send these on their behalf? |  |
| do you have access restrictions of the kind of inquiries a device can make? |  |
| how does the device behave if a inquiry request is denied / fails? |  |
| Command | |
| what type of actions can be sent to a device (provide examples)? |  |
| do you support sending multiple actions as a batch request to the device? |  |
| what is the message payload format (binary, string, JSON etc.?) |  |
| is the message received by a device gateway or by the device directly? |  |
| what is the max size of a message sent from backend to device? |  |
| what is the average size of a message sent from backend to device? |  |
| do you want to support sending large message over the wire (> 192 KB)? |  |
| Notifications | |
| what type of notifications can be sent to a device (provide examples)? |  |
| do you support sending multiple actions as a batch request to the device? |  |
| what is the message payload format (binary, string, JSON etc.?) |  |
| is the message received by a device gateway or by the device directly? |  |
| what is the max size of a message sent from backend to device? |  |
| what is the average size of a message sent from backend to device? |  |
| do you want to support sending large message over the wire (> 192 KB)? |  |
| Sizing/Configuration | |
| Will you need multiple instances of the IoT Hub for different environments (e.g, development, test, etc.)? |  |
| Will you need multiple instances of the IoT Hub for business or policy reasons (e.g., different instances for different business units)? |  |
| In what geographies are the connected devices? |  |
| How much data is flowing from devices over time (e.g., per day)? |  |
| How distributed is the data flow from devices over time (e.g., is the same amount of data flowing per hour or are there peak periods)? |  |
| In what geographies are the connected devices? |  |
| Reliability | |
| What are the availability requirements for the processing pipeline (IoT Hub)? |  |
| If IoT Hub is down is failover necessary? |  |
| General | |
| Provide metrics around response time, availability etc. |  |
| Data Storage, Complex Event Processing, Reporting and Machine Learning | |
| what will the data be primarily used for (analytics, auditing etc.)? |  |
| will the data be captured for each region separately or it needs to be shared across regions? |  |
| does the data need to be encrypted? |  |
| What are the message sizes and formats? |  |
| Does the data need to be cached or preprocessed in the device? |  |
| Does the data transformation need to happen in the device? |  |
| Which formats does the data need to be published? |  |
| Data Archival requirements |  |
| Is there any PII data that is being stored |  |
| Does the on-premise data need to be imported to the cloud to combine with the telemetry data collected using the Iot solution |  |
| do we use any tools for data analytics, data warehousing, data mining etc.? |  |
| are there any reporting requirements (Please provide example of reports) |  |
| do you require real/near-realtime time data analytics / processing? |  |
| Do you require micro-batch processing and if so, are you using Apache Spark or similar? |  |
| do you require traditional batch data analytics / processing? |  |
| what is the average size of data that you want to capture? |  |
| do you already have a data collection and analytics strategy? (if yes please provide details) |  |
| is the data archived, if so what are the retention policies |  |
| is the data replicated in other repositories or a single data warehouse is used (please attach data architecture if available) |  |
| Does all the data need to be stored in cold storage or only the aggregated data? How long |  |
| How much historical data & responses are available for any form of supervised machine learning? |  |
| How much of the source data is needed for insights? Tradeoffs associated with sending too much data over wire (costs) versus data needed for insights & machine learning |  |
| Security, Access and Compliance | |
| do you have confidential data being transmitted to / from devices? |  |
| do you use pre-shared keys during device manufacturing? |  |
| which protocol are used for transport layer security (TLS -SSL, PSK-TLS etc., 6LoWPAN ) |  |
| do you encrypt message before transmitting? (If yes please specify encryption algorithms used) |  |
| what are the authorization policies for devices to request information (if any)? |  |
| do you leverage an user directory for authentication / authorization (e.g. Azure AD) |  |
| do you have any compliance requirements? |  |
| how do you protect key material on the device? |  |
| How do you provision the device? |  |
| How do you know you are talking to your device? Tamperproof mechanisms |  |
| How do you colect/maintain security events from device? |  |
| Do you deploy certificates to the device? What kind and what is the purpose |  |
| Are custom security tokens or certificates required? |  |
| Is data moving to and from the devices encrypted? What encryption schemes are in place? What will need to be in place? |  |
| What process (with respect to the devices/field gateways) would need to be in place if a device or field gateway was compromised? (ability to deactivate device, ignore or reroute data/commands) |  |
| Do the devices have an embedded identity? (e.g., EIN, Serial Number, or use a SIM card) |  |
| Application Integration | |
| do you have any consumer applications (Web, Mobile) that will interact with devices (example: changing temperature on thermostat from a mobile device) Provide details on how commands / actions are being performed by these) |  |
| is the consumer application available globally or regionally (provide scale and no of user interacting) |  |
| how many users can be connected to these consumer applications at a time? (provide average concurrency rates) |  |
| what is the acceptable latency for device updates for actions users perform in the applications? (example: changing temperature on a thermostat) |  |
| In case of Mobile – provide more details on the application:  Platforms and languages for the apps  Connectivity  Any expected lifecycle of usage? |  |
| what scenarios do you want to enable through the mobile / web application? (today / future) |  |
| do you have any internal / partner applications that leverage telemetry data to provide analytics and reporting (provide details on how data is being used by these applications) |  |
| what scenarios do you want to enable through the internal / partner application? (today / future) |  |

# Resources and Tools

## Big data and analytics

<https://azure.microsoft.com/en-gb/solutions/big-data/>

## Business intelligence

<https://azure.microsoft.com/en-us/solutions/business-intelligence/>

## Data warehouse

<https://azure.microsoft.com/en-gb/solutions/data-warehouse/>

## Remote monitoring with IoT

<https://azure.microsoft.com/en-gb/solutions/remote-monitoring/>

## Predictive maintenance with IoT

<https://azure.microsoft.com/en-gb/solutions/predictive-maintenance/>

## Cognitive Services

<https://azure.microsoft.com/en-us/services/cognitive-services/>

## Microsoft Azure IoT Starter Kits

Start innovating today with kits that include development boards that are ‘Azure Certified for IoT’, sensors and actuators. Simple user-friendly tutorials help you to seamlessly connect your devices to the cloud with Microsoft Azure IoT.

<https://azure.microsoft.com/en-gb/develop/iot/starter-kits/>

Related Azure services and Microsoft products (HDInsight, SQL Data Warehouse, Machine Learning)

<https://azure.microsoft.com/en-gb/services/hdinsight/>

<https://azure.microsoft.com/en-gb/services/sql-data-warehouse/>

<https://azure.microsoft.com/en-gb/services/machine-learning/>