Business in Real-Time Using Azure IoT and Cortana Intelligence Suite

Driving Your Digital Transformation

Bob Familiar
Jeff Barnes

Apress®

Business in Real-Time Using Azure IoT and Cortana Intelligence Suite

Driving Your Digital Transformation

Bob Familiar

Jeff Barnes

Apress®

Business in Real-Time Using Azure IoT and Cortana Intelligence Suite: Driving Your Digital Transformation

Bob Familiar
Sudbury, Massachusetts, USA

Jeff Barnes
Miami, Florida, USA

ISBN-13 (pbk): 978-1-4842-2649-0
ISBN-13 (electronic): 978-1-4842-2650-6
DOI 10.1007/978-1-4842-2650-6

Library of Congress Control Number: 2017943349

Copyright © 2017 by Bob Familiar and Jeff Barnes

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

Trademarked names, logos, and images may appear in this book. Rather than use a trademark symbol with every occurrence of a trademarked name, logo, or image we use the names, logos, and images only in an editorial fashion and to the benefit of the trademark owner, with no intention of infringement of the trademark.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Cover image designed by Freepik

Managing Director: Welmoed Spahr
Editorial Director: Todd Green
Acquisitions Editor: Natalie Pao
Development Editor: James Markham
Technical Reviewers: Alina Stanciu and Jim O'Neil
Copy Editor: Kezia Endsley
Coordinating Editor: Jessica Vakili

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail rights@apress.com, or visit http://www.apress.com/rights-permissions.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/978-1-4842-2649-0. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

This book is dedicated to all those who seek to "live life in the fast lane" by exploiting technology to help drive digital disruption and positively transform their businesses so they may truly operate at Internet speed.

To my incredible wife, Mandy, who is a continuous stream of inspiration, and to my children, Ariana and Bobby, who never cease to amaze me with their talent, insight, and intelligence.

—Bob Familiar

This book is the culmination of many long, sacrificed nights and weekends. To that end, I would like to thank my wife Susan, and my children, Ryan, Brooke, and Nicholas, for their constant love, support, and encouragement.

—Jeff Barnes

Contents at a Glance

About the Authors.....xv

About the Technical Reviewers xvii

Acknowledgments.....xix

Introduction.....xxi

Chapter Overview xxiii

■ Chapter 1: Business in Real-Time 1

■ Chapter 2: DevOps Using PowerShell, ARM, and VSTS 21

■ Chapter 3: Device Management Using IoT Hub..... 95

■ Chapter 4: Sensors, Devices, and Gateways..... 127

■ Chapter 5: Real-Time Processing Using Azure Stream Analytics..... 169

■ Chapter 6: Batch Processing with Data Factory and Data Lake Store..... 227

■ Chapter 7: Advanced Analytics with Azure Data Lake Analytics..... 291

■ Chapter 8: Advanced Analytics Using Machine Learning and R..... 351

■ Chapter 9: Data Visualizations, Alerts, and Notifications with Power BI..... 397

■ Chapter 10: Security and Identity..... 475

■ Chapter 11: Epilogue 517

Index..... 519

Contents

About the Authors xv

About the Technical Reviewers xvii

Acknowledgments xix

Introduction xxi

Chapter Overview xxiii

Chapter 1: Business in Real-Time 1

A Platform Approach 1

Real-Time Business Platform 2

Internet of Things (IoT) and Big Data 5

Real-Time Business Reference Architecture 8

Devices 9

Device Hub 9

Stream Processing 10

Data Management, Storage, and Messaging 10

Advanced Analytics 10

Microservices and API Gateways 10

Visualization, Alerts, and Notifications 11

Identity 11

Automation 11

Security 12

Microsoft Azure IoT and Cortana Intelligence Suite 12
Automation 13
Security and Identity 13
Devices 13
Device Hub 14
Stream Processing 14
Data Management, Storage, and Messaging 14
Advanced Analytics 14
Microservices and API Gateway 14
Visualization, Alerts, and Notifications 15
Worker Health and Safety: A Reference Implementation 15
Backstory 15
Solution Architecture 15
Downloading the Repository 18
Summary 19
Chapter 2: DevOps Using PowerShell, ARM, and VSTS 21
People 22
Process 22
Infrastructure as Code 23
Continuous Integration 23
Continuous Delivery 24
Tools 24
DevOps and Azure 24
PowerShell 25
Azure PowerShell 28
Azure Resource Manager 29
ARM Templates 30
Visual Studio Team Services 34
Summary 93

Chapter 3: Device Management Using IoT Hub 95
The Device Management Lifecycle 95
Planning 97
Provisioning 98
Configuring and Monitoring 104
Retiring 112
Summary 126
Chapter 4: Sensors, Devices, and Gateways 127
Sensors 127
Programmable Logic Controllers 128
Devices 130
GSM Modems 130
RFID 133
Bluetooth Beacons 135
Get Smart 136
Microcontroller Software 138
Edge Gateways 140
Summary 168
Chapter 5: Real-Time Processing Using Azure Stream Analytics 169
The Lambda Architecture 169
What Is Streaming Analytics? 171
Real-Time Analytics 171
Streaming Implementations and Time-Series Analysis 172
Predicting Outcomes for Competitive Advantage 174
Stream Processing: Implementation Options in Azure 174
Choosing a Managed Streaming Analytics Engine in Azure 176
Streaming Technology Choice: Decision Considerations 178
Pain Points with Other Streaming Solutions 178

Reference Implementation Choice: Azure Streaming Analytics 178
Advantages of Azure Streaming Analytics 178
Development and Debugging Experience Through Azure Portal 180
Scheduling and Monitoring Built-In 180
Why Are Customers Using Azure Stream Analytics? 181
Key Vertical Scenarios to Explore for Azure Stream Analytics 181
Our Solution: Leveraging Azure Streaming Analytics 182
Streaming Analytics Jobs: INPUT definitions 182
Streaming Analytics Jobs: OUTPUT Definitions 184
Planning Streaming Analytics Outputs 185
Streaming Analytics Jobs: Data Transformations via SQL Queries 186
Azure Streaming Analytics SQL: Developer Friendly 187
Azure Streaming Analytics (ASA): SQL Query Dialect Features 187
SQL Query Language 187
Supported Data Types 187
Data Type Conversions 188
Temporal Semantic Functionality 188
Built-In Operators and Functions 188
User Defined Functions: Azure Machine Learning Integration 189
Event Delivery Guarantees Provided in Azure Stream Analytics 189
Time Management Functions 190
Azure Stream Analytics: Unified Programming Model 191
Azure Stream Analytics: Examples of the SQL Programming Model 191
The Simplest Example 191
Tumbling Windows: A 10-Second Tumbling Window 191
Hopping Windows: A 10-Second Hopping Window with a 5-Second “Hop” 191
Sliding Windows: A 10-Second Sliding Window 192
Joining Multiple Streams 192
Detecting the Absence of Events 192

The Reference Implementation 193
Business Use Case Scenario 193
Summary 226
Chapter 6: Batch Processing with Data Factory and Data Lake Store 227
Azure Data Factory Overview 227
Pipelines and Activities 228
Linked Services 230
Datasets 230
Pipelines 231
Scheduling and Execution 231
Pipeline Copy Activity End-to-End Scenario 231
Monitoring and Managing Data Factory Pipelines 236
Data Factory Activity and Performance Tuning 239
Azure Data Lake Store 241
Hadoop Access 241
Security Layers 241
Implementing Data Factory and Data Lake Store in the Reference Implementation 243
Summary 290
Chapter 7: Advanced Analytics with Azure Data Lake Analytics 291
Azure Data Lake Analytics 292
Getting Started with Azure Data Lake Analytics 294
Implementing ADLA in the Reference Implementation 319
Reference Implementation Summary 348
Summary 348
Handles Virtually All Types of Data 348
Productive from Day One 349
No Limits to Scale 349
Enterprise Grade 349
Reference Implementation 349
Just Scratching the Surface 350

Chapter 8: Advanced Analytics Using Machine Learning and R 351

What Is Machine Learning? 351

Understanding Machine Learning 351

Brief History of Machine Learning 355

Industry Applications of Machine Learning 355

Overview of Azure Machine Learning 357

The Traditional Data Science Landscape 358

Democratizing Machine Learning 358

Azure Machine Learning Studio 359

Microsoft R Server Overview 364

Processing limitations of Open Source R 364

Enter Microsoft R Server 365

Extend Machine Learning Experiments with the R Language Module 366

R Tools for Visual Studio 367

Implementing Azure Machine Learning and R in the Reference Implementation 367

Business Case for Machine Learning 367

Reference Implementation: Assumptions 368

Choosing a Machine Learning Algorithm 368

Summary 396

Chapter 9: Data Visualizations, Alerts, and Notifications with Power BI 397

The Modern Reporting Landscape 397

Overview of Power BI 399

The Power BI Service 399

Power BI Desktop 399

Power BI Mobile 404

Power BI Embedded 405

Power BI ReST APIs 406

Power BI Custom Visuals 406

Power BI Natural Language Query 407

Power BI Cortana Integration 408

Cloud Reporting Cost Architectures 409

Alerts and Notifications 410
Azure Event Hub 411
Streaming Analytics: Output to Event Hub 413
Azure Functions 414
Reference Implementation 417
Implementation Overview 417
Summary 472
Chapter 10: Security and Identity 475
Threat Modeling 475
Threat Modeling Zones and IoT 476
Local Zone 476
Device Zone 477
Cloud Gateway Zone 478
Cloud Services Zone 479
Security Protocols 481
Azure Security Center 481
Data Encryption 482
Key Management 482
Identity 483
Authentication and Authorization 483
Multi-Tenancy 483
Summary 516
Chapter 11: Epilogue 517
Index 519

About the Authors

Bob Familiar is the National Practice Director for Application Services at BlueMetal, an Insight company. As National Director, Bob leads a team of seasoned principal architects who are responsible for industry research and development, rapid prototyping and outreach to the technology community, and providing strategy and guidance to BlueMetal clients on their most important and challenging projects. Bob is an accomplished software professional, evangelist, and author. Bob has been working in the software industry for over 30 years and holds a patent in object relational architecture and technology and is the author of *Microservices, IoT, and Azure,* available from Apress.

Jeff Barnes is a Cloud Solution Architect on the Microsoft Partner Enterprise Architect Team (PEAT), where he helps global Microsoft partners with pre-sales Azure architecture technical support. Jeff has been in the IT industry for over 30 years and has been with Microsoft for the last (18+) years. During that time, he has held several key software architect roles at Microsoft, including consulting, evangelism, and partner divisions. Prior to Microsoft, Jeff worked for several Fortune 500 financial, manufacturing, and retail companies. Jeff holds a Bachelor's degree in Management Information Systems from Florida State University and is the author of *Microsoft Azure Essentials: Azure Machine Learning by Microsoft Press*.

About the Technical Reviewers

Alina Stanciu is a senior program manager of the Microsoft Azure IoT team, where she drives deep engagements with Microsoft partners worldwide to accelerate their migration path to Azure IoT services. Alina works directly with customers and top-tier partners to get product feedback for continuous innovation of the Azure IoT platform. Her support of the Azure IoT community ranges from training, to workshops and hackathons, to deep architectural reviews. She not only leads training sessions for internal events, such as Microsoft MVP Summit, Microsoft TechReady, and Microsoft C+E University, but she also participates in external events, such as the Microsoft Worldwide Partner Conference or IoT Solutions World Congress. Before joining the Azure IoT team, Alina worked for the Microsoft Cloud Infrastructure Operations, where she delivered hyper-scale operations management platform for Microsoft Data Centers.

Jim O'Neil has over 25 years of experience in the software industry, spanning the disciplines of software prototyping, military and commercial software development, technical support, and developer evangelism. Jim's current passion is the intersection of devices (mobile and IoT) with cloud infrastructure, focusing on agility, scalability, and reliability. Jim joined BlueMetal, an Insight Company, in January 2014 after six years in the Microsoft Developer Evangelism organization where he hosted workshops, hackathons, and conferences to bring new technologies to a wide developer audience. In his spare time, Jim is the cofounder of New England GiveCamp, an annual hackathon supporting non-profits in the New England region; a perennial facilitator at National Junior Classical League Conventions; and the de facto family genealogist.

Acknowledgments

We would like to thank Kevin Miller, Alina Stanciu, and Nayana Singh, for their guidance and support and the Azure IoT engineering team, for their in-depth technical reviews and highly constructive feedback.

A warm and heartfelt thank you to Jim O'Neil for his thorough review covering both the technical content as well as educating us on the proper use of the Oxford comma.

The following people, all members of the BlueMetal team, have been instrumental in making this book possible through their support and technical expertise—Michael Griffin, Raheel Retiwalla, Ron Bokleman, Mike Shir, Scott Jamison, Matt Jackson, Priya Gore, and Rich Woodbury.

A big thank you to the team at Apress for their awesome support and guidance on this journey.

Introduction

It is often said that change is the only constant in the modern business world. In fact, this idea has been with us for centuries. In 1531, writing in his *Discourses on Livy*, Niccolò Machiavelli observed that, "Whosoever desires constant success must change his conduct with the times."

It is, in my view, too early for us to determine if the Internet of Things (IoT) will come to be seen as the harbinger of another great evolution of human productivity, rivaling those spurred by the industrial revolution in the 18th and 19th centuries, or by the move to computer automation and information technology (IT) in the 20th. However, it is already clear that IoT is a disruptive technology, and with any disruption comes both challenge and opportunity.

Like many advances in technology, IoT is a new name for a collection of ideas that have been developing for many years. The central enabling technologies of IoT solutions—sensors, telemetry streams, data storage and analysis, and system-level command and control—have existed for decades. Certainly, the systems of the NASA space program in the mid-20th century were IoT systems without that name. What has changed is that maturing technologies have been matched by dramatic economic changes, and together these have increased capabilities while simultaneously driving down costs. Today IoT solutions can be built which would have been economically infeasible only a short time ago.

At this point, the potential impact of IoT solutions includes most areas of human endeavor—industrial systems; logistics and manufacturing; smart homes, buildings, and cities; autonomous vehicles; efficient and personalized healthcare; tailored retail experiences; and many others, including systems that enrich fans' experiences in following their favorite sports figures. Bob and Jeff have been early practitioners of IoT solutions in many of these areas, and what they have learned along the way informs and enriches this book.

IoT solutions typically connect devices, provide analysis of the data from those devices (usually paired with data from other data repositories), and then operationalize the insights derived from that analysis to act. That action could be sending command and control messages back to the connected devices, or initiating workflows in existing business systems. Very often IoT solutions connect systems that have not previously been connected, and machine learning on this newly visible data can create not only new efficiencies in existing environments, but also transformed understanding of what is possible.

Many business leaders have adopted a "disrupt or be disrupted" approach to IoT, leveraging the deeper understanding IoT solutions provide to lead a digital transformation of their organizations and position them for future success.

At Microsoft, our ongoing goal for Azure IoT is to simplify the creation of solutions that provide compelling customer value. This book explains how businesses can leverage our platform to realize their own digital transformations.

In this book, Bob and Jeff illustrate how solutions come together across the landscape of devices, sensors, device management, real-time analytics, and predictive analytics. They also provide a needed focus on how to create actionable intelligence from the system, exposed in this case via mobile dashboards and real-time text and e-mail alerts. They have created a cookbook for creating a complete solution based on a wide variety of Azure platform services, and they share with the reader all the individual recipes that make up the solution along the way. I'm delighted that they have chosen to share their expertise in this way.

April 2017

Kevin Miller
Principal Program Manager, Azure IoT
Microsoft Corporation

Chapter Overview

Chapter 1: Business in Real-Time

This chapter provides a business context for the technical topics covered in the book. What business conditions are driving the need for real-time data? What is the impact on the organization if these new mission-critical applications, big data stores and advanced analytics are deployed to the cloud? How will they be managed and maintained? Is there a logical approach and a technology roadmap that will point the way to a successful transformation to a real-time business? A principled approach and reference architecture are introduced that provide the roadmap on how to design and implement a highly scalable, secure IoT and advanced analytics SaaS solution on Azure.

Chapter 2: DevOps Using PowerShell, ARM, and VSTS

DevOps is best summed up as the union of people, process, and tools. It's an alignment of the development and operations teams, the automation of development, test, and release processes, and the selection of a consistent set of tools used to facilitate automation of the build, test, and release cycles. The goal of adopting a DevOps approach is to create a streamlined product development lifecycle that removes, to the greatest degree possible, errors that are introduced through manual steps. This chapter details how you can use Azure PowerShell, Azure Resource Manager (ARM) templates, and Visual Studio Team Services to automate the provision, build, and deploy steps of Azure hosted services.

Chapter 3: Device Management Using IoT Hub

Connecting people, places, and things to the cloud, while not trivial, may be one of the easier aspects of IoT as the techniques and protocols are very well defined. The real work begins when you have thousands of devices connected to the cloud and you need to manage the day-to-day operations of this extremely distributed system. In addition to monitoring and managing the cloud services that are providing analytics, storage, dashboards, alerts, and notifications, you also need to monitor and manage your beacons, devices, and edge gateways. This chapter examines the Device Management features of Azure IoT Hub that support command and control, device twin, and direct methods.

Chapter 4: Sensors, Devices, and Gateways

This chapter provides a glimpse into the world of sensors and devices. It touches on some of the more common sensor and device scenarios that you will encounter and how they relate and work together to create a consistent, reliable network of connected things. Patterns for the implementation of device firmware are covered.

Chapter 5: Real-Time Processing Using Azure Stream Analytics

In this chapter, we examine the use of Microsoft Azure Streaming Analytics to create jobs to process our incoming data streams from our various sensors, perform data transformations and enrichment, and finally, provide output results in various data formats.

Chapter 6: Batch Processing with Data Factory and Data Lake Store

In this chapter, we examine the use of Azure Data Factory and Azure Data Lake and where, why, and how these technologies fit within the capabilities of a modern business running at Internet speed. We first cover the basic technical aspects and capabilities of Azure Data Factory and Azure Data Lake. Following that, we detail three major pieces of functionality for our reference implementation—how to leverage reference data in our Stream analytics jobs, how to retrain an Azure machine learning model, and how to move to data from Azure blob storage to Azure Data Lake.

Chapter 7: Advanced Analytics with Azure Data Lake Analytics

The Data Lake analytics tools and capabilities help make it easier and more efficient to solve today's modern business analysis and reporting problems. It is more efficient because it offers virtually unlimited storage, with immediate access to that storage for running analytical operations on top of it. Data Lake offers the ability to persist the raw data in its native form and then run transformational and analytical jobs to create new analysis, summarizations, and predictions across structured and unstructured data. This analysis is always based on the original data. All this adds up to a "faster-time-to-value" for a modern business seeking to maximize its true potential. In this chapter, we examine the use of Azure Data Lake Analytics (ADLA), which is Microsoft's new "big data" toolset that runs on top of Azure Data Lake.

Chapter 8: Advanced Analytics Using Machine Learning and R

We are truly living in exciting times as three major trends are converging in the IT industry today—big data and the Internet of Things (IoT), cloud computing and cheap cloud-based storage, and business intelligence capabilities. Some would say that the combination of these forces is helping to usher in the fourth Industrial Revolution. It has been predicted that artificial intelligence (AI) and machine learning (ML) capabilities will be incorporated into an ever-increasing number of platforms, applications, and software services as we approach the next year. These new AI and ML capabilities will enable a new generation of business and it professionals to take advantage of artificial intelligence and machine learning capabilities without having to understand exactly how they work. In this chapter, we explore the exciting new world of machine learning and predictive analytics using Azure Machine Learning and the R programming language.

Chapter 9: Data Visualizations, Alerts, and Notifications with Power BI

In this chapter, we explore the use of data visualizations, alerts, and notifications to help today's modern business provide useful communications to their employees and customers to successfully manage their operations in real time. We start the chapter with a brief look at today's modern reporting landscape, then take a look at how Microsoft technologies like Power BI and Azure Functions can help provide quick and easy solutions. We then demonstrate enabling these technologies as part of our reference implementation scenario. We conclude the chapter by demonstrating the use of a C# .NET "SIMULATOR" application to automatically generate thousands of sample test data transactions through our Azure Cloud applications. The simulated data will be processed in real time using Azure Streaming Analytics. We also implement a Power BI dashboard that provides outputs for our new Lambda cloud architecture. Lambda architectures are designed to handle massive quantities of data by taking advantage of both batch and stream-processing methods. Our new Power BI dashboard will display outputs for all three "temperatures" of the Lambda architecture processing model and provide visualizations for our cold, warm, and hot data paths.

Chapter 10: Security and Identity

Cybercrime and IoT security have been front and center in the news this past year. Deploying an IoT solution using public cloud platforms requires an understanding of the surface areas of vulnerability and the attack vectors that cyber criminals might leverage to define your security strategy. Security is a not a solution; it is an ongoing process that requires discipline and constant analysis, review, and action. Identity is all about who can access your applications, APIs, and the underlying data that's at the heart of your IoT solution. You will want the ability to provide users some level of self-service for registration, password management, and profile updates while maintaining restrictive protocols for application capabilities and access to data. This chapter outlines a framework for analyzing potential threat vectors and the tools and protocols that Azure provides to mitigate these threats. We also examine how to implement a multi-tenant application using Azure Active Directory B2C.

Chapter 11: Epilogue

In this chapter, we reflect on each of the topics covered in the book and provide some advice on how you can get started down the path to your digital transformation.

CHAPTER 1

Business in Real-Time

Every business today is going through a digital transformation due to disruptive forces in the market, from born-in-the-cloud competitors to the increasing demands of customers, partners, and employees to engage through modern digital experiences. They are evolving from relying only on historical data to learning to use both historical and real-time data to drive innovation, evolve business strategy, and automate critical business processes.

As businesses evolve and transform to take advantage of real-time data, they will drive impact through operational efficiencies as well as create new revenue opportunities. For example, a product manufacturer can gather information about how their products in remote locations are performing and automate the scheduling of field service engineers only as needed. Retail outlets can provide real-time inventory to drive an Omni-channel shopping experience for their customers. Companies that have a need to increase worker safety can track environmental conditions such as temperature, humidity, and wind speed along with employee biometrics such as heart rate, body temperature, and breathing rate to be able to determine if an employee's physical condition would create a worker safety issue.

In each of these scenarios, companies are finding that to stay competitive, improve operational efficiencies, and engage their customers more deeply, they must learn to leverage modern software development patterns and practices. They are transforming to become Software-as-a-Service (SaaS) providers skilled in the dark arts of data science and the Internet of Things (IoT). They are learning to create applications that connect people, places, and things. They are providing real-time data visualization, alerts, and notifications. They are integrating these connected products with existing line-of-business systems and providing seamless authentication for customers, partners and employees through immersive, beautiful experiences that work on any device and are available 24/7. They are transforming to become a *Real-Time Business*.

A Platform Approach

To maximize the investment in transforming to a Real-Time Business, it is necessary to increase reuse of common business and technology capabilities. This is most effectively achieved by leveraging a platform approach. A platform approach seeks to create a common underlying set of capabilities that are accessed through managed APIs so that many verticalized applications can be built on this common substrate, as depicted in Figure 1-1.

© Bob Familiar and Jeff Barnes 2017

B. Familiar and J. Barnes, *Business in Real-Time Using Azure IoT and Cortana Intelligence Suite*,
DOI 10.1007/978-1-4842-2650-6_1

Figure 1-1. *IoT and the Advanced Analytics Platform model*

Regardless of vertical industry, there are a set of common, reusable components that you need to operationalize an IoT solution. Once those capabilities are in place, a business can build many applications that target different types of users and different vertical markets. One of the huge benefits of this approach is that all the data from all those vertical markets ends up in a common advanced analytics sub-system where cross-business insights can be drawn. These insights result in what we call *X-Factor applications,* those new revenue opportunities that were undiscoverable due to the siloed nature of the previous business model.

Real-Time Business Platform

Transforming to a Real-Time Business requires an investment in people, process, and tools. To meet the expectations of your customers and provide the business with a platform for driving impact at velocity, a new approach is required in the design, development, and deployment of your software products. Real-Time Business solutions are SaaS applications that support frequent release cycles, work on any device, and provide a multi-tenant authentication scheme and secure access to the underlying information for customers, partners, and employees.

A Real-Time Business implies that you are adopting the latest sensor, beacon, and smart device technologies for connecting products, physical environments, and people to generate real-time data. It implies that you are leveraging advanced analytics techniques such as stream processing, map reduce, and machine learning to perform real-time analytics on the constant stream of data flowing into the system. Through that analysis, you will be able to provide visual and system level alerts and notifications and high-value data visualization for consumers of the information. In addition, you will be able to integrate with existing lines of business systems to automate critical business processes.

As depicted in Figure 1-2, to be successful on this journey, organizations must apply a set of unwavering principles that provide the guideposts and rules that are used to influence each product decision along the way.

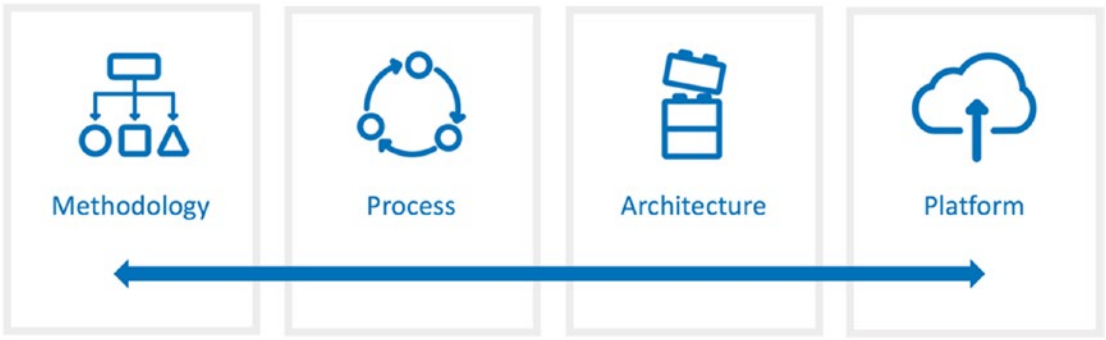

Figure 1-2. *Methodology, Process, Architecture, Platform*

These principles include a well-defined software development methodology, a standard software development process that includes not only development but also operations, the adoption of software architecture patterns that provide scalability, elasticity, and agility, and a software platform that provides the foundational services for modern software.

Methodology: Lean Engineering

Lean engineering has its root in manufacturing where the primary concept is to maximize customer value while minimizing waste, i.e., creating more value for customers with fewer resources. Applied to the software development lifecycle, the product team engages customers early in the development process by operationalizing a minimal-viable-product (MVP) and asking for feedback. Using this approach, the software product team can more easily adjust or even course-correct and thus increase product quality.

The lean engineering lifecycle, as depicted in Figure 1-3, is called Build-Measure-Learn and promotes continuous delivery, continuous analytics, and continuous feedback.

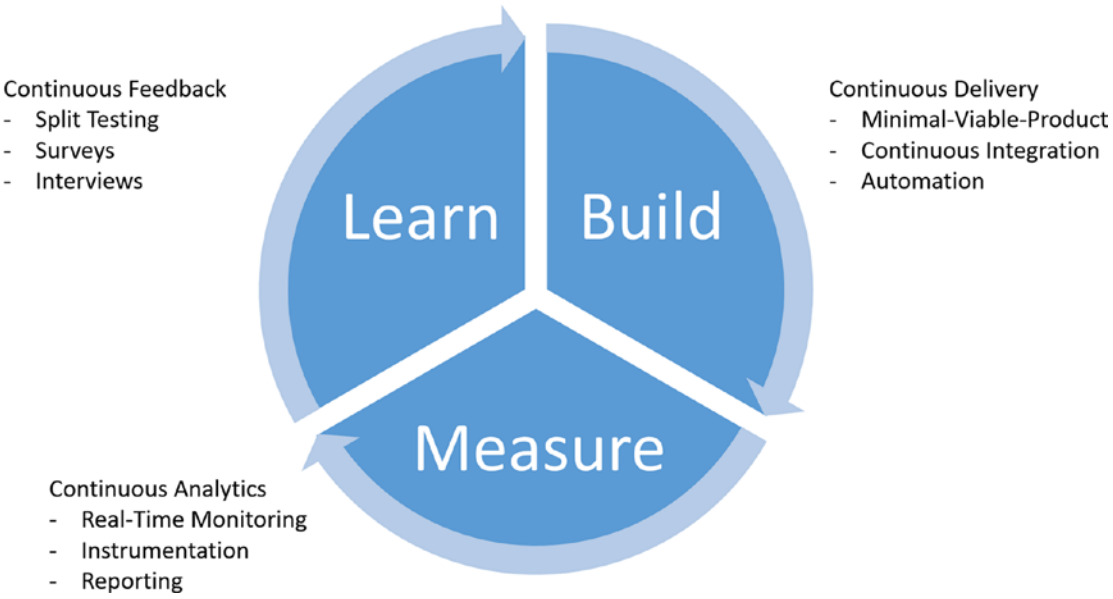

Figure 1-3. *Lean engineering cycle*

The Build phase represents the development and deployment activities; the Measure phase focuses on monitoring and reporting on the health of the software; Learn is all about engaging the customer to gather feedback that then is used to drive the next iteration of the product development lifecycle. The creation of dashboards, either custom or provided by third-party tools, provide the real-time and historical analytics from which you can derive insights quickly and steer the product development effort in the direction that meets your customer's needs.

Process: DevOps

The term *DevOps* is a mashup of development and operations. The mashup implies a deeper, more collaborative relationship between development and operations organizations. The goal of that collaboration is to define how people, process, and tools combine to automate a software development lifecycle.

DevOps, as depicted in Figure 1-4, implies the creation of cross-functional teams, combining developers, testers, and architects along with operations who together own the entire deployment pipeline from build through test to staging through to production.

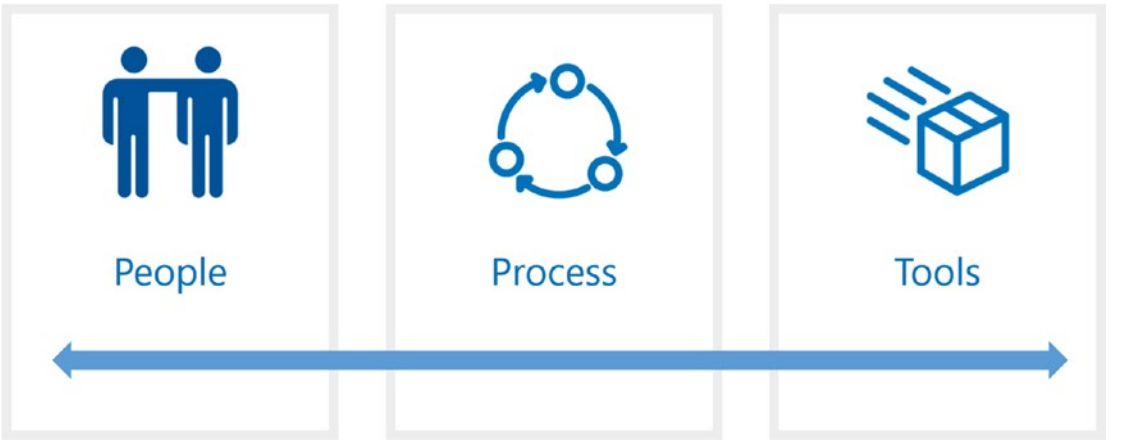

Figure 1-4. *DevOps framework*

It requires that these teams work collaboratively to adopt common processes and tools. This simple explanation has massive implications to an organization. It does not happen overnight and should be approached in a phased manner using small teams that adopt the new methods and best practices and then transition to become subject matter experts, transferring their knowledge to the rest of the staff.

Architecture: Microservices

The key attributes of microservices are:

- *Autonomous and Isolated*: Microservices are self-contained units of functionally with loosely coupled dependencies on other services and are designed, developed, tested, and released independently.
- *Reusable, Fault Tolerant, and Responsive*: Microservices must be able to scale appropriately depending on the usage scenario. They must be fault-tolerant and provide a reasonable timeframe for recovery if something does go awry. Finally, they need to be responsive, providing reasonable performance given the execution scenario.
- *Programmable*: Microservices rely on APIs and data contracts to define how interaction with the service is achieved. The API defines a set of network-visible endpoints, and the data contract defines the structure of the message that is either sent or returned.
- *Configurable*: Microservices are configurable. To be both reusable and able to address the needs of each system that chooses to employ its capabilities, a microservice must provide a means by which it can be appropriately molded to the usage scenario.
- *Automated*: The lifecycle of a microservice should be fully automated, from design all the way through deployment.

Microservice architecture has become a popular pattern in the development of highly scalable, fault-tolerant SaaS applications as it provides a more robust application architecture, supports a high velocity release cycle, and places an emphasis on managed APIs as the means to access the underlying data, analytics, and business operations. One of the benefits of this approach is that the APIs have the potential to become a new revenue channel for the business.

Platform: Cloud

Modern applications require an infrastructure and software platform that can provide high availability, fault tolerance, elastic scale, on-demand storage and compute, APIs, and tools to fully automate every interaction with the platform. The cloud platforms must also provide the foundational building block services that span IoT, big data, and predictive analytics.

Internet of Things (IoT) and Big Data

IoT is not new. The ability to connect devices to networks, gather telemetry, and display that information to garner insight and act has been around for some time. NASA pioneered the concept of data being collected by sensors and sent across space and time to be analyzed in near-real-time so that status could be visualized, insights gleaned, and action taken in an emergency (see Figure 1-5).

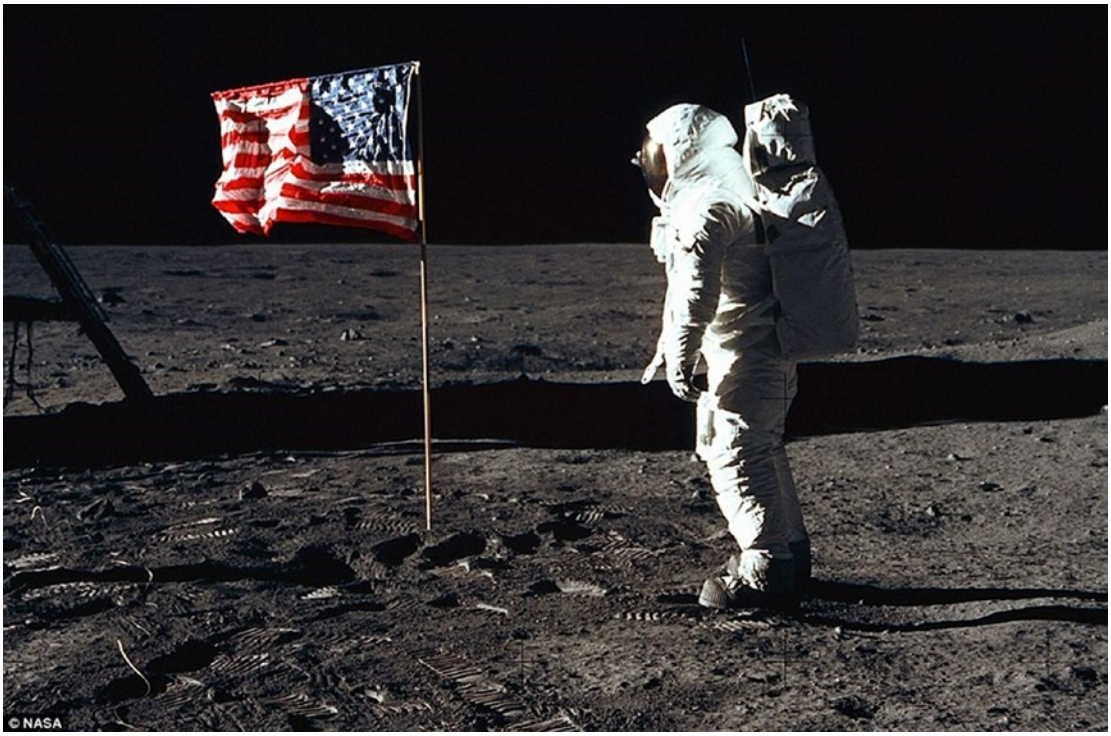

Figure 1-5. *Neil Armstrong (image credit NASA)*

NASA's Mission Control, seen in Figure 1-6, consisted of hundreds of people, each with his own collection of monitors providing data visualization of key metrics coming from the command module or an astronaut's suit. That data was an immediate measure of mission status and safety. Truly amazing when you think about what that organization accomplished given the state of technology at the time.

Figure 1-6. *NASA's Mission Control (image credit NASA)*

The Tipping Point

The one thing that NASA had that made them unique was a budget. Billions of dollars enabled NASA to put humans on the moon and, in the process, define IoT for the rest of us.

It is not likely that you have a NASA-sized budget, but you are in luck. It is no longer necessary to break the bank to IoT-enable your products and connect them to the cloud to gather telemetry, transform and store the data, gather insight, and act. Sensors and miniature microprocessor boards are inexpensive and getting cheaper and more powerful all the time. The ability to develop the code to gather sensor readings, connect to a secure cloud endpoint, and send messages has never been easier.

What is driving this thirst for IoT is data. This data will reveal the quality of the product and how it is used by customers, as well as give you the ability to calculate mean time to failure for its components and provide immediate business value through the automation of scheduled preventive maintenance. Using predictive and prescriptive analytics, you can provide an enhanced customer experience, increase product quality, and create a competitive advantage.

Big Data

What you learn very quickly with IoT is that once you have sensor-enabled people, places, and things, then it becomes all about the data. There are approximately 2 billion PCs on the planet in 2017 and about 10 billion mobile devices. By 2020, it is projected there will be over 50 billion connected devices driving Exabytes of data into the cloud.

How are we going to be able to ingest, transform, store, and analyze this data, and even more importantly, how are going to query and visualize the data so that we can quickly draw insights and act? We need to learn new skills and leverage new cloud capabilities to deal with this influx of massive amounts data.

Advanced Analytics

If data is the ore, then knowledge is the gold. One of the goals of collecting all this data is to be able to mine insight from real-time data visualizations that improve business value and automate critical business processes. A typical IoT data processing pattern is to set up cold, warm, and hot path routes for the data where the cold path provides long-term storage, the warm path provides storage for real-time dashboards, fast batch and slow batch processing, and the hot path delivers the messages via a queue to a microservice that provides real-time notification services for alerts and alarms.

Predictive analytics is a popular choice for real-time solutions because it is a data-mining process focused on predicting a future state. Data models are created from historical or sample data. These models are used along with statistical algorithms to examine real-time data streams and make a prediction. Manufacturing companies can use predictive analytics to determine when component parts of their products are about to fail and use that to automate the process of scheduling a field service engineer to visit and perform maintenance. A medical clinic may use predictive analytics to examine genomic data to see if a population is predestined for a medical condition. A retail outlet may use this approach applied to real-time inventory levels to predict when it will run out of a popular product.

Real-Time Business Reference Architecture

To realize the benefits a Real-Time Business solution requires a set of business and technical capabilities that define the process by which you ingest device events, perform advanced analytics, gather insight, and act.

The business will need to develop a model that demonstrates how processes will be improved to reduce cost and what new revenue channels will be identified and leveraged to increase profit. For example, a manufacturer can reduce costs by only sending field technicians to perform maintenance when needed. That same manufacturer may find that they can offer more flexible service contracts and thus increase sales and monthly service contract revenue.

A product development team who is responsible for supporting this new business model will need a reference architecture that provides a roadmap for how they can construct a real-time system from connected devices and cloud services.

Figure 1-7 depicts a reference architecture for Real-Time Business. Each component of the architecture represents the combination of foundational cloud services for implementing a real-time data pipeline, advanced analytics, big data storage, APIs, and the supporting automation scripts and security protocols. Together these components provide the technical capabilities of a Real-Time Business platform.

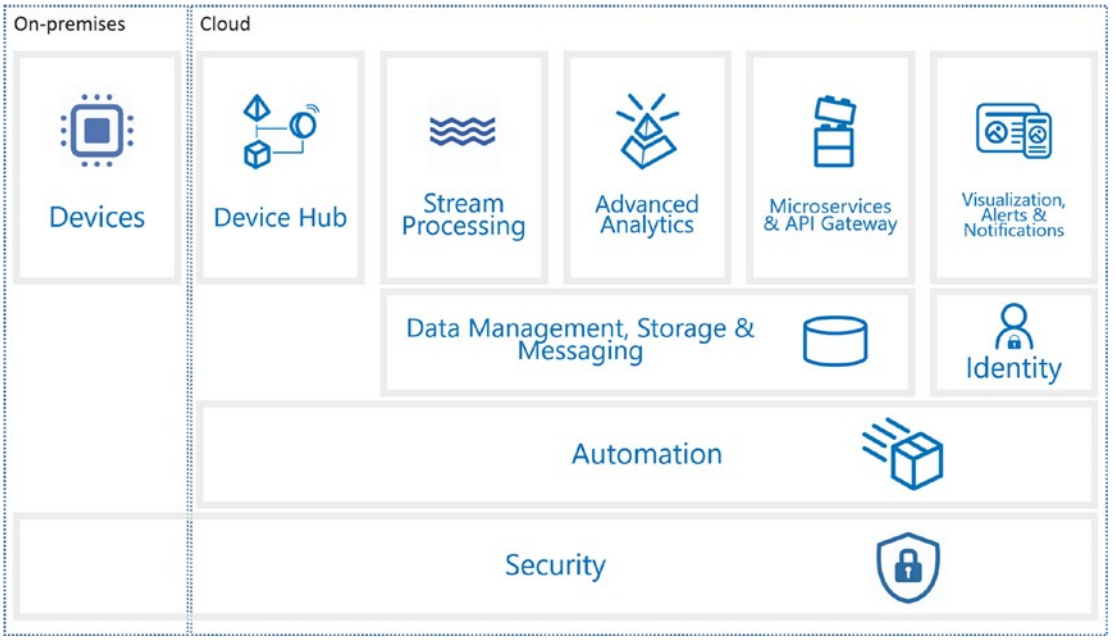

Figure 1-7. *Real-Time Business reference architecture*

Devices

Each IoT scenario requires analysis and strategy on how best to sensor-enable the product, the environment, or the people to efficiently and securely gather the data needed to drive the business case. This may involve RFID tags for location and product identification; environmental sensors for temperature, humidity, or wind speed; mechanical sensors for tracking gear revolutions or hydraulic pump iterations; or in the case of people, biometric sensors for heart rate, skin temperature, or blood glucose.

In addition to the sensors, you will need devices with an embedded operating system such as Linux or Windows 10 IoT, to host the code that manages the connectivity to sensors and the physical environments in which the device is deployed, secure connectivity via wired, wireless and/or cellular network, secure authentication to the cloud, device-to-cloud messaging such as heartbeat and telemetry, and cloud-to-device messaging for command and control and device management. Advanced scenarios may add analytics, filtering, business rules and alerts, and notification at the edge. The device will also participate in the device-management protocols for managing device state, firmware upgrades, and other remote control operations.

Device Hub

A *device hub* is a cloud-hosted service that provides secure device connectivity, telemetry ingestion, and remote command and control. This service should provide these capabilities at scale so that, as the number of connected devices grows, the service never fails. This service may provide a transient store for all incoming messages. This transient store allows real-time analytics of the message events, analysis of messages over short periods of time, as well as the ability to go back in time to reassess the events.

Device management is typically incorporated into the device hub service. Device management provides the ability to register devices using their unique identifier such as serial number. Once registered, the device will be able to connect securely to the device hub for communication purposes. Device management may leverage the concept of a *device twin,* which is a digital representation of the state of the device kept synchronized with physical devices. The device twin provides the ability to synchronize property values of the device in the cloud with the devices in the field, essentially performing desired state configuration (DSC) at scale.

Stream Processing

Stream processing is a cloud-hosted service that provides real-time analytics on incoming telemetry. The service allows you to query across the incoming messages in real time, select messages of a certain type or that contain certain values, apply aggregation and calculations over time (windowing), transform the messages, identify alarm conditions, and then act on the result of the analytics. In most cases, this service routes the resulting message to a storage location, API, or message queue for further processing.

Data Management, Storage, and Messaging

To get the most out of the real-time data now coursing through the cloud, you will want to provide various types of storage and data management, each optimized for the next step in the data processing pipeline.

- *Blob Storage*: Blob storage is typically used for event archival.
- *Store and Forward Messaging*: To provide integration with event-driven microservices or on-premises line-of-business systems, queues and pub/sub mechanisms provide loosely coupled messaging.
- NoSQL and Relational Databases: For time-based query capabilities and integration with traditional applications and modern dashboards, you can leverage relational or NoSQL storage services.
- *Data Lake*: A Hadoop Distributed File System (HDFS) that provides big data storage and cross-language query access and can be used with advanced analytics engines.
- *Extract, Transform, and Load*: You may look to use an ETL service to perform data integration and transformation operations.

Advanced Analytics

Advanced analytics is a catch phrase for all the possible big data analytics you may look to perform on your real-time data. This may involve combining the real-time data with historical and reference data and leveraging the distributed query capabilities of a Hadoop engine, using ETL tools to integrate into a data warehouse or train a predictive model to automate preventive maintenance. It is also possible to use the latest advances in Azure Cognitive APIs and artificial intelligence bots along with this real-time data to create new immersive, conversational experiences for your customers.

Microservices and API Gateways

A microservice provides a business or platform capability through a well-defined API, data contract, configuration, and an underlying data storage necessary to function. It provides this function and only this function. It does one thing and it does it well. Microservices represent business capabilities defined

using domain-driven design, implemented using object-oriented best practices, tested at each step in the deployment pipeline, and deployed through automation as autonomous, isolated, highly scalable, resilient services in a distributed cloud infrastructure. An IoT solution may have three types of microservices:

- *Transactional*: Transactional microservices are responsible for writing messages to an appropriate store.
- *Event Driven*: Event-driven microservices listen on message queues and act on the event of a message arriving on the queue. These microservices are typically used to drive alert and notification business processes or integrate with other lines of business systems that require special message handling.
- *API Contracts*: These microservices leverage ReST endpoints and JSON data models and provide the cross-cutting concerns and business capabilities that you want exposed to any consuming application.

API gateways provide secure API proxies that wrap your ReST APIs, organize APIs into products, provide restricted access to API products via the definition of developer groups, provide a subscription capability, provide policy injection, throttling, quotas, etc., and provide analytics at the product, API, and operation levels. API gateways provide the ability to segregate your APIs into private, semi-private (partner access), or public access and then monetize the APIs to create a new revenue channel.

Visualization, Alerts, and Notifications

Now that you can ingest, analyze, and store your real-time data, you will want to create customer-, partner-, and employee-facing applications that provide impactful data visualizations, visual and device-centric alerts and notifications using your APIs and event-driven microservices. You may leverage third-party services such as Twilio and Send Grid to provide text, voice, and e-mail notifications. In addition, you may look to use cloud-hosted mobile notification services for real-time updates on mobile devices.

Identity

You will want to provide access to the applications you create to your customers, partners, and employees. The identify service provides a single sign-on or a multi-tenant authentication and authorization mechanism such that the person logging into your application can only see the data and application functionality that their role provides. The identity service can provide these features in both a Business-to-Consumer (B2C) and Business-to-Business (B2B) model, including integration with your company's directory services.

Automation

Designing, developing, deploying, and operationalizing an IoT solution requires the adoption of an automated approach to the software product lifecycle. The popular term for this today is DevOps. DevOps implies that you are organized into a software product team model that places an emphasis on product quality as the code moves through an automated deployment pipeline. The team has a well-defined process and uses a set of tools to automate its work, reducing errors and improving quality. The team leverages the cloud platform to automate the provisioning of cloud infrastructure, performing build, test, and release management, and the monitoring and gathering of runtime health metrics.