

# Building Data-Driven Internet of Things (IoT) Applications

## A four-step primer

### IOT DEMANDS NEW APPLICATIONS

Automated homes. Connected cars. Smart cities. The Internet of Things (IoT) will forever change the way businesses interact with consumers and other businesses. In 2020, 25 billion connected “things” will be in use, reports Gartner.<sup>1</sup> IDC predictions are even higher as analysts estimate IoT will grow from 15 billion devices in 2015 to 30 billion devices in 2020.<sup>2</sup>

Anticipation is building for intelligent innovation. Two-thirds of consumers expect to buy connected technology for their homes by 2019, according to Acquity Group (Accenture Interactive); nearly half expect to buy wearable technology.<sup>3</sup> A Gartner survey of almost 500 IT and business leaders shows that “more than 40 percent of organizations expect the IoT to have a significant impact in the short term, and this number rises to more than 60 percent in the long term. Industries dealing with tangible equipment, products and services (such as retailers and manufacturers) are more likely to see the IoT as having a significant impact.”<sup>4</sup>

Yet as the cost of computing continues to decline and networking becomes ubiquitous, how will your business balance what IDC characterizes as the “deluge of data from devices that are connected to the corporate network?”<sup>5</sup> How will you create business impact from IoT, which Gartner defines as the network of dedicated physical objects (things) that contain embedded technology to sense or interact with their internal state or external environment?<sup>6</sup> The fastest, most effective way to take advantage of the IoT is to evolve application development to specifically address how IoT applications will report data from, prescribe actions for, and control connected devices.

---

<sup>1</sup> Gartner. “Gartner Says 4.9 Billion Connected ‘Things’ Will Be in Use in 2015,” November 11, 2014

<sup>2</sup> IDC. “IDC Predictions 2015: Accelerating Innovation — and Growth — on the 3rd Platform,” December 2014

<sup>3</sup> Forbes. “Internet of Things By The Numbers: Market Estimates and Forecasts,” August 22, 2014

<sup>4</sup> Gartner. “Survey Analysis: The Internet of Things Is a Revolution Waiting to Happen,” January 2015

<sup>5</sup> IDC. “IDC Reveals Worldwide Internet of Things Predictions for 2015,” December 3, 2014

<sup>6</sup> Gartner. “IT Glossary.” 2015

## ENABLE CONTINUOUS EVOLUTION

Whether your business is interested in optimizing utilization or driving new revenue streams, evolving operations and service delivery to take advantage of IoT is an iterative process that will challenge your legacy application development and deployment practices. Your IT organization will need to rethink processes to enable customers and employees to view, interact with, update, and control information about connected devices.

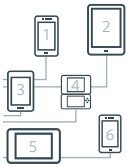
Because IoT is disruptive, it will impact the way your software development team:



**Builds more apps** – Delivering different applications to support analytical, operational, and customer interfaces with IoT data



**Iterates apps** – Providing frequent updates to each application and moving applications beyond basic reporting to prescribing recommendations to taking action and sending instructions to connected devices, as user requirements by audience are better understood



**Mobilizes apps** – Controlling mobile delivery and supporting the consumption of IoT reports from connected devices, for consumers and increasingly for operational or field-based employees

Without a solid foundation, organizations will find it more difficult, if not impossible, to build, iterate, and mobilize data-driven IoT applications. In addition to solutions that effectively integrate high-velocity and high-volume data, organizations interested in intelligent innovation need an application runtime environment that supports continuous integration, continuous delivery, and automated deployment.

Pivotal Cloud Foundry (CF) is a leading platform as a service, delivering middleware abstraction and automation that reduces code deployment time from weeks to daily or hourly. Pivotal CF also supports mobile services, accelerating mobile application development. To date, seven Fortune 500 industrial companies have chosen Pivotal Cloud Foundry for IoT build outs, improving their application development life cycles.

*“As we look to open up our Predix® software platform for the Industrial Internet in 2015, Pivotal Cloud Foundry will be a cornerstone of our cloud platform strategy. Predix will be the leading platform for secure industrial connectivity, data collection, industrial insights, and cloud deployment.”*

— Harel Kodaesh, Vice President and Chief Technology Officer, GE Software

## SUPPORT HIGHER VOLUME AND VELOCITY OF DATA

IoT brings several new realities in terms of how data is created and used. A data-centric approach is required for IoT applications, forcing organizations to adapt to these new realities:

**IoT data is distributed** – The nature of connected devices puts data creation on numerous systems

**IoT data sets will grow** – As more sensors are deployed, more data is created

**IoT data usage will change** – Use-cases for IoT are just beginning to emerge and will evolve

Larger volumes of rapidly flowing data from connected devices require new types of visualizations, analysis, and logic. As an IoT application evolves, so do the technologies required to handle IoT data. The following four modifications to existing data architectures can provide organizations investing in IoT with new insights while transforming applications:

### 1. Allow iterative data flow logic

IDC reports “80 percent of an analytics project typically involves gathering and then preparing the data for analysis.”<sup>7</sup> Without an adaptable platform to use myriad data sources, processes, and destinations, organizations face costly rework. IoT applications designed to evolve as additional requirements emerge save time and money. They can accommodate new data sources (e.g., weather, GPS, sensors, etc.), filters, sorts, and patterns.

Spring XD is a unified, distributed, and extensible system for data ingestion, real-time analytics, batch processing, and data export. It simplifies the development of big data applications by supporting numerous data source types and independently modifiable processing modules. It also reduces the time developers spend building and adapting data pipelines for new sources, transformations, and destinations. A distributed runtime, Spring XD delivers the scalability and fault tolerance high-volume, high-velocity IoT applications require and uses a declarative language that business analysts and data scientists can use without needing to program and compile code.

*“Spring XD provides a convenient way for developers to get started building a Big Data application, providing a ‘one stop shop’ for building and deploying such applications.”*

— Charles Humble, head of the editorial team at InfoQ.com

### 2. Persist data for temporal analysis

The volume of data IoT produces will increase exponentially, outstripping the growth rate of traditional enterprise data sets. Data produced by mobile things will increase from 18 percent of the digital universe in 2014 to 27 percent in 2020, predicts IDC.<sup>8</sup> As sensors connect more devices, the importance of gaining real-time views of IoT data may equal the imperative to collect data for longer-term analysis, trend detection, and predictive modeling.

<sup>7</sup> IDC White Paper (sponsored by GE Software). “A Software Platform for Operational Technology Innovation,” June 2014

<sup>8</sup> EMC. “The Digital Universe of Opportunities,” April 2014

Although organizations can now capture much greater amounts of data, many businesses only want to filter and store the data necessary to meet IoT application requirements. Hadoop has improved the economics of storing high-volumes of data, but actionable insights from trend detection and predictive modeling require an analytics stack. Pivotal HD is that stack. By integrating with HAWQ, the world's most advanced enterprise SQL-on-Hadoop analytic engine, Pivotal HD supports IoT historical analysis using familiar, broadly supported SQL skills and tool sets.



Gaining more insight from its big data is why GE in collaboration with Pivotal created the first-ever industrial data lake approach, which provides data management solutions for the industrial Internet to enable rapid insights across fleets of assets and industrial processes for data-driven business decisions. In a recent press release, GE announced it used this approach to “integrate terabytes of full flight data for the first time in industry to produce measurable cost savings of 10x and significantly reduce analysis time from months to days.”<sup>9</sup>

*“Working with Pivotal, we have created a unique industrial data approach that merges information technology with operational technology to better match the productivity and efficiency needs of our customers so they get the most value out of their mission-critical information.”*

— Bill Ruh, vice president, GE Software

### 3. Use machine learning for predictive analytics at scale

Because IoT data can now be easily captured, transmitted, processed and stored, data scientists can more quickly analyze historical trends and build predictive models with machine learning. These models then become the basis for evolving IoT applications beyond basic reporting functionality to prescriptive recommendations using logic to improve business outcomes, for example, inventory management or predictive maintenance applications. The IoT analytics market is expected to grow at a CAGR of 30 percent over the next five years, according to IDC,<sup>10</sup> but skill set shortages and scalability challenges when applying predictive models to large IoT data sets can slow implementation and the realization of business benefits.

Pivotal HD integration with HAWQ enables organizations to leverage existing SQL skills and tools to analyze IoT data stored in Hadoop. Whether businesses choose to team with data scientists for customer-facing or machine automated activities, data science helps ensure IoT applications are more accurate and effective. Data scientists at Pivotal Data Labs drive company and sector-level transformations that pave the way for data to become a new source of competitive differentiation. Pivotal Labs data scientists review

<sup>9</sup> GE. “GE Announces First Data Lake Approach for Industrial Internet to Better Access, Analyze and Store Industrial-Strength Big Data,” August 10, 2014

<sup>10</sup> IDC. “IDC Predictions 2015: Accelerating Innovation — and Growth — on the 3rd Platform,” December 2014

architecture, business goals, data science methodology and tool use, as well as paths to operationalization. They further accelerate adoption of analytics tools by building agile, analytic machine-learning models for customers across industries. Data scientists' preferred approach to quickly build algorithms that run against large-scale IoT datasets is MADLib, an SQL-based, open source library for scalable in-database analytics.

*GE is on its way to knowing “where any locomotive is at any time, what kind of problems it has, what kind of problems it will have in the future, what kind of weather it is going through, whether the operator is conserving fuel or not,” and a host of other data... I think in five years, we’ll be thinking very differently about the potential for rail in North America.”*<sup>11</sup>

— Sham Chotai, CTO of GE’s Transportation Division

#### 4. Apply real-time analysis at high-velocity and high-volume

IoT applications rely on high-speed access to incoming data for real-time analysis. Defined predictive models allow IoT data to be scored in real-time so businesses can apply historical data insights and meet incoming data requirements. Yet increasing numbers of devices fitted with sensors as well as more sensors per device have begun to tax legacy data approaches and require organizations to make unnecessary speed and capacity tradeoffs that businesses can’t afford.

Pivotal GemFire, a distributed in-memory data grid, enables GE Power & Water’s Thermal Remote Monitoring and Diagnostics Center to ingest 100,000 time series data points per second and store over five terabytes of data in memory. During a typical year, the Center receives a real-time flow of an average of five time series data points per minute from about 1.2 million unique sensors across over 1,500 turbines globally—requiring both near real-time and batch analytics.<sup>12</sup> In contrast to a fragmented architecture that processes smaller subsets of data, Pivotal enables fast, fleet-wide analytics to detect issues and optimize operations.

*“We anticipate that in-memory technologies will become a central component of solutions for applications such as remote monitoring & diagnostics.”*

— GE Global Research and GE Power & Water

#### WHY STREAM PROCESSING IS IMPORTANT FOR MACHINE LEARNING

Machine Learning algorithms need extensive training and learning based on historical data sets before they can analyze real-time inputs. Data streaming pipelines allow the implementation of closed-loop scenarios, where the data being ingested is used for real-time insights, but also immediately become available as historical data, allowing continuous learning and self-improving algorithms. Spring XD is highly available, supports stream processing at scale and allows business analysts and data scientists to easily add or change a data pipeline dynamically, in order to accommodate new data sources, processing steps or destinations.

<sup>11</sup> Fast Company: “Behind GE’s Vision for the Industrial Internet of Things,” June 18, 2014

<sup>12</sup> IEEE Xplore. “Bridging high velocity and high volume big data through distributed in-memory storage & analytics,” October 2014

IoT has the potential to make a tremendous impact on your business interactions with customers, partners, and employees. McKinsey Global Institute estimates the total economic impact will be \$2.7 trillion to \$6.2 trillion per year in 2025.<sup>13</sup> Yet transformation requires taking a first step then evolving and iterating applications so they go beyond simply reporting to prescribing and ultimately controlling a network of connected devices and the vast volumes of data produced.

Pivotal is powering enterprise transformation by serving as a trusted IT partner to enterprises building, evolving, scaling and adapting IoT applications. Contact us to learn how Pivotal Cloud Foundry and our portfolio of big data solutions can advance your software-driven experiences.

### LEARN MORE

To learn more about our products, services and solutions, visit us at [pivotal.io](https://pivotal.io)

---

<sup>13</sup> McKinsey & Company. "Disruptive technologies: Advances that will transform life, business, and the global economy," May 2013

Pivotal offers a modern approach to technology that organizations need to thrive in a new era of business innovation. Our solutions intersect cloud, big data and agile development, creating a framework that increases data leverage, accelerates application delivery, and decreases costs, while providing enterprises the speed and scale they need to compete.

Pivotal 3495 Deer Creek Road Palo Alto, CA 94304 [pivotal.io](https://pivotal.io)

Pivotal, Pivotal Cloud Foundry, and Cloud Foundry are trademarks and/or registered trademarks of Pivotal Software, Inc. in the United States and/or other Countries. All other trademarks used herein are the property of their respective owners. © Copyright 2015 Pivotal Software, Inc. All rights reserved. Published in the USA. PVTL-WP-04/15