

# Creating the Foundations for the Internet of Things





# Using big data and fast data to drive advantage



Companies that manufacture and support physical 'things' are facing a revolution driven by improvements in connectivity, storage and processing efficiency. The objects themselves have evolved from being passive recipients of support to become active players in their own lifecycle. Previously, the only way to see if something was broken was to send someone out to check. Now, devices are able to issue alerts that they could be about to fail or need service, requiring a support call before the problem has even occurred.

The term given to this revolution in connectivity is The Internet of Things. The term describes the way these new devices and objects are connected to the internet and stream information, enabling improved support and smarter decision-making.

To be ready for the Internet of Things there are four broad challenges that an organization needs to address:

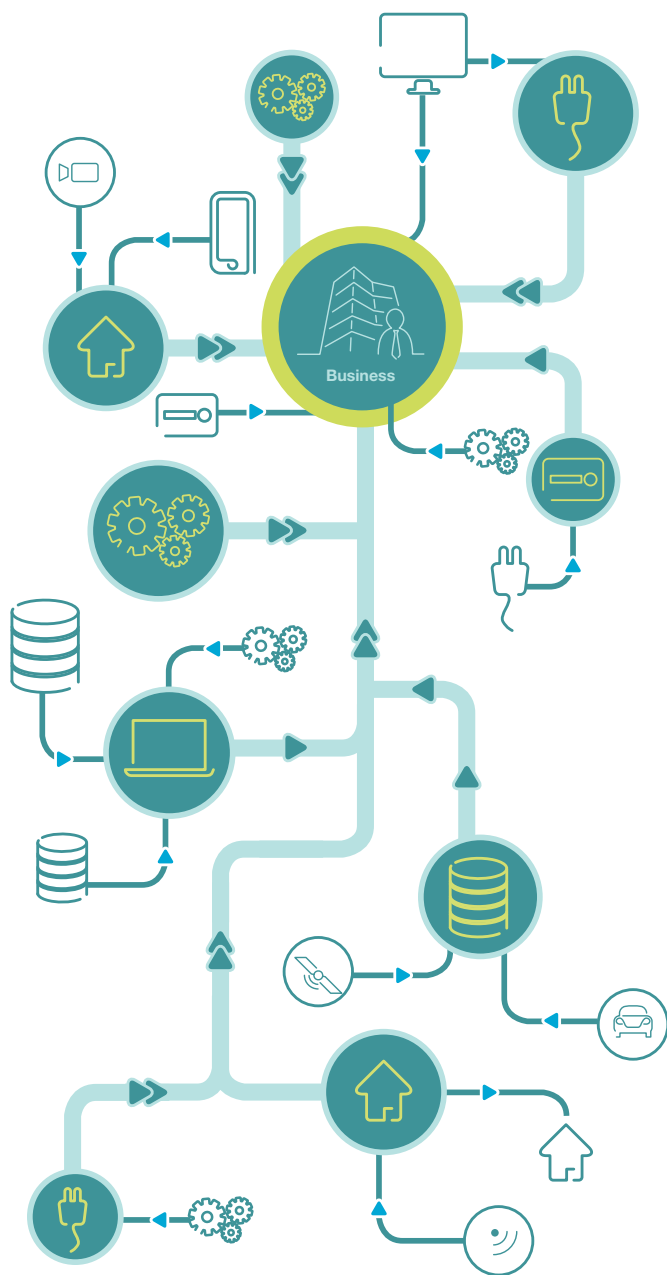
1. Storing large volumes of information
2. Handling high volume streams of information from devices
3. Undertaking predictive analytics based on historical data
4. Using machine learning to drive adaptive analytics to react in real time.

The value comes from the analytics. Global manufacturer, and leader in the Internet of Things, General Electric touts a principle they call "The Power of One". For example, a one percent operational improvement, powered by better use of big data, would create a \$27bn reduction in system inefficiency in freight utilization. A one percent reduction in process inefficiency in healthcare would lead to \$63bn in savings, and a one percent improvement in predictive diagnostics would drive a \$66bn value in efficiency improvement in gas-fired power plant fleets.

If one percent can have such an impact, the potential value of the Internet of Things is almost inestimable. The biggest challenge today is creating an infrastructure to deliver that value.

# Fast makes Big

The Internet of Things brings a simple but fundamental challenge: lots of devices that are constantly sending out information.



To be ready for this explosion of information your company needs to be able to do two things: ingest at speed and store cheaply. If a device produces 1kb of data every second (a reasonably modest amount) this will generate 315 gigabytes of data each year. If a company has 1,000 devices this creates 320 terabytes of information each year. If you were to use traditional data warehouse technologies – which can cost anything between \$30k and \$80k a terabyte – this would cost \$9m to \$25m to store. Clearly, a different approach is required.

This different approach is to utilize a big data technology like Pivotal HD which dramatically reduces the cost, so that 320 terabytes can now be stored for under \$1m. This makes the return on investment much faster which helps to spur innovation and drives new services to market more quickly.

### Traditional HD Cost



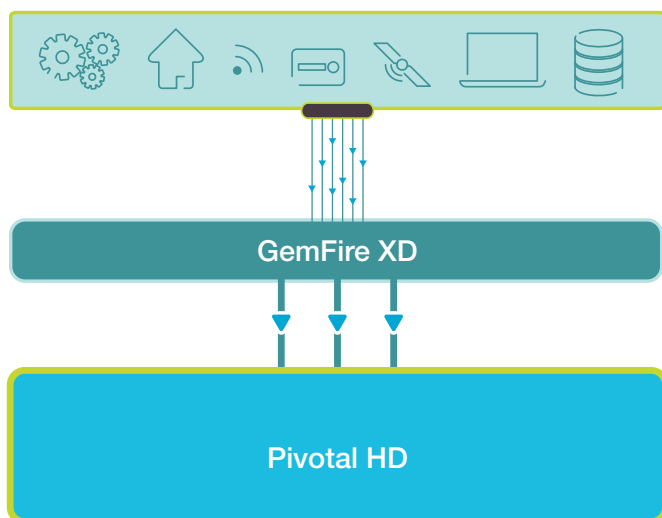
## Pivotal HD Cost



## Fast ingest for now and the future

The first challenge, however, is to handle very fast ingestion speeds. Traditional enterprise data warehouse (EDW) approaches have focused on the batch movement of data: large bundles of information being loaded in one go. The problem with this, when approaching the Internet of Things, is that devices are constantly running. So a batch solution can never respond to the 'now,' only to the past and misses a significant opportunity for adding value.

This is not a new problem, however. High-performance websites have been facing this challenge for many years and have solved it by using in-memory database technologies which then 'archive' information once it is no longer needed for active processing. Pivotal's Gemfire XD follows this approach, providing a lightning fast in-memory ingest and analytics engine which automatically archives information to Hadoop for later predictive analytical processing.



This approach also delivers a steady stream of information which is easier to handle than attempting to load massive volumes every night hoping they will be processed in time.

### Analytics not reporting

Once you have the foundation to handle fast data ingestion and big data storage it then becomes possible to extract value from that information. To do this you need to stop thinking about information and BI as being simply about reports and dashboards. There is some value to be had by reporting against device information but the real value comes through the application of data science. Reporting looks backwards at what has happened but data science looks forwards.

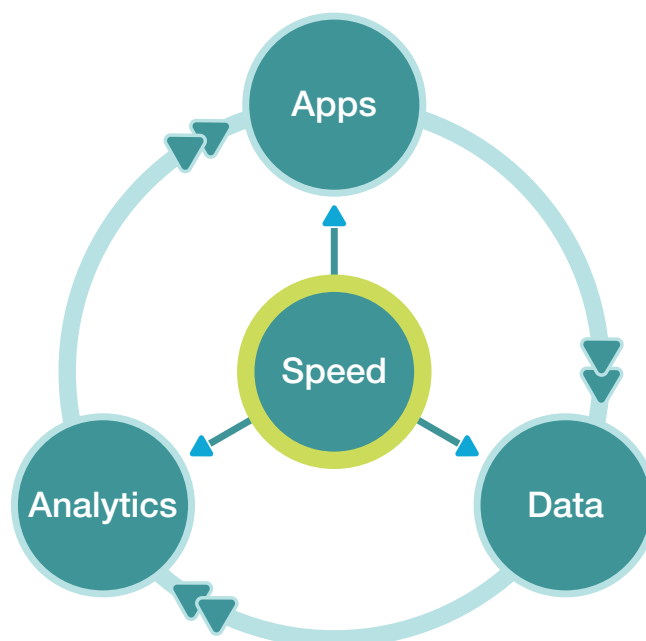
Data science is about applying new types of analytics, designing new algorithms and tuning existing ones to turn these massive data sets into insight. This requires new sets of skills but the good news is that you don't need to hire and manage PhD level mathematicians. Working together, the Pivotal and Capgemini data science teams can either design wholly new algorithms or tune existing ones to add predictive analytics on top of device data. Technologies such as MADLib help your internal teams take control of the power of advanced analytics without having to become experts themselves.

The Capgemini and Pivotal teams will work with your BI and business leaders to create the right types of predictive analytics and ensure the insight created is actionable and will deliver value.

### Dashboards for monitoring and action

Once the analytics has been done, you face the challenge of how to represent that information to your customers and operations teams. This is where the use of Pivotal's HAWQ and SQLFire technologies are used to distill information from the advance analytics into more traditional structured data stores. This allows information to be used in the creation of dashboards and also to create new applications that are driven from the information and insight derived.

This information can be combined with other critical enterprise information such as customer priority, contractual SLAs or inventory levels to create a fully integrated view of what is required and what should be done.



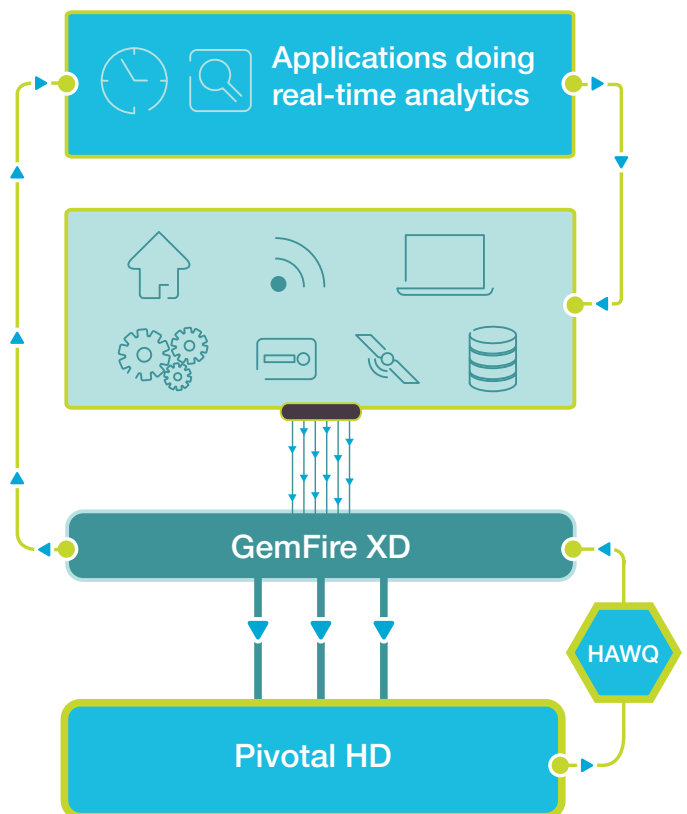
This ability to construct applications and to integrate information back into operational systems helps drive significantly higher levels of operational effectiveness than are possible today. These dashboards and applications will help deliver pre-emptive maintenance and monitor device health. They will also provide a potential new revenue channel for companies looking to offer value-added services alongside and supplementing the devices themselves.

By pushing analytical models that are calculated based on the large history within Pivotal HD you can create a new generation of algorithms that are routinely tuned via Pivotal HD analytics and react in real time to the flow of information.

### Reacting as it happens

The final stage for an organization, when looking at the Internet of Things, is how to leverage predictive analytics at the point of action. While Hadoop is cost effective it cannot react to changes in real time. This means that while Pivotal HD can predict that a device is liable to fail and that fixing it tomorrow would be more cost effective, it cannot deal with events as they happen, for instance if a device is broken due to a physical mishap.

This is where the ability to handle real-time ingestion of data moves beyond merely helping to smooth out loading and towards adding the next stage of value. By pushing analytical models that are calculated based on the large history within Pivotal HD you can create a new generation of algorithms that are routinely tuned via Pivotal HD analytics and react in real time to the flow of information.



This gives you the ability to combine heavy statistical analysis with the latest information to further optimize operations and service.

For instance, one of the leading suppliers of railroad equipment is collecting and analyzing real-time sensor performance data from locomotive operations and from the train tracks

themselves to perform diagnostics and root cause analysis that prescribes predictive maintenance before a problem occurs.

Agriculture and farming organizations are looking to use sensor data from tractors to identify where to best plant seeds for higher crop yields. They are measuring the moisture content of the soil, nutrient density, exposure to sun and shade, spacing between rows, and other factors to automate and improve the planting process.

### Where to start

The first step to achieving maximum benefit from the Internet of Things is to create foundations. This means putting in place a next-generation information architecture designed to handle the Internet of Things and combine the information with traditional enterprise data sources.

Capgemini and Pivotal have created the Business Data Lake (BDL) reference architecture to address the challenges of big and fast data. The BDL enables device information to be

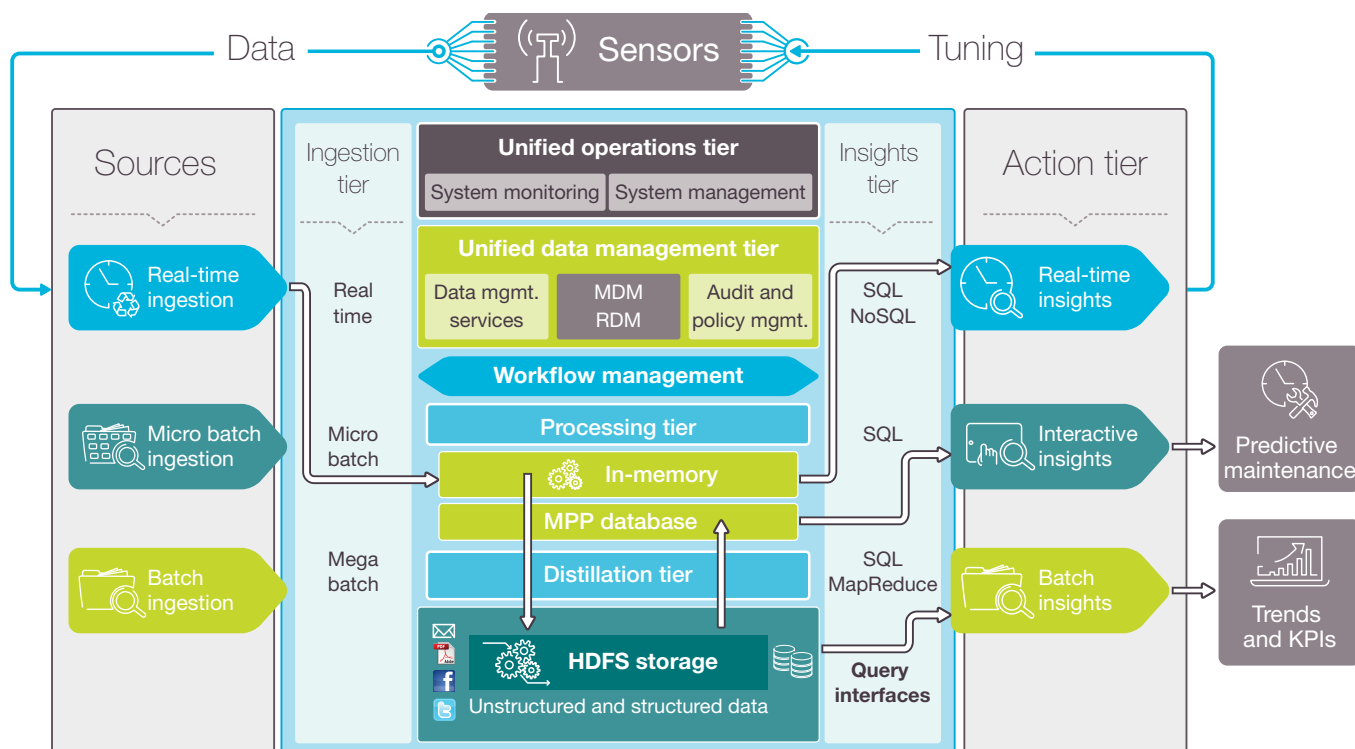
on-boarded and integrated in a way that scales separately to traditional BI requirements, minimizing costs while enabling sharing.

Gemfire XD and Pivotal HD form the foundation, respectively ingesting and storing your device information and so ultimately delivering value from the data.

### Results at the heart of your business

The Internet of Things is a growing opportunity for organizations, one that is built around the ability to analyze and react to information at two different paces. The first is at the pace of the ‘thing’ itself, to tune, to optimize and to help it react more effectively. The second is the pace of the business, to reduce costs, optimize outcomes and better assure its future.

An architectural approach which addresses both of these opportunities embeds the results into the heart of a business. This turns the Internet of Things from a massive data silo into a core feature of your basic operations.







Find out more at [www.capgemini.com/bdl](http://www.capgemini.com/bdl)  
and [www.pivotal.io](http://www.pivotal.io)

Or  
contact us at [bim@capgemini.com](mailto:bim@capgemini.com)



## About Capgemini

With almost 140,000 people in over 40 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2013 global revenues of EUR 10.1 billion.

Together with its clients, Capgemini creates and delivers business and technology solutions that fit their needs and drive the results they want. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.