

# InfoWorld DeepDive



## The best hardware and software of the year

INFOWORLD'S 2015 TECHNOLOGY OF THE YEAR AWARD  
WINNERS SHOWCASE THE **BEST TOOLS** AND **TECHNOLOGIES** FOR  
DEVELOPERS, IT PROFESSIONALS, AND BUSINESSES

## Deep Dive

# Technology of the Year 2015:

## The best hardware and software of the year



*InfoWorld's 2015 Technology of the Year Award winners showcase the best tools and technologies for developers, IT professionals, and businesses.*

Like InfoWorld itself, InfoWorld's Technology of the Year Awards have always been about change. We keep an eye out for the platforms and tools pushing against the barriers in application development, mobile, cloud computing, and in other corners of information technology, and we bring them in for review. At the end of the year, we get together and decide which are the very best.

It should be obvious to everyone that technology is evolving faster than ever, and we found ourselves working overtime in 2014 trying to keep up. New languages, frameworks, and [platforms for building enterprise mobile apps](#) — invariably powered by Node.js — demanded our attention. The arrival of Java 8 prompted refreshes of the [major Java IDEs](#). The PaaS wars led us to the inevitable dilemma: Cloud Foundry or [OpenShift](#)? OpenShift took the crown this year, but keep an eye on [IBM Bluemix](#) in 2015.

### Little big iron

Even data center hardware is evolving at a break-neck pace as the vendors work to shove more compute power, more IO, and more storage into smaller spaces.

One of our hardware winners, HP Moonshot,

packs a whopping 45 server cartridges — ARM, Atom, Opteron, or Xeon — into a 4.3U chassis. With energy-sipping cartridges designed for specific workloads such as virtual desktops and distributed data processing, Moonshot is a triumph of high density and high efficiency.

Another winner, the Dell PowerEdge R730xd, brings together so much flash, disk, and elegance, it makes an ideal platform for software-defined storage solutions such as VMware's remarkable [Virtual SAN](#) (also a winner). The R730xd shows that innovation still lives in 2U, two-socket servers.

It's fascinating what some vendors can do with commodity parts. The magic, of course, is in the software. In the Tintri VMstore, for example, the Tintri OS uses inline deduplication, compression, and other tricks to turn one part MLC flash and 10 parts SATA disk into something akin to an all-flash array for virtual machines. Lightning-fast flash delivers the IO, while disk provides the capacity.

Managing storage with Tintri might seem magical because Tintri operates exclusively in terms of virtual machines and virtual disks, rather than intermediate concepts like LUNs and volumes. After all, isn't it high time we assumed that all machines are virtual machines?

You might think so ... until something better



**It should be obvious to everyone that technology is evolving faster than ever, and we found ourselves working overtime in 2014 to keep up.**

## Deep Dive

comes along — say, like Linux containers, which offer the isolation of a virtual machine in a much lighter footprint. Think tiny, OS-less VMs with an instant-on button.

### Container Craze

It's a rare product that earns a Technology of the Year Award before reaching version 2. But have we ever seen anything like [Docker](#)? A new tool for building, running, and sharing Linux containers, Docker addresses so many pain points in application development and application lifecycle management, it would be difficult to overestimate its impact. Docker makes it easy to ensure that applications are built the same way every time and run exactly the same in development, testing, and production.

Docker helps streamline and simplify devops. Developers own what's inside the container, and operators own what's on the outside.

Docker is even changing how applications are designed (as loosely coupled containers running microservices) and how the operating system is conceived (as a minimalist runtime for containers). It's an ambitious project with big issues remaining to be solved, but it has broad industry support and a rapidly growing user base. It's exactly the sort of technology that the Technology of the Year Awards were made for.

After so much experimentation in 2013, it was an oddly quiet year in mobile devices, with Apple's [iPhone 6](#) and [iPhone 6 Plus](#) the clear highlights. Apple impressed us even more [with Handoff](#), a feature in the latest versions of OS X and iOS that allows you to begin a note or an email or another task on one device, and pick up right where you left off on another. Dubbed "[liquid computing](#)" by my InfoWorld colleague Galen Gruman, Handoff is the kind of innovation that really matters to users and an easy pick for a Technology of the Year Award.

### Something old, something new

The Technology of the Year winners circle is usually a gathering of both established products and young upstarts. We've seen the iPhone on this list before, and the same can be said for [Node.js](#), the nimble JavaScript server, and [GitHub](#), the new center of the universe for open



source code. One of this year's winners, [Microsoft Office for iOS](#), might count as both established product and upstart. It's certainly the first non-Windows Microsoft product to win our award — another sign of how fast the technology landscape is changing.

Technology of the Year veteran Hadoop has owned the big data stage for years, but the time has finally arrived for the elephant to share the spotlight. Recent versions of Hive, a companion Apache project, finally deliver on the promise of real-time, SQL-like queries of Hadoop data, moving us closer to the day when the open source big data stack can replace the traditional enterprise data warehouse.

The push toward real-time and event stream processing is also shifting attention to projects such as [Apache Spark](#) and [Apache Storm](#). Spark offers an ingenious way to do distributed processing in memory, and it can be run on Hadoop or a stand-alone cluster. Storm is a distributed stream processing system that integrates with any message queue and any database, and work is underway to integrate it with Hadoop.

So just as Hadoop has been freed from the bonds of MapReduce and batch processing, these fresh new options for distributed, real-time processing have emerged to take their place alongside it. They also take a place alongside 30 other 2015 Technology of the Year Award winners.

To learn about all 32 of our winning products, see our slide show. — DOUG DINELEY



**It's a rare product that earns a Technology of the Year Award before reaching version 2. But have we ever seen anything like Docker?**

## Deep Dive



## Apple iPhone 6

The iPhone line was feeling tired the last couple of years. Apple finally refreshed the iPhone's design this year, [going big in the iPhone 6](#) and bigger in the [iPhone 6 Plus](#), its first phablet and [already the best on the market](#). What a difference! Both devices showcase Apple's attention to user experience, with sensuous cases and a choice of view modes to accommodate eyes of all ages.

Under the hood, the new iPhones showcase impressive, industry-changing technology. The Apple Pay service — embedded into the iPhone's hardware and OS — quickly became the darling of credit card processors because it's secure and easy to use, a sorely needed combination in today's payments technology.

Apple's Touch ID fingerprint sensor now has APIs to enable direct access by third-party apps, a capability that banks and others are quickly adopting to secure sensitive transactions and escape the never-ending forgotten-password business.

The new iPhones also bring iOS 8.1, [which makes business communications more compatible](#) than ever with Microsoft Exchange and improves on Microsoft's own Exchange clients in many regards. It enables a [new extensions capability](#), which lets apps work together more easily — most people will see this in the [newfound ability for cloud storage services](#) like Dropbox and Box to plug directly into apps.

What you get in the new iPhones is the most capable and sophisticated mobile operating system married with the most capable mobile hardware today. The emphasis is on capabilities, with the iPhone elegantly doing more of them than any competitor. — GALEN GRUMAN



## Apple Handoff

The most compelling client technology of the year is the [Handoff capability](#) introduced in Apple's iOS 8 and OS X 10.10 Yosemite, running on Lightning-equipped iOS devices and 2012-or-later Mac models.

I call it [liquid computing](#): the ability to transfer a task you're working on from one device to another, with both the content and state retained. You don't need a server or Internet connection. Bluetooth identifies which

**I call it liquid computing: the ability to transfer a task you're working on from one device to another, with both the content and state retained.**

of your devices are available for handoff, and Wi-Fi Direct transfers the data and its context straight to the desired device.

Apple's implementation is proprietary to its hardware and operating systems, though third-party apps on those platforms can use the APIs. That makes it work out of the box but limits the promise to Apple users only.

However, the concept is one that any platform provider can adopt — in fact, Google now has its own version in the works for Android and Chrome OS. Perhaps the open source community can find platform-agnostic ways to implement similar APIs across apps that leverage the already-standard Bluetooth and Wi-Fi Direct transport protocols Handoff relies on.

Breaking out of the client/server box that today's federated devices are kept in has [huge potential for flexible person-centered workflows](#) in many environments, as well as some security advantages. — GALEN GRUMAN

## Deep Dive

## Microsoft Office for iOS

**A**t the beginning of 2014, Office for iPad was an OK suite, with simple core capabilities. Today, it's a very good suite (composed of Word, PowerPoint, and Excel) that could serve as the only Office suite for many users. Although Apple's iWork suite (Pages, Keynote, and Numbers) for iPad and iPhone is a little better, [Office is no longer playing second fiddle](#) and is a thoroughly capable tool in its own right. Last month Office became available in a capable version for the iPhone, no longer only the iPad.

It still surprises me to write that, given [Microsoft's history of poor mobile apps](#), but it's true.

If Microsoft delivers an equally capable Office for Android this winter, as expected, and for Windows Phone and Windows 10 tablets this fall, [as is widely hoped](#), we'll finally have the modern, touch-savvy version of Office [we've wanted for years](#). Office's future as the dominant productivity suite will be all but assured.

Sure, business use of Office for iOS (and for other mobile devices) requires an Office subscription, but given how much IT organizations pay each year for Microsoft maintenance plans, I bet the ultimate cost will stay the same, and IT will be happy to get the policy-manageable Office app in users' hands on every device they have. — GALEN GRUMAN



**Head and shoulders above any Windows laptop I've ever tested, the Dell XPS 15 Touch sets the standard for screen quality, keyboard mashability, and trackpad responsiveness, complemented by performance that defines the top of the line.**



## Dell XPS 15 Touch

**H**ead and shoulders above any Windows laptop I've ever tested, the [Dell XPS 15 Touch](#) sets the standard for screen quality, keyboard mashability, and trackpad responsiveness, complemented by performance that defines the top of the line. Toss in an amazingly long-running battery and a surprisingly svelte and sophisticated case, and you have a package that outshines the MacBook Pro Retina. It's simply the best Windows laptop ever made.

If you don't mind squinting, Dell's newly announced 13-inch little brother, the XPS 13, has many of the same features in a considerably smaller case. But if you want the best of the best and have the cash to match, the XPS 15 with 3,200-by-1,800 touchscreen runs rings around the competition. — WOODY LEONHARD

## Deep Dive



## HTML5

Despite all the [logistical](#), technical, and [philosophical](#) struggles that have dogged [HTML5](#) and [its implementations](#) over the course of its seven-year gestation period, the Web standard is at last an actual, ratified standard. It's not buried in the pages of an obscure ISO spec document, but is in active use on millions of websites, in billions of browsers, and in countless desktop and mobile applications.

HTML5 owes a good deal of its success to the browser-makers — mainly the teams at Google and Mozilla whose perpetual update cycle allowed bleeding-edge HTML5 features to be used in the real world. Unlike the ill-fated XHTML, HTML5 became an easily adopted default for new websites and applications, and it provided a unified way to handle a plethora of tasks previously relegated to external plug-ins, such as pixel-accurate drawing, video and audio, external data storage, geolocation, and speech synthesis and recognition.

Not every feature in HTML5 has made people happy, though. The Encrypted Media Extension standard, for

**Not every feature in HTML5 has made people happy, though. The Encrypted Media Extension standard, for instance, bristled hairs in the free-software and open-Web communities (though Tim Berners-Lee himself gave it the thumbs-up).**

instance, bristled hairs in the free-software and open-Web communities (though Tim Berners-Lee himself [gave it the thumbs-up](#)). But the whole of HTML5 has brought a Web that is closer to being a platform, one that runs across a panoply of devices and with fewer external dependencies than ever.

— SERDAR YEGULALP



## Famo.us

**F**[amo.us](#) is the only JavaScript framework that includes an open source 3D layout engine fully integrated with a 3D physics animation engine that can render to DOM, Canvas, or WebGL. The tools you need to build Famo.us apps and sites will always be free and available to everyone. Integrations between Famo.us and Angular, Backbone, Cordova, and jQuery are under way. Famo.us uses Node.js and Grunt in its tooling. A well-written Famo.us iOS or Android app can perform as well as a native app, while being much easier and faster to develop.

**In addition to the framework and tools, Famo.us provides free training through interactive online lessons, branded as Famo.us University and currently offering four online courses.**

In addition to the framework and tools, Famo.us provides free training through interactive online lessons, branded as Famo.us University and currently offering four online courses: Famo.us 101 (basics); Famo.us 102 (layouts, transitionables, and animations); Layouts (header-footer, grid, flexible, and sequential); and Famo.us/Angular (Angular bindings with Famo.us layouts and transitionables). I expected a course in using the Famo.us physics engine by now, but no such luck.

— MARTIN HELLER

## Deep Dive

## AngularJS

**A** [ngularJS](#) is a lightweight, open source JavaScript framework for building Web applications with HTML, JavaScript, and CSS, maintained by Google and the community. It offers powerful data binding, dependency injection, guidelines for structuring your app, and other useful features to make your

**It offers powerful data binding, dependency injection, guidelines for structuring your app, and other useful features to make your Web app testable and maintainable.**

Web app testable and maintainable. Its most notable feature, two-way data binding, reduces the amount of code

written by relieving the server back end from templating responsibilities. Instead, templates are rendered in plain HTML according to data contained in a scope defined in the model.

You can bind an Angular module to a given section of an HTML document using the `ng-app` tag, and a controller using the `ng-controller` tag; the actual controller code lives in JavaScript code that is typically maintained in a separate file and included using a `script src` tag. Data binding locations in HTML markup are signified by mustache markup — for example,

`<span>{{remaining()}}...</span>`, where the contents of the mustaches will be updated whenever the value of the code inside them changes.

The `ng-submit` tag can redirect a form submit action to an Angular method. AngularJS provides built-in services on top of XMLHttpRequest as well as various other back ends using third-party libraries. For instance, the AngularFire library makes it easy to connect an Angular app to a [Firebase](#) back end. — MARTIN HELLER



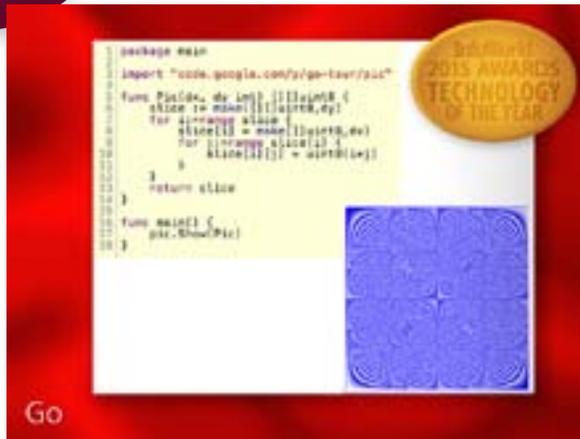
## Node.js

**B**uilt on Chrome's V8 JavaScript runtime, [Node.js](#) is a platform that allows developers to easily construct fast, scalable network applications. Node.js uses an event-driven, nonblocking I/O model that renders it lightweight and efficient, compared to, say, Java Web Pages or ASP.NET. Node.js also lets developers make code asynchronous without the mess of threads and synchronization. Node.js is well suited for data-intensive, real-time applications that run across distributed devices.

All is not entirely sweetness and light in the Node community, however. [As reported elsewhere on InfoWorld](#), "Node.js devotees who are dissatisfied with

Joyent's control over the project are now backing their own fork of the server-side JavaScript variant, called io.js or iojs." According to Mikeal Rogers of Digital Ocean, the idea of the fork is "to get the community organized around solving problems and putting out releases." — MARTIN HELLER

## Deep Dive



## Go

The [Go programming language](#) is an open source programming language from Google that makes it easy to build simple, reliable, and efficient software. It's part of the programming language lineage that started with Tony Hoare's Communicating Sequential Processes, and it includes Occam, Erlang, Newsqueak, and Limbo. The top differentiating feature of the language is its extremely lightweight concurrency, expressed with goroutines. The project currently has more than 500 contributors, led by Rob Pike, a Distinguished Engineer at Google, who worked at Bell Labs as a member of the Unix Team and co-created Plan 9 and Inferno.

Go's concurrency mechanisms make it easy to write programs that get the most out of multicore and networked machines, while its novel type system enables flexible and modular program construction. Go compiles quickly to machine code, yet has the convenience of garbage collection and the power of runtime reflection. It's a fast, statically typed, compiled language that feels like a dynamically typed, interpreted language.

Goroutines, channels, and select statements form the core of Go's highly scalable concurrency, one of the strongest selling points of the language. The language also has conventional synchronization objects, but they are rarely needed.

Goroutines are, to a rough approximation, extremely lightweight threads. Channels in Go provide a mechanism for concurrently executing functions to communicate by sending and receiving values

**Go's concurrency mechanisms make it easy to write programs that get the most out of multicore and networked machines, while its novel type system enables flexible and modular program construction.**

of a specified element type. A select statement chooses which of a set of possible send or receive operations will proceed. It looks similar to a switch statement but with all the cases referring to communication operations. – MARTIN HELLER



## Docker

All great innovations revolve around a simple idea. [Docker](#), the open source application containerization system that started on Linux (and will [head to Windows eventually](#)) is based on an idea that's as surpassingly simple and transformative as they come. Take an application, wrap it in a container that allows it to be easily deployed on any target system, and deploy it anywhere the Docker host is running. Apps run with the isolation of VMs, but they can be instantiated far more quickly, and they consume far less overhead.

Docker has done more than demonstrate a nifty way to package and isolate apps. It has sparked new approaches to devops, to turning applications into microservices and deploying them at scale, and even to designing the underlying operating system, whereby Linux itself is rebuilt around containerized applications as a basic unit of construction (see: [CoreOS](#), [Red Hat Atomic](#)).

That said, Docker is young and protean. Questions linger regarding the project's treatment of issues like networking and security in the long run. But a wide range of third-party contributors stand behind both the core technology and the needed auxiliary functionality like [orchestration](#); [most every cloud vendor](#) is on board with Docker as a key technology; and the speed at which Docker is evolving makes a bold statement about its future. – SERDAR YEGULALP

## Deep Dive

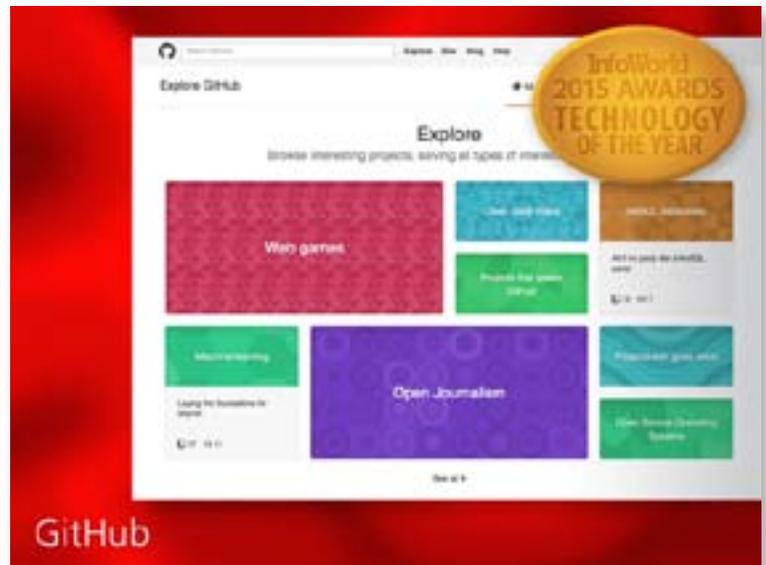
## GitHub

**G**itHub is one of about 18 public Git hosting sites, and it supports both public, open source projects and private, proprietary code. Public repositories are free; private repositories cost money to host, but only about \$1 per repo per month. Each repo can be up to about 1GB; that isn't a terrible limit if you restrict yourself to storing source code and a reasonable number of small images, but you can run out of space quickly if you try to store binary builds, media, external dependencies, backups, or database dumps. GitHub will warn you if you push files past 50MB and will reject files exceeding 100MB.

Where GitHub differentiates itself is in the social aspects of coding and in its client software. While you can't really use Git effectively unless you can drop down to the command line at need, the GitHub client does a good job of implementing the Git features you need on a daily basis, and it automatically updates itself. In addition, the client integrates with GitHub's very nice, free Atom programming editor, which in turn integrates well with GitHub repositories.

**Where GitHub differentiates itself is in the social aspects of coding and in its client software.**

The social aspects of GitHub – following people, forking and watching projects, making pull requests, reporting issues, and sharing Gists – are important enough that I tell developers they should be on GitHub no matter where else they keep or use code repositories. Plus, some of the most important and popular Open Source projects are on GitHub, including Bootstrap, Node.js, Angular, jQuery, D3, Ruby on Rails, and the Go language. – MARTIN HELLER



## AnyPresence

**T**he goal of [AnyPresence](#) is not only to help enterprises build back ends for their mobile apps; AnyPresence combines app building, back-end services, and an API gateway. AnyPresence has an online designer that generates back-end and mobile app code, as well as customized mobile API code. All of the generated code can be downloaded, edited, and run on compatible platforms. To cite one of AnyPresence's favorite customer examples, MasterCard has used AnyPresence to enable partners to easily build mobile apps against MasterCard's Open API services.

AnyPresence generates app UIs (or starter kits, if you wish) for jQuery, Android (XML layout), and iOS (storyboard), and it generates app SDKs for Java, Android, HTML5, Windows Phone, Xamarin, and iOS. The design environment refers to the generated JavaScript/HTML5 SDK as "jQuery."

## Deep Dive

In fact, what AnyPresence actually generates is CoffeeScript that uses the Underscore, Backbone, and jQuery libraries.

AnyPresence generates back-end servers for Ruby on Rails and Node.js. It can generate deployments to Heroku (usually for back ends), to Amazon S3 (usually for HTML5 apps), and to native iOS and Android apps with or without Apperian security. The generated code can always be downloaded and deployed elsewhere. AnyPresence was clearly the best of numerous [mobile back-end platforms I reviewed](#) in the past year, with a good selection of enterprise data source integrations and a nice hosted build environment for app UIs, app SDKs, and back-end servers. It's especially valuable to companies that want to expose their APIs to partners. – MARTIN HELLER

## Red Hat OpenShift

Red Hat's [OpenShift](#) PaaS was designed to provide rapid self-service deployment of common languages, databases, Web frameworks, and applications. One of its current differentiators is that continuous integration (using Jenkins) is a standard part of its workflow; another is that it automatically scales

applications across nodes. It's useful for application development, devops, testing, and production deployment, and it can run in a public or private cloud or on-premise.

Origin — the bleeding-edge, community-supported, free open source version — has daily updates and runs on your hardware using Fedora as the underlying operating system. It's

not really intended for production environments, but can provide a good, fast, free development environment that runs on a laptop or desktop. Red Hat offers OpenShift Online in the Amazon cloud by subscription, updated on a monthly basis based on a snapshot of OpenShift Origin, plus bug fixes. OpenShift Enterprise is based on stable cuts from OpenShift Origin taken two or three times a year.

Some of OpenShift's attractions include automatic application scaling, Git integration at the source code level (with automatic deployment triggered by a git push), and Gear idling, which allows OpenShift to support a very high density of applications compared to, say, Cloud Foundry. – MARTIN HELLER

**AnyPresence was clearly the best of numerous mobile back-end platforms I reviewed in the past year.**



## JetBrains WebStorm

JetBrains' [WebStorm](#) is a modestly priced IDE for HTML, CSS, JavaScript, and XML, with support for projects and version control systems including GitHub. WebStorm is more than an editor, though it's a very good editor. It can check your code and give you an object-oriented view of your project.

Code inspections built into WebStorm cover many common JavaScript issues as well as issues in Dart, EJS, HTML, Internationalization, LESS, SASS, XML, XPath, and XSLT. WebStorm supports code checkers JSHint, JSLint, ESLint, and JSCS.

In addition to debugging Node.js applications as well as tracing and profiling with Spy-js, WebStorm can debug JavaScript code running in Mozilla Firefox or Google Chrome. It gives you breakpoints in HTML and JavaScript files, and it lets you customize breakpoint properties.

When debugging, a feature called LiveEdit allows you to change your code and have the changes immediately propagate into the browser where you are running your debug session. This saves time and helps avoid the common problem of trying to figure out why your



Red Hat OpenShift

## Deep Dive

change didn't do anything, only to discover that you forgot to refresh your browser.

For unit testing, WebStorm bundles the JsTestDriver plug-in. This was originally a Google project, but JetBrains is now contributing to it. In addition, WebStorm can integrate with the Karma test runner. For either testing method, WebStorm tracks code coverage. – MARTIN HELLER

## JetBrains IntelliJ IDEA

JetBrains' [IntelliJ IDEA](#) is a Java IDE available both in an open source Community edition and in a paid-for Ultimate edition. What makes IntelliJ IDEA so compelling is its many innovative development accelerators that speed the process of getting code out of your head and into the computer. For example, its multicursor capability relieves you from having to repeatedly enter the same text at multiple locations: Set a cursor at each spot the text must be added, and type the text once — it appears simultaneously everywhere you specified. Other productivity enhancements include the find action — type open, and IntelliJ will find all operations in the IDE that pertain to the action of opening something (it's even clever enough to recognize that you might have meant "importing").

Granted, we wish the Community edition were equipped with the sorts of J2EE development tools found only in the Ultimate edition: database tools, support for frameworks such as JPA and Hibernate, deployment tools for application servers like JBoss AS, WildFly, and Tomcat. Nevertheless, the Community edition makes a fine Java application development platform that also gives you Android tools, as well as support for other JVM languages like Groovy, Clojure, and Scala (the last two via free plug-ins). Whichever version of IntelliJ IDEA you use, you'll find a rich array of tools designed to simplify otherwise tedious development chores. – RICK GREHAN



## Microsoft SQL Server 2014

Microsoft [SQL Server 2014](#) was the most significant SQL Server release since 2008, and it carried two main themes: cloud and speed. For me, the high note is definitely speed, specifically OLTP performance, which Microsoft has addressed with a number of new features. In-memory tables, delayed durability, buffer pool extension, and updateable columnstore indexes are at the top of this list.

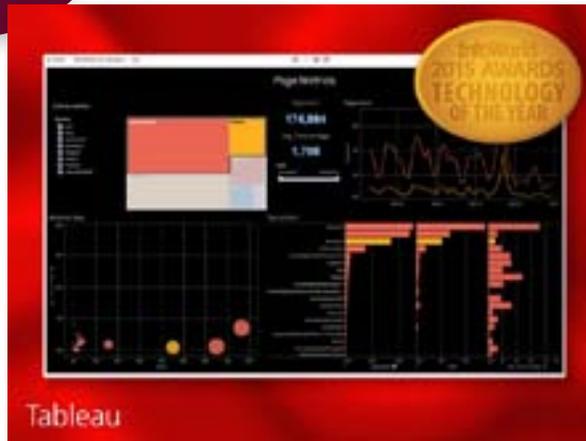
In-memory tables offer a turbo boost through a combination of optimized algorithms, optimistic concurrency, eliminating physical locks and latches, and of course storing the table in memory. This feature is brand new and still quite limited (do your homework first), but as the limitations are removed it will start reaching a much wider audience.

Updateable columnstore indexes have matured since they were added in 2012 and now that they're updateable (i.e., you no longer have to drop and recreate them), they will be much more usable to the general public. Resource Governor finally gets physical I/O control, where you can limit the amount of I/O per volume for a process. This will keep those I/O hogs from taking over your system.

Last and definitely least in my book are the Azure enhancements. You can now back up your database to Azure blob storage. You can even use Azure blob storage to store the data and log files of an on-premises database. While each feature may have its place, I think they come with more caveats than benefits. I don't believe that housing your database files across the Internet will be an advantage to many shops. – SEAN MCCOWN



## Deep Dive



Tableau

## Tableau

**T**ableau is on the cutting edge of the latest generation of business intelligence and reporting tools. The product differentiates itself from its antecedents with a strong focus on easy-to-use visual analytics and powerful features for exploratory data analysis. Tableau is available as a traditional desktop application, a Web-based tool, and a cloud offering. Users of the desktop app can create workbooks of charts and analyses that can be published to Tableau Server for browser-based access by any authorized user.

Tableau Software was founded in 2003, as a spin-off from Stanford University, where researchers had been working on new techniques for visualizing and exploring data. Dating back to the Stanford days, the goal behind Tableau was to combine structured queries and graphical visualization into an easy-to-use combination, allowing nonprogrammers to interactively explore data contained in relational databases, data warehouse cubes, spreadsheets, and other data stores.

Today, Tableau has joined the big data world by integrating data stored in Hadoop and other NoSQL stores. Recent Tableau releases include Mac support and features such as “story points” for linking data into a narrative structure, updated mapping tools, and a “visual data window” for visually defining joins between multiple tables.

– PHIL RHODES

**The product differentiates itself from its antecedents with a strong focus on easy-to-use visual analytics and powerful features for exploratory data analysis.**



Neo4j

## Neo4j

**A** graph database is an excellent place to stash data that involves relationships among its elements: Social networks, shipping routes, and family trees are a few examples that spring to mind. Neo Technology's [Neo4j](#) is a graph database that's simultaneously short on learning curve and long on scalability. Neo4j can be run either as an embedded Java library or in client/server fashion, which means it can tackle tasks small to large (a Neo4j database can house up to 34 billion nodes and an equal number of relationships).

Even better, Neo4j is ACID-compliant and provides two-phase commit transaction support. Its API, which is accessible from a variety of popular languages, includes a number of important “shortest path” algorithms, which you can modify with cost-evaluation functions to create “cheapest path” equivalents. If you'd rather not control your Neo4j database from a language API, you can always turn to its declarative graph query language, Cypher, the Neo4j equivalent to SQL.

Neo4j is available in a free open source edition, as well as an enterprise edition that adds clustering, caching, backups, and monitoring capabilities. Want to know more? Take a few minutes to explore Neo Technology's excellent, interactive [online documentation](#), which guides you through graph database concepts, lets you execute Cypher code against a temporary database, and allows you to view the results graphically in real time.

– RICK GREHAN

**Neo Technology's Neo4j is a graph database that's simultaneously short on learning curve and long on scalability**

## Deep Dive



Apache Cassandra

## Apache Cassandra

**C**assandra is a distributed database that began life as sort of a combination of Google's Bigtable and Amazon's Dynamo, but has since evolved to incorporate architectural elements from both key-value and column-oriented data stores. Cassandra clusters can grow to more than 1,000 nodes, simultaneously providing high throughput and significant data protection.

Cassandra claims near-linear scaling of read and write operations as nodes are added to a cluster. Work and responsibility are equally distributed among the nodes: an I/O request can be sent to any cluster member, and there is no single point of failure. While Cassandra supports eventual consistency, consistency of both read and write operations can be tuned. Recent releases of Cassandra have added row-level isolation on write operations, thus providing full write consistency on a per-row basis.

Best of all, CQL — the Cassandra Query Language — is a route for developers

familiar with relational database systems to begin work with Cassandra. CQL was deliberately fashioned after SQL, and CQL's designers have done a fine job of incorporating Cassandra's unique capabilities into the language (such as expressing Cassandra's "lightweight transaction" mechanism as a CQL element) without making CQL too obscure for new developers. In fact, CQL is becoming the primary programming interface for Cassandra and even supports prepared statements — and their accompanying performance benefits.

— RICK GREHAN

**Best of all, CQL - the Cassandra Query Language - is a route for developers familiar with relational database systems to begin work with Cassandra.**



Apache Hadoop

## Apache Hadoop

**H**adoop, the primogenitor of big data tools, has a plethora of newer, younger competitors nipping at its heels these days. But the Hadoop developers are not resting on their heels, and the technology continues to reinvent itself to remain at the forefront of innovation in the data management and analytics space.

The biggest development was the modularization effort that decoupled the MapReduce API from the underlying scheduling and resource management facilities, giving the world YARN. Using YARN, a Hadoop cluster can serve as a general-purpose compute cluster, capable of hosting jobs using a wide range of compute models. YARN can host graph processing jobs using Apache Giraph, Bulk Synchronous Parallel computations using Apache Hama, and Apache Spark jobs. There is even an implementation of MPI for YARN.

But YARN is not the only news from the Hadoop camp. Over the past few releases, Hadoop has added improvements and new features related to security, encryption, rolling upgrades, Docker support, tiered storage in HDFS, improved REST APIs, POSIX extended file system attributes in HDFS, improved Kerberos integration, scheduler improvements, and much more. While Hadoop may be thought of as the graybeard of big data platforms, it remains a spry, nimble, fast-moving project with plenty of youthful vigor and energy. Hadoop competitors are finding that catching up to — or passing — Hadoop is no easy challenge.

— PHIL RHODES

**Hadoop competitors are finding that catching up to - or passing - Hadoop is no easy challenge.**

## Deep Dive

## Splice Machine

So far, the big data era has been characterized by two hard truths. On the one hand, nothing beats an RDBMS for real-time operations and analytics, but you can forget about scaling. On the other hand, nothing beats Hadoop for scaling, but you can forget about real-time anything. If only you could have your RDBMS and horizontal scalability too. Well, guess what?

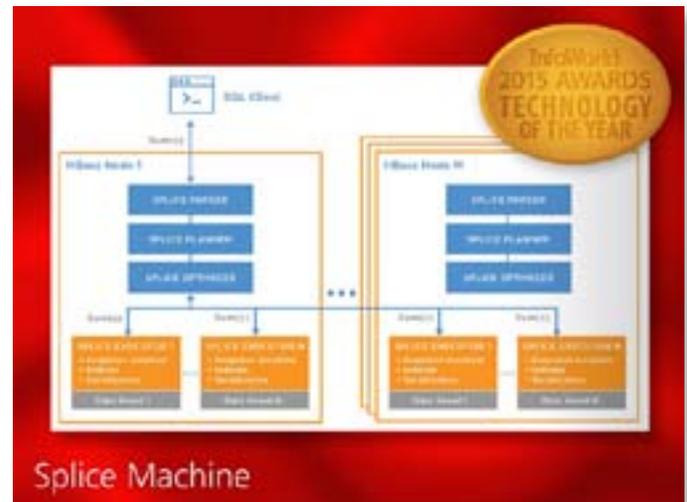
[Splice Machine](#) disrupts the big data status quo by exposing a transactional SQL query layer atop Hadoop's big data stack, essentially blending the best of SQL with the scale-out

capability of NoSQL. This nirvana is achieved by replacing the storage

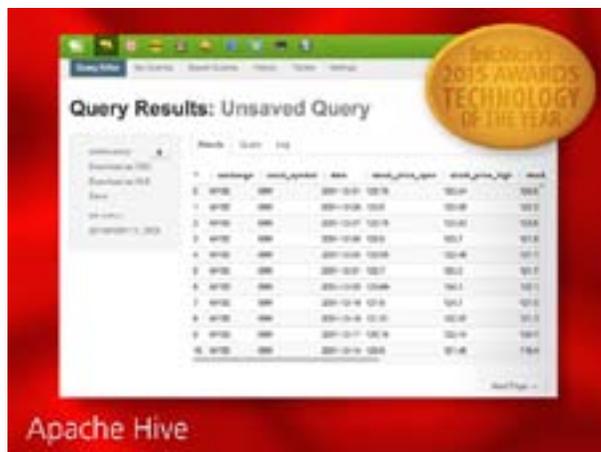
layer inside Apache Derby with HBase and tweaking Derby's planner, optimizer, and executor components to function within the distributed architecture. Splice Machine compiles SQL byte code, pushes it to the HBase nodes, and manages the parallel processing to splice results back together in the end.

Splice Machine combines the ability to create real-time, ACID-compliant OLTP applications with the capacity to analyze and process massive amounts of data in real time. Best of all, the skill set requirements boil down to SQL and HDFS — no need to climb the learning curves for MapReduce and Java programming. It's not open source or cheap, but Splice Machine unlocks the future potential for running concurrent transactional and operational workloads on Hadoop. — JAMES R. BORCK

**Splice Machine disrupts the big data status quo by exposing a transactional SQL query layer atop Hadoop's big data stack.**



Splice Machine



Apache Hive

## Apache Hive

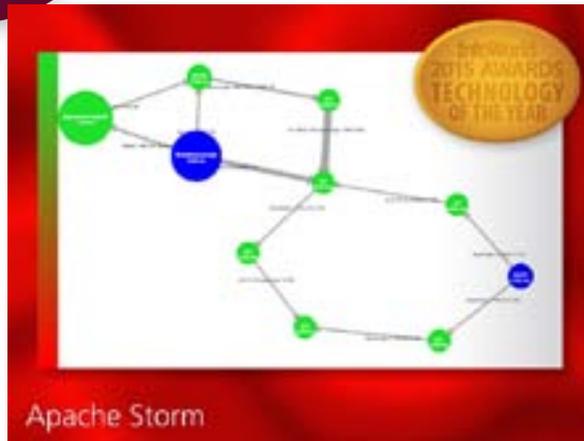
Traditional data warehouses are under pressure. Companies now want to analyze unprecedented amounts of data, and they aren't willing to wait six months for the answer. [Apache Hive](#) is a good place to store EDW (enterprise data warehouse) data that isn't frequently used, relieving some of the stress on the data warehouse. It is also a great ETL (extract, transform, load) tool. By landing data in Hadoop and processing it in Hive before loading into the EDW, processing loads are lower too. By keeping less frequently used data in a "warm" location like Hive, it is still accessible to auditors, regulators, and data

scientists, continuing to add value to the enterprise, but at lower storage and processing costs.

Though typically deployed to augment the data warehouse, the newest Hive, 0.14, boasts impressive capabilities itself. Traditionally a write-once, read-often system, updates and inserts were always problematic in Hive. With version 0.14, Hive added SQL transactions for insert, update, and delete, with ACID semantics, bringing Hive much closer in capability to traditional EDWs. The Apache community is rapidly closing the gap too: Full SQL 2011 semantics and subsecond queries are coming up next. The day may not be too far away when Hive is your sole EDW solution.

— STEVEN NUNEZ

## Deep Dive



Apache Storm

## Apache Storm

**S**torm is one of a new breed of real-time stream computing platforms that have emerged over the past year or two. The project began life with a company called BackType, which was eventually acquired by Twitter. After the acquisition, Twitter open-sourced the project, and it had an immediate impact on the big data landscape. Developers looking to perform real-time, incremental computations over streams of data immediately jumped on the Storm bandwagon, and the project has done nothing but pick up steam since.

After being hosted on GitHub and maintained largely by Twitter, Storm was eventually transitioned to the Apache Software Foundation incubator. Storm graduated the incubator and became an Apache Top Level Project (TLP) in September 2014. The latest release, 0.9.3, was delivered in November.

Storm provides scalability for performing distributed computation on streaming data. Components (spouts and bolts) are assembled into “topologies” that run forever by default, with the individual component instances running on multiple nodes within the cluster. The sequencing of computational steps is defined as a Directed Acyclic Graph where messages (tuples of fields) are passed from component to component according to the execution graph. This architecture facilitates easy recovery from failure and permits Storm clusters to scale to support massive streams of data.

– PHIL RHODES

**Developers looking to perform real-time, incremental computations over streams of data immediately jumped on the Storm bandwagon, and the project has done nothing but pick up steam since.**



Apache Spark

## Apache Spark

**T**he current darling of the Hadoop world, [Apache Spark](#) provides a suite of applications bound together by a common data structure. Designed by the AMPLab at UC Berkeley to solve machine learning problems on a Hadoop cluster, Spark takes a strictly functional approach to programming. Spark’s design allows you to easily share code and predictive models among the stack components by passing the underlying data structure, the RDD (resilient distributed dataset). Written in Scala, Spark also offers language bindings for Python and Java, and third-party DSLs are available for Clojure and Groovy.

Spark has four components in its ecosystem: SparkSQL, MLlib, Spark-Streaming, and GraphX. MLlib is Spark’s sweet spot, greatly increasing the speed with which machine learning models can be built. Compared to some of the Hadoop machine learning veterans like Mahout, Spark is hands-down the winner in almost every respect. With a large, vibrant community that is actively adding new algorithms on a three-month release cycle, it won’t be long before Spark catches up in functionality to Python and R, for machine learning, but able to do it at scale. – STEVEN NUNEZ

**Spark has four components in its ecosystem: SparkSQL, MLlib, Spark-Streaming, and GraphX. MLlib is Spark’s sweet spot, greatly increasing the speed with which machine learning models can be built.**

## Deep Dive



## RStudio

Open source statistical language R has been the domain of science labs and quantitative finance geeks since the late 1990s. Today, R's use by mainstream companies to parse line-of-business data has made it one of the fastest-growing languages (as measured by the Tiobe Index).

**RStudio offers an accessible, graphical interface that tames R's quirky syntax with features such as code completion, debugging, charting, and report development.**

Backing the surge is Boston-based RStudio and its eponymous R development environment. In contrast to R's bare-bones, command-line console, RStudio offers an accessible, graphical interface that tames R's quirky syntax with features such as code completion, debugging, charting, and report development.

With RStudio you can load and clean data, run a variety of statistical functions across the set, and output tables with insightful graphics using a few keystrokes. RStudio even helps manage R's extensive library of pluggable packages — from ODBC connectors and data massaging tools to statistics routines and time-series forecasting. Further, RStudio's Shiny framework provides a toolkit for building interactive, data-driven Web apps for browser-based reporting.

RStudio is available in a free open source edition and in a professional edition with additional security and resource management features. Either way, RStudio is a must-have for productivity with R.

— JAMES R. BORCK



## HP Moonshot

Hewlett-Packard has delivered a number of unique hardware innovations with the introduction of [HP Moonshot](#). A next-gen, high-density, energy-saving, cartridge server system, Moonshot drives up the efficiency of specific workloads — such as cloud computing, big data processing, and virtual desktop infrastructure — by carving processing resources into more discrete and specialized units.

At launch, the Moonshot system came fully loaded with 45 individual processor cartridges running Intel Atom processors. The latest crop of server cartridges includes the much-anticipated HP ProLiant m400, featuring a 64-bit ARM processor and 64GB of memory. Other new cartridges include the HP ProLiant m700, with four AMD Opteron X2150 APUs and 32GB of ECC RAM, and the HP ProLiant m710, with an Intel E3-1282Lv3 processor and integrated Intel Iris Pro Graphics.

Connecting 45 processor slots together along with two additional long slots for the network switches required a significant engineering effort. Four separate interconnect fabrics provide physical pathways for different types of communication within the Moonshot chassis. Utilizing one of the underlying data pathways, the 2D Torus fabric, allows multiple ProLiant m800 cartridges to be clustered together for high-bandwidth, low-latency data throughput to support complex analysis of large volumes of data. Networking options include 1G and 10G

switches for both internal and external connections.

— PAUL FERRILL

**Moonshot drives up the efficiency of specific workloads — such as cloud computing, big data processing, and virtual desktop infrastructure.**

## Deep Dive

## Dell PowerEdge R730xd

Dell is no stranger to new and innovative server designs, and the company did it again this year with the [Dell PowerEdge R730xd](#). Part of Dell's 13th Generation lineup, this 2U server brings an extreme amount of flexibility and versatility to an industry-standard system. Highlights include the new disk housing that holds up to 18 1.8-inch SSDs while still leaving room for 10 2.5-inch SAS disks. That doesn't count the two disks on the rear of the box for loading an operating system or the dual internal SD slots designed to serve as a redundant boot device for VMware ESXi. Both SD slots can be mirrored for redundancy, giving the system an extra measure of protection against failure. The latest Dell PERC disk controllers support JBODs for use with Microsoft Storage Spaces or VMware VSAN.

The PowerEdge R730xd excels in every category, but especially in performance and scalability. With 18-core Intel Xeon CPUs and new DDR4 memory parts, a single server can pack 72 processing threads and 1.5TB of memory. That's an incredible amount of horsepower in a 2U box. — PAUL FERRILL



Dell PowerEdge R730xd

## VMware Virtual SAN

VMware's [Virtual SAN](#) brings the control and management of locally attached storage into the hypervisor kernel, making it possible to use the combination of SSD and traditional disk in vSphere cluster nodes as resilient, high-performance shared storage. VSAN achieves high availability by replicating storage objects (virtual machine disks, snapshot images, VM swap disks) across the cluster, allowing admins to specify the number of failures (nodes, drives, or network) to be tolerated on a per-VM basis. At the same time, the solution addresses IO latency by leveraging flash-based storage devices for write buffering and read caching, along with support for 10GbE network connectivity.

**The net result – as we discovered in our testing – is a level of performance you would expect from a high-end, dedicated storage system, yet significantly lower overall storage costs.**

The net result — as we discovered in our testing — is a level of performance you would expect from a high-end, dedicated storage system, yet significantly lower overall storage costs. Plus, the vSphere Web Client contains everything you need to monitor and manage the system from the convenience of a Web browser. — PAUL FERRILL



VMware Virtual SAN

## Deep Dive

## Tintri VMstore

Leveraging MLC flash for high-speed IO and low-cost disk for capacity, the [Tintri VMstore](#) was designed from the ground up to store virtual machines. Tintri tosses old-fashioned terms like “LUN” and “volume” to the wayside and puts the VM at the center of storage management. Admins get an unencumbered view of VM performance, and storage management tasks operate directly on the level of virtual machines and virtual disks.

The Tintri OS inspects all network traffic into and out of each storage appliance, leveraging deduplication, compression, and other techniques to keep active data in flash, and providing a deep level of understanding about each virtual machine and its performance. This information is used to provide QoS down to the individual VM, making it possible to run mixed workloads on the same data store while delivering the performance required to each one.

Tintri initially supported only VMware, but has expanded coverage with the latest release to include Microsoft’s Hyper-V and Red Hat Enterprise Virtualization. The VMstore implements a number of features included in these virtualization management systems – such as cloning, snapshots, and replication – but brings to bear all of the flash-specific speed and optimization advantages of Tintri OS. This is not your father’s general-purpose storage array. – PAUL FERRILL

**Tintri tosses old-fashioned terms like “LUN” and “volume” to the wayside and puts the VM at the center of storage management.**



## ioSafe 214

A fireproof and waterproof NAS... what more could you want for your precious files? Fast and full-featured, with cloud backup, iSCSI support, a Web file/calendar server, folder encryption, and the ability to run a wide range of pluggable apps including Asterisk, Drupal, DNS, and Git, the [ioSafe 214](#) is a great storage system. In fact, it’s a Synology DiskStation under the hood. But it’s also a hedge against disaster, protected by an ablative ceramic housing that can survive heat up to 1,550 degrees Fahrenheit for 30 minutes or a dunking in 10 feet of salty or fresh water for 72 hours.

The icing on the cake is ioSafe’s policy of providing Drive Saver recovery services as part of its extended warranty. All together, it’s a great way to provide essential storage services to a small business or workgroup and a superb solution for business continuity. – BRIAN CHEE

## Deep Dive



Lancope StealthWatch

## Lancope StealthWatch

One of the best methods for detecting hackers who have pwnd your IT environment is network traffic flow analysis, aka netflow. Most servers don't talk to other servers, and most workstations never talk to other workstations. If you understand your legitimate network traffic flows, you can more quickly identify malicious hackers as they explore your network in atypical ways.

To do this you need a good netflow analysis tool. The best netflow product we've come across is Lancope's [StealthWatch](#).

StealthWatch works by monitoring traffic flows on routers, switches, and other network devices, collecting network flow statistics via sFlow, NetFlow, and other standard methods, then proactively identifying known bad traffic. It then allows you to quickly create baselines and alerts.

StealthWatch monitors each endpoint device using 90 different attributes (who it's communicating with, how long, how much data is being sent or downloaded) and quickly allows you to create logical groups (workstations, Web servers, SQL servers). Each group is assigned a "concern index," which is basically a criticality ranking. Yellow and red rankings are the ones you'll want to explore first. StealthWatch provides excellent graphical representations and makes it easy to find and sort detailed data. It's a netflow analyst's dream tool and one of our top products of the year. — ROGER A. GRIMES



Fluke LinkSprinter

## Fluke LinkSprinter

Simply because the link light comes on doesn't mean the network device you installed is ready for prime time. Only through testing and validation can you trust that all is well as you go into production. Unfortunately, handheld network-testing tools have been too expensive, too slow, and too cumbersome to carry as part of your regular toolkit — until now. Fluke Networks' [LinkSprinter](#) makes validation test equipment so affordable, there's no excuse not to hang a piece of test gear on every technician.

The LinkSprinter 100 (wired Ethernet) and LinkSprinter 200 (Wi-Fi) provide quick and dirty test indicators in the form of tri-state LEDs, but thanks to a built-in Web server, your browser is now the place to get more detailed test information. Best of all, the companion LinkSprinter Cloud Service provides a common data collection point for any number of LinkSprinters, with filtering tools to make it easy to create PDF or CSV

reports on any subset of your testing data. Installers can even include the data jack number combined with data from LLDP or CDP switches to give you a fairly decent start on your network inventory. ■

— BRIAN CHEE

**Fluke Networks' LinkSprinter makes validation test equipment so affordable, there's no excuse not to hang a piece of test gear on every technician.**