

The Official Conference, Trainings, and Expo of The Apache Software Foundation

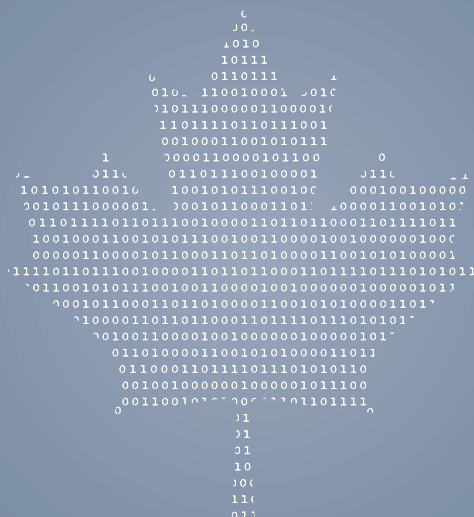
Apache Con North America

2011

November 7-11 Vancouver, BC

Conference Program Guide

[Updated 20111104]



Presented by: The Apache Software Foundation
Produced by: Stone Circle Productions, LLC

SPONSORED BY



Platinum



Gold



FuseSource
A Progress Software Company

spring source
A division of VMware

Silver

Open Source Enterprise Solutions, Cloud Computing
and Community Leadership

<http://www.apachecon.com/>

Thank You to Our Sponsors

Platinum Sponsor

cloudera

Gold Sponsors



Silver Sponsors



Bronze Sponsors



Community Sponsors and Exhibitors



Media and Community Partners



Welcome to ApacheCon

Hello and Welcome to ApacheCon!

As with past ApacheCons, this year's conference promises to be a great event, brings the best minds in Open Source together with an impressive attendee roster from across the globe.



WELCOME

We'll be kicking off the week with pre-conference Trainings, conference proper Wednesday through Friday. You will be able to pick up your delegate badge at the Main Foyer (Lobby Level, Westin Bayshore) from 9AM Monday and Tuesday, from 7.30AM Wednesday, and 8AM Thursday and Friday.

Whether you're a rookie or a rocket scientist, the week is an excellent opportunity to connect with Apache enthusiasts, share, network, collaborate, and discuss real-world issues, from big data to community development to keeping the machine running.

With a conference program packed with presentations that deliver the most up-to-date information, our trainers and presenters will share relevant strategies, experiences, techniques, trends, and best practices --both on stage and throughout the conference. Keynotes from David A. Wheeler, Eric Baldeschweiler, and David Boloker highlight the evolving benefits, challenges, and opportunities with next generation Apache innovations.

Evening events start on Tuesday with Apache MeetUps and roll through the week with the Welcome Reception, Expo kickoff, PGP keysigning, more MeetUps, and Lightning Talks, where you'll hear something fun, off-the-cuff, and, at times, controversial. Those of you who simply can't get enough are invited to participate in the post-con BarCampApache, coordinated with members of the local Open Source community.

Very special thanks to all our sponsors, exhibitors, and partners for their support for both the conference and community-at-large - we are most grateful for your generosity. Many thanks also to the ApacheCon Planners, Stone Circle Productions, and the countless volunteers for their efforts in making our event a success.

We wish you a fun, productive stay in Vancouver, and hope you enjoy all that ApacheCon has to offer.

--Sally Khudairi, Conference Chair

Table of Contents

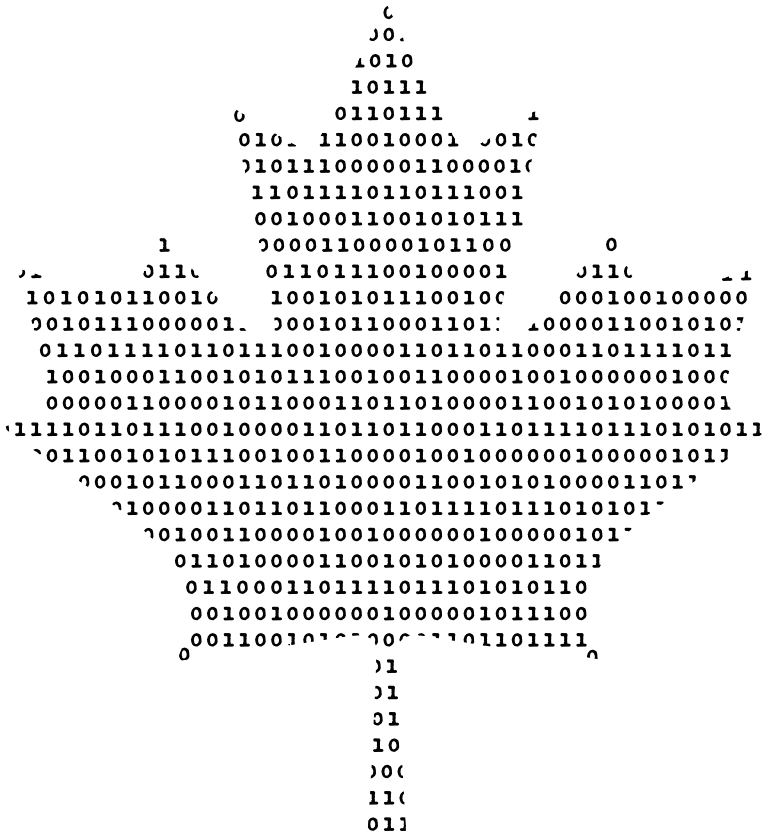
- Sponsors & Exhibitors 4
- Community & Media Partners 16
- Keynotes 23
- Special Events & Meetups 24
- Schedule-at-a-Glance 25
- Sessions A-Z 29
- Speakers 59



Cloudera - Platinum Sponsor

Cloudera is the leading provider of Apache Hadoop-based software and services and works with customers in financial services, web, telecommunications, government and other industries. The company's products, Cloudera Enterprise and Cloudera's Distribution including Apache Hadoop, help organizations profit from all of their information. Cloudera's Distribution including Apache Hadoop is the most comprehensive Apache Hadoop-based platform in the industry. Cloudera Enterprise is the most cost-effective way to perform large-scale data storage and analysis and includes the tools, platform and support necessary to use Hadoop in a production environment.

For more information visit
 » <http://www.cloudera.com>





Hewlett-Packard - Gold Sponsor

HP Open Source and Linux solutions are designed to solve your mission-critical IT problems today! We sell over a million Linux servers a year, delivering high performance, scalability, security, storage, and backed by our commitment to the open source community. Learn more about our market leading Linux platforms and Open Source Middleware Stacks at www.hp.com/go/linux. Also be sure to check out our new FossBazaar and Fossology initiatives which are aimed at further encouraging the adoption of Open Source in the enterprise. Our solutions are delivered and supported by over 6500 Open Source and Linux professionals across 160 countries.
<http://hp.com/go/opensource>



Hortonworks - Gold Sponsor + Exhibitor

Hortonworks is focused on accelerating the development and adoption of Apache Hadoop. Together with the Apache community, we are making Hadoop more robust and easier to install, manage and use. We also provide support and training for Apache Hadoop.
<http://www.hortonworks.com>



HotWax Media - Gold Sponsor

As the global leader in Apache Open For Business (OFBiz), HotWax Media provides web-based enterprise resource planning (ERP), e-commerce, and business automation consulting services. From enterprise-level e-commerce to warehouse, manufacturing, and financial management, HotWax Media and OFBiz help optimize business processes. Open source is the smart choice today for businesses ranging from small-and medium-sized organizations to global Fortune 100 companies. Using Apache OFBiz, HotWax Media delivers top quality enterprise software for a fraction of the price of outdated, proprietary vendors. For more information visit <http://www.hotwaxmedia.com/index.html>

**Adobe Systems – Silver Sponsor**

We help our customers create, deliver, and optimize compelling content and applications – improving the impact of their communications, enhancing their brands and productivity, and ultimately bringing them greater business success. Together, we're turning ordinary interactions into more valuable digital experiences every day, across media and devices, anywhere, anytime.

<http://www.adobe.com/solutions/customer-experience/web-experience-management.html>

**FuseSource – Silver Sponsor + Exhibitor**

FuseSource is the leader in open source integration and messaging, and offers tested, integrated and certified distributions of Apache projects. The FuseSource products are fully supported by the people who build and influence the projects to combine the speed and innovation of open source software with the reliability and expertise of commercially provided enterprise services.

<http://fusesource.com/>

**SpringSource – Silver Sponsor + Exhibitor**

VMware delivers virtualization and cloud infrastructure solutions that enable IT organizations to energize businesses of all sizes. With the industry leading virtualization platform – VMware vSphere™ – customers rely on VMware to reduce capital and operating expenses, improve agility, ensure business continuity, strengthen security and go green. With 2009 revenues of \$2 billion, more than 190,000 customers and 25,000 partners, VMware is the leader in virtualization which consistently ranks as a top priority among CIOs. VMware is headquartered in Silicon Valley with offices throughout the world and can be found online at www.vmware.com.

<http://www.springsource.com/>

Facebook – Bronze Sponsor

Founded in February 2004, Facebook is a social utility that helps over 500 million people communicate more efficiently with their friends, family and coworkers.

Over the past few years, the company has developed a number of open source infrastructure technologies to support the site's growth including HipHop for PHP, Apache Cassandra, Apache Hive, and Apache Thrift as well contributed to numerous open source projects including Hadoop. Facebook engineers actively participate in various Apache projects, developing technologies that facilitate the sharing of information through the social graph.

<https://developers.facebook.com/opensource/>

**Lucid Imagination – Bronze Sponsor + Exhibitor**

Lucid Imagination is the commercial company for Apache Lucene/Solr enterprise search technology. We provide the planet's best search solution development platforms built on the power of Solr/Lucene open source search. LucidWorks Enterprise makes the power of Solr/Lucene open source search more accessible to the broad range of application developers and slashes the learning curve for search solution development. Unlike "black box" products, LucidWorks Enterprise allows organizations of all sizes and types to continuously tune their search to fit the ongoing needs of their users and achieve a consistently lower cost of growth.

Lucid Imagination also offers free software for developers, documentation, commercial-grade support, high-level consulting, and comprehensive training. Customers include AT&T, Sears, Ford, Verizon, Cisco, Zappos, Raytheon, The Guardian, The Smithsonian Institution, Salesforce.com, The MotleyFool, Macy's, Qualcomm, Taser, eHarmony, and many other household names around the world. Lucid Imagination is a privately held venture-funded company. The investors include Shasta Ventures, Granite Ventures, Walden International, and In-Q-Tel. The Lucid Imagination founding team consists of several key contributors and committers to the Lucene project, as well as experts in enterprise search software development.

<http://www.lucidimagination.com/>



Acunu - Community Sponsor

Acunu are world experts in Big Data applications and offer products and services for Apache Cassandra and beyond. Acunu has reinvented the storage stack for Big Data workloads and brings fundamental innovation to Apache Cassandra. The Acunu Data Platform supercharges and simplifies Apache Cassandra deployments, making them business-ready at a fraction of the cost -- it deploys out-of-the box, offers dramatically improved performance and reliability with commodity hardware or cloud based deployments, and comes with management and monitoring tools to meet enterprise operational requirements. Acunu is backed by leading European investors and has a unique mix of researchers and engineers from Cambridge, Oxford and Stanford Universities. Acunu has offices in San Francisco, San Antonio and London. Find out more.

<http://www.acunu.com>



The Apache Software Foundation - Community Sponsor + Exhibitor

Established in 1999, the all-volunteer Foundation oversees nearly one hundred fifty leading Open Source projects, including Apache HTTP Server — the world's most popular Web server software. Through the ASF's meritocratic process known as "The Apache Way," more than 300 individual Members and 2,500 Committers successfully collaborate to develop freely available enterprise-grade software, benefiting millions of users worldwide: thousands of software solutions are distributed under the Apache License, and the community actively participates in ASF mailing lists, mentoring initiatives, and ApacheCon, the Foundation's official user conference, trainings, and expo. The ASF is a US 501(3)(c) not-for-profit charity, funded by individual donations and corporate sponsors including AMD, Basis Technology, Cloudera, Facebook, Google, IBM, HP, Matt Mullenweg, Microsoft, SpringSource, Talend, and Yahoo!. For more information, visit

<http://www.apache.org/>

Acunu Data Platform

The Big Data Platform powered by Apache Cassandra



Find out more and get your **free copy of the
Cassandra High Performance Cookbook**
at the Official Apache Cassandra MeetUp on
Thursday (8pm, 10th Nov)

Acunu are proud sponsors of
the Official Apache Cassandra MeetUp

Find out more about Acunu and the
Acunu Data Platform at www.acunu.com

Apache Cassandra is a trademark of the Apache Software Foundation





AppDynamic - Exhibitor

AppDynamics is the leading provider of application performance management for modern application architectures in both the cloud and the data center, delivering solutions for highly distributed, dynamic, and agile environments. Companies such as Priceline, TiVo, and ZipRealty use AppDynamics to monitor, troubleshoot, diagnose, and scale production applications, and over 30,000 people have downloaded our Lite product. AppDynamics offers both Dev and Ops teams a powerful solution to reduce cost and ensure the performance and availability of their revenue-critical applications. <http://www.appdynamics.com>



Bazaarvoice - Community Sponsor

Bazaarvoice is a Software as a Service (SaaS) company that turns social media into social commerce for the world's best global brands by enabling authentic customer-powered marketing. We achieve that by first letting consumers create and share ratings, reviews, questions, answers, and stories about products and brands on client websites. We then syndicate that content across the Web and through the channel to reach shoppers wherever they are. This customer-generated content has a tremendously positive influence on purchase decisions. It is also rich with customer intelligence, which we analyze and provide to our clients so they can make real-time, information-based decisions that affect all aspects of their business - marketing, product R&D, and operations. This strategy of putting customers at the center of your business helps you achieve measurable results across your entire enterprise. Founded in 2005 and headquartered in Austin, Bazaarvoice has offices in Amsterdam, London, Munich, Paris, Stockholm, and Sydney. <http://www.bazaarvoice.com/>

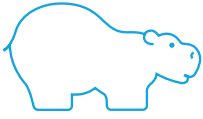


CrowdVine - Community Sponsor

CrowdVine, founded in 2007, builds simple and powerful social software to help people connect. While we believe our products are very good, the killer feature of social software is people. That's why when

you go to our features page, you'll see a list of reasons to use our software, not just a list of tools. Everything we do is aimed at making it easier for people to meet. Once they've done that, it's our job to get out of the way.

<http://www.crowdvine.com/>



HIPPO

Hippo - Community Sponsor

Hippo develops the Apache licensed Hippo CMS and Hippo Portal application. Hippo is an avid believer of the Apache way of developing, and participates pro-actively in ASF Projects. Hippo's developers are stimulated to get involved in Apache in order to become committed in relevant projects, sharing knowledge with others in true Apache spirit. You can find code initiated by Hippo people in many projects, including Apache Cocoon, Apache Jetspeed and Apache Slide. Hippo is a member of the Java Community Process and its developers help define new Java specifications.

Hippo CMS is already widely used in the Netherlands and across Europe. With the acquisition of the assets of Blue Sunrise, another great company active in the Apache community, Hippo obtained a strong threshold into the USA. Hippo is now well on the way to becoming a household name in the US market place, and is rolling out vastly to set the global standard in open source content management and portal applications.

For more information, visit the company website at <http://www.onehippo.com>. or get involved in the Hippo community at <http://www.hippocms.org>.
<http://www.onehippo.com>



Jahia - Community Sponsor

Jahia's next-generation, open source CMS stems from a widely acknowledged vision of enterprise application convergence - web, search, document, social and portal - unified by the simplicity of web content management. Structured around Jahia Extended Content Management (xCM), the Jahia Studio IDE, Jahiapps social and collaboration

capabilities and based on a powerful Composite Content Management platform, Jahia brings a unique modularity to the Java development arena. It is the tool of choice for building innovative user experiences for virtually any type of web application at a fraction of the cost of competing solutions. Founded in 2002 in Switzerland, Jahia has its North American headquarters in Washington DC, and offices in Canada and throughout Europe. Jahia has over 350 customers in more than 20 countries across the globe. For more information, please visit <http://www.jahia.com>

Microsoft® **Microsoft - Community Sponsor**

At Microsoft, we're motivated and inspired every day by how our customers use our software to find creative solutions to business problems, develop breakthrough ideas, and stay connected to what's most important to them. We are committed long term to the mission of helping our customers realize their full potential. Microsoft's open source strategy is grounded in recognition of the value of openness to working with others -- including commercial open source vendors, developers and open source communities -- to help customers and partners succeed in today's heterogeneous "world of choice." <http://www.microsoft.com/en-us/openness/>



Ning - Community Sponsor

Ning is the leading online platform for the world's organizers, activists and influencers to create social experiences that inspire action. Based in Palo Alto, Calif., Ning offers an easy-to-use service that enables people to create custom branded social networks. Ning was co-founded by Chairman Marc Andreessen and is privately held. For more information, visit: <http://www.ning.com>



OpenDirective - Community Sponsor

OpenDirective emerged from Oxford University's OSS Watch to facilitate open innovation in software, and form the link between OSS Watch in the academic

research sector and product development of the commercial sector. OpenDirective is committed to identifying software products and opportunities that are reaching the chasm between the early adopters and the early majority, with a focus on mobile and accessibility-enabling technologies that stem from research or are developed for internal use, and are not necessarily developed as products in themselves. By facilitating open innovation through engagement with Open Source software, OpenDirective helps make companies more innovative, capable, and cost efficient. OpenDirective is led by Ross Gardler and Steve Lee, who, for more than ten years, have actively participated in various Open Source communities that include The Apache Software Foundation, Mozilla, GNOME and Ubuntu.
<http://www.opendirective.com/>



Oracle - Community Sponsor

Oracle provides the world's most complete, open, and integrated business software and hardware systems, with more than 370,000 customers—including 100 of the Fortune 100—representing a variety of sizes and industries in more than 145 countries around the globe. Oracle's product strategy provides flexibility and choice to our customers across their IT infrastructure. Now, with Sun server, storage, operating-system, and virtualization technology, Oracle is the only vendor able to offer a complete technology stack in which every layer is integrated to work together as a single system. In addition, Oracle's open architecture and multiple operating-system options gives our customers unmatched benefits from industry-leading products, including excellent system availability, scalability, energy efficiency, powerful performance, and low total cost of ownership.
<http://www.oracle.com/us/technologies/open-source/index.html>



Talend - Community Sponsor + Exhibitor

Talend is the recognized market leader in open source data management & application integration. Talend provides fully certified and tested application

integration solutions powered by popular Apache projects including Apache CXF, Apache Camel, Apache Karaf and Apache ActiveMQ. As a leading provider of Apache-powered software, Talend believes in making a strong investment in the community. Talend currently employs engineers to work on many Apache projects, including CXF, Camel, Karaf, ActiveMQ as well as Apache Santuario, Apache Hadoop, Apache ACE and others.

Talend revolutionized the world of data integration when it released the first version of Talend Open Studio in 2006. Talend's data management solution portfolio includes data integration, data quality, and master data management. Unlike proprietary, closed solutions, which can only be afforded by the largest and wealthiest organizations, Talend makes middleware solutions available to organizations of all sizes, for all integration needs.
<http://www.talend.com>



WSO2 - Exhibitor

WSO2 delivers the only complete open source middleware platform. With its revolutionary componentized design, WSO2 middleware adapts to the project for a lean, targeted solution to enterprise applications. Fully cloud-native, the WSO2 middleware platform is also the only open source platform-as-a-service for private and public clouds available today. With WSO2, seamless migration and integration between servers, private clouds, and public clouds is now a reality.
<http://www.wso2.com>

Get Up to Speed on Apache Hadoop

Cloudera, the leading provider of Hadoop-based software and services, offers comprehensive training and certification exams.



Upcoming classes:

Cloudera Developer Training & Certification for Apache Hadoop

Dallas	Nov 30–Dec 2
SF Bay Area	Dec 5–8
New York City	Dec 19–22
Los Angeles	Jan 9–12
Raleigh	Jan 9–12
Chicago	Jan 17–20
Columbus, OH	Jan 23–26

Cloudera Administrator Training & Certification for Apache Hadoop

Columbia, MD	Nov 21–23
SF Bay Area	Nov 21–23
Herndon, VA	Jan 4–6
Los Angeles	Jan 18–20
Chicago	Jan 25–27

Cloudera Training for Apache Hive and Pig

SF Bay Area	Dec 12–13
Chicago	Jan 23–24
New York City	Feb 6–7

Cloudera Training for Apache HBase

SF Bay Area	Dec 14–15
New York City	Feb 8–9

For more information and more classes:
<http://www.cloudera.com/info/training>
training@cloudera.com
1-888-789-1488

cloudera



ActiveState - Community + Media Partners

ActiveState is the world leader in development, management, distribution, and cloud solutions for dynamic language applications. The company's products and services for Perl, Python, Tcl and other web languages are used by over 2 million developers and 97% of the Fortune 1000 to build and run applications from mission-critical to open source projects. ActiveState's end-to-end solutions include development tools, language distributions, a cloud platform, indemnification, OEM solutions, support and other services. With these, customers such as Cisco, CA, HP, Bank of America, Siemens, and Lockheed Martin are able to accelerate productivity, minimize risk, and ensure compliance.

<http://www.activestate.com/cloud>



Admin - Community + Media Partners

Explore the new world of system administration! Each issue of ADMIN magazine delivers technical solutions to the real-world problems you face every day. Subscribe now!



The BitSource - Community + Media Partners

The BitSource, Technology's Premier Source of Information, is an online publication and research lab providing in-depth enterprise software and service product reviews, interviews and tutorials. The BitSource brings together best-in-class research, reviews, and most importantly: the people of the technology world.



Conferencevault - Community + Media Partners

Conferencevault is the gateway site for IT professionals and Technology enthusiast dedicated to providing current and relevant information on all Information technology-centric conferences, training, events and opportunities across the globe. From Calls for papers, to conference searches by city, state or country, Conferencevault makes it easy for its users to be involved in technology.



DZone - Community + Media Partners

DZone is a social linking and blogging network for

the developer and IT communities. According to PC Magazine, DZone is a developer's dream—a vast network of user-submitted links to message boards, news, coding tricks, and more.

Launched in June, 2006, DZone is in Alexa's top 3000 sites, surpassing established leaders like DevX, Syscon, FTP Online and TheServerSide.com. DZone is the only vertically focused site regularly listed among the web's largest social bookmarking sites. In its first year of operation DZone sent over 5 million visitors to other developer websites. Today, DZone has curated topic pages for Java, Solr/Lucene, Cloud Computing, PHP, Agile, Mobile, and much more. www.dzone.com



FeatherCast - Community + Media Partners

Created in 2006 by Apache Software Foundation (ASF) Members/co-hosts Rich Bowen and David Reid, FeatherCast highlights the Foundation's various people, projects, and plans. The independently-run series of podcasts draws developers, users, Open Source enthusiasts, and general IT audiences to hear perspectives on activities relevant to the Apache Community. More than 40 leading voices have contributed their insight on popular topics such as ASF-incubated projects, ApacheCon, ActiveMQ, and Google's Summer of Code. With episodes added regularly, catch future FeatherCasts on new ASF Members, intellectual property issues, WS-*, and more at FeatherCast.org and via iTunes.



LinuxFest Northwest - Community + Media Partners

April 28th & 29th 2012 in Bellingham Washington
LinuxFest Northwest is a non-profit technical Expo produced annually during the last weekend of April since 2000. It will be held April 28th & 29th on the campus of Bellingham Technical College and is free of admission cost. LinuxFest Northwest has a community atmosphere, and offers key presentations and exhibits on a wide variety of free and open source technology topics. Over 90 workshops and presentations, ranging from a beginning level to the highly technical, draw tech savvy professionals and entrepreneurs from throughout northwestern United States and British

Columbia. Event speakers include influential and internationally known open source experts.



The Linux Gazette - Community + Media Partners

The Linux Gazette is the Free International Online Linux Monthly, with 42 mirrors in 18 countries and 7 translated versions (Indonesian, Portuguese, Russian, Spanish, and Thai.) Aimed at a readership including everyone from brand-new Linux users to professionals in the field, we have been "Making Linux a Little More Fun" since 1995, and cover technical topics, community news, hot Linux tips, and geek humor. We're also a peer-reviewed journal, suitable for academic publication credit.



Linux Pro Magazine - Community + Media Partners

Linux Pro Magazine keeps the emphasis on real-life, practical techniques and delivers in-depth technical articles on a range of topics, including reviews of new products, practical advice on tools and strategies for sys admins, programming tips, and discussions of advanced desktop techniques. Each issue includes a Linux DVD! Sign up for your \$3 trial subscription now!



Manning Publications - Community + Media Partners

Manning Publication publishes computer books for professionals--programmers, system administrators, designers, architects, managers and others. Our focus is on computing titles at professional levels. We care about the quality of our books. We work with our authors to coax out of them the best writing they can produce. We consult with technical experts on book proposals and manuscripts, and we may use as many as two dozen reviewers in various stages of preparing a manuscript. The abilities of each author are nurtured to encourage him or her to write a first-rate book.



MUCMD - Community + Media Partners

About the Meaningful Use of Complex Medical Data Symposium (MUCMD)

MUCMD will bring together scientists, engineers, and clinicians from diverse backgrounds to explore the opportunities and challenges introduced by the

growing abundance of digital data captured during the delivery of clinical care. With electronic health records and myriad other sources of healthcare-related digital data ushering the era of “big data” into medicine, the new challenge is to meaningfully use this deluge of clinical data to deliver improved patient care through exciting medical breakthroughs, new treatments and protocols, novel decision support tools, and insights that will transform how care is delivered at the bedside. The symposium will address four broad areas: 1) design of effective data acquisition and management systems (data collection/rationalization); 2) secure, HIPAA-compliant sharing of research data (sharing: HIPAA, data security, and legal issues); 3) modern analytical techniques for large clinical data sets (sophisticated analysis of this unique “big data”); and 4) translation of research into novel decision support tools (transition to bedside decision support). The symposium will feature talks from academia and industry, poster presentations, and panel-led discussions, and will also present the results from a “grand challenge” to derive meaning from a novel real-world critical care data set. For more information, visit <http://mucmd.org/>



OSCON - Community + Media Partners

For 13 years, OSCON, the O'Reilly Open Source Convention, has put open source to work building the future. This year, we've added two new conferences. OSCON Data and OSCON Java have graduated from tracks at previous OSCONs to become two distinct events, co-located with OSCON. OSCON's program encompasses 20 tracks to explore all things open source. New tracks have been added to OSCON this year, including Citizen Science, Geek Lifestyle, Open Data, Programming, and Emerging Languages, while Healthcare returns for its second year. OSCON Java tracks include Client, Cloud, Craftsmanship, JVM, and Trends and OSCON Data tracks include Analytics and Visualization, Hadoop, NoSQL Databases, Real-Time and Streaming, Relational, and Roulette. OSCON Data and OSCON Java premier July 25-27, while OSCON runs the entire week of July 25-29, 2011, at the Oregon Convention Center in Portland.

ApacheCon members can use discount code `os11apc` to save 15% on your registration.

OStatic - Community + Media Partners

OStatic aims to help users find the right Open Source software for their personal and business needs, evaluate Open Source alternatives to Proprietary Source applications and services, and collaborate with a network of trusted peers. As a member of the award-winning GigaOM Network, the OStatic blog delivers originally created content from a team of experienced editors and technologists. OStatic also provides a set of tools and services to help increase the development and adoption of Open Source Software and in 2008 OStatic was selected as PC Magazines Top 100 Sites on the Web.



ReadWriteWeb - Community + Media Partners

ReadWriteWeb is one of the most popular technology blogs in the world, known for offering insightful analysis about each day's Web technology news and trends. Ranked as one of the world's top 20 blogs by Technorati, it is syndicated daily by The New York Times.



Techvibes.com - Community + Media Partners

Techvibes.com is a hyperlocal technology blog, events calendar, job board, and company directory, with more than 200,000 visitors and over 375,000 pageviews every month. Spanning across ten major cities in Canada, Techvibes' readers are venture capitalists, entrepreneurs, and trendsetters in the tech industry.



TheNextWeb.com - Community + Media Partners

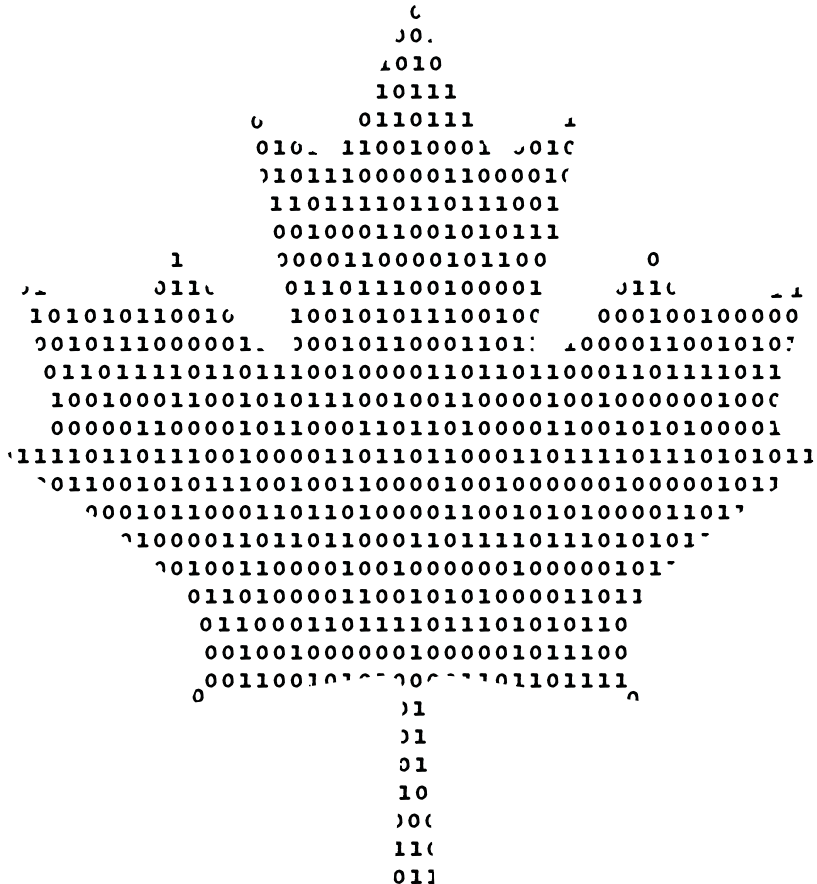
TheNextWeb.com is a worldwide leading source for technology news, business and culture. Over the last few years this site has grown into one of the world's leading technology blogs. We live to bring news and views from an international perspective. TNW has four million unique visitors per month, over 600k Twitter followers, 55k Facebook fans and the accolade of the 9 most influential blogs in the world (according to Technorati). With a brilliant team of editors & writers across Europe, Australia, Middle East, Africa, Asia and North & South America, you can rest assured that if it's worth hearing about: you'll hear it from us first.



WSO2Con - Community + Media Partners

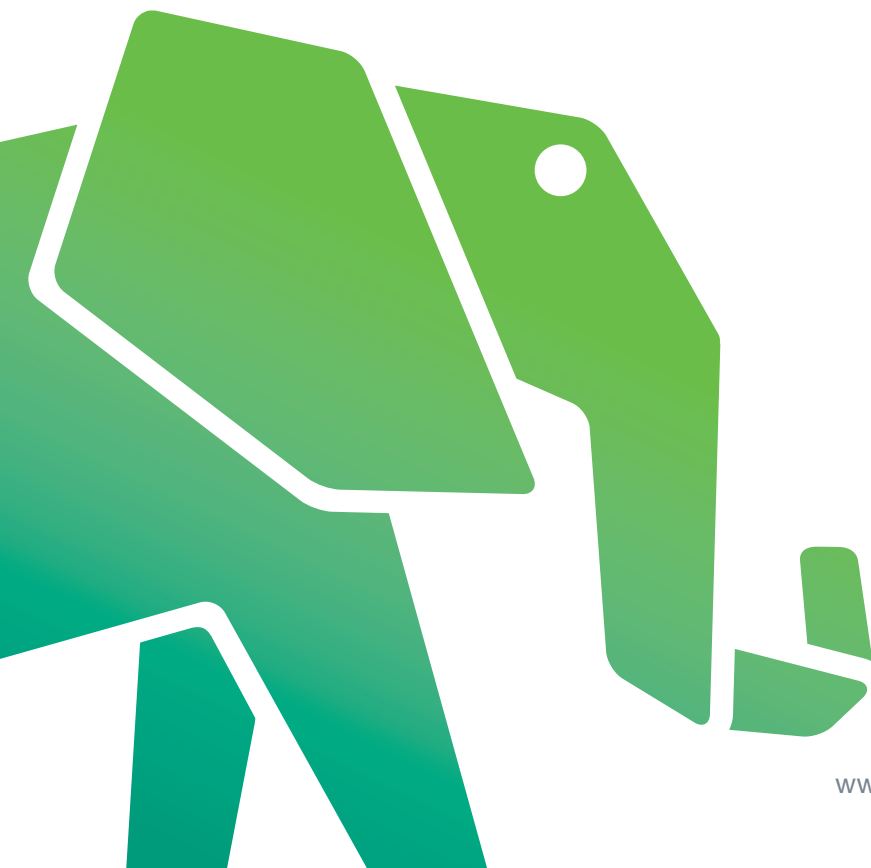
Architects, Developers and technology enthusiasts, WSO2Con 2011, focuses on you! Newbie or seasoned professional, WSO2Con offers a great week of learning, sharing and the opportunity to connect with experts who have shaped the future enterprise.

WSO2Con 2011 sessions cover everything from building lean enterprise middleware solutions to cloud computing, with special focus on real world customer stories, in-depth workshops and inspirational keynote presentations by industry leaders.



Architecting the Future of Big Data

- Hortonworks Data Platform
Powered by Apache Hadoop
- Hadoop Support
- Hadoop Training & Certification



Wednesday | November 9 | 9:30-10:20

David A. Wheeler | Building in Security and Innovation



The Apache Software Foundation develops and maintains software that the world depends on. But how can it help create software that with stands later attack? How can it counter malicious developers and repository subversion? And how can it encourage and enable innovation while doing so? This talk will broadly discuss how to resist attack while enabling innovation.

Thursday | November 10 | 13:30-14:20

Eric Baldeschwieler | The Apache Way Done Right: The Success of Hadoop



Apache Hadoop is emerging as the leading platform for managing big data in the enterprise. Hortonworks CEO Eric Baldeschwieler will discuss how Hadoop came to be an Apache project from its original inception at Yahoo!, the impact the community has had in driving innovation and spawning enterprise adoption around the technology, and future directions for Hadoop and the Hadoop community moving forward.

Friday | November 11 | 11:30-12:20

David Boloker | Watson, a Reasoning System: based on Apache Inside!



Much research work over the next decade will be driven by those seeking to solve complex problems employing the cloud, multicore processors, distributed data, business analytics, and mobile computing. In this talk I'll discuss some past approaches but also look at work being done in the labs on languages like X10 that extend the value of Java through parallelism, technologies that drive cross-stack interoperability, and approaches to handling and analyzing both structured and unstructured data.

Hackathon | 7-8 November

Sponsored by:   

If you are interested and have the desire and skills to code and develop Open Source software, please join us for this free Hackathon event. Meet Apache project contributors and work on projects together all day long.

BarCamp Apache | 8 & 12 November

BarCamp Apache is our free Apache-focused unconference. It will gather people interested in The Apache Software Foundation and it's various projects to share and learn in a completely open environment. BarCamp Apache will be intense with discussions, demos and interaction from participants packed into an ad-hoc, participant-driven event. BarCamp Apache is free.

Welcome Reception & PGP Keysigning Wednesday 9 November | 18:00-19:30

Sponsored by:     

Join us Wednesday evening to relax and meet and geek with the people you have seen all day but have been too busy with the conference to really sit down and talk to. Have a bit of food and a glass of beer or wine and enjoy yourself!

The Fast Feather Track

The Fast Feather Track provides space for projects that are too new or fast-moving to fit into another track. So if you want to get the skinny on an Apache Incubator project, the lowdown on the Labs, discover a great new feature in a project or some new technology out there, then this is your chance. These twenty-minute sessions are all about the technology!

Lightning Talks | Thursday 10 November | 18:00-19:00

Sponsored by: 

Come for the Talks, stay for the **FREE BEER**, courtesy of **Facebook!** This lively session has become a fun and expected ApacheCon tradition! Thursday evening, after the day's sessions are done, we will gather in the main hall for Lightning Talks. As a community, taking time to laugh together is as important as taking time to code together. The Lightning Talks are spontaneous and each year has its own unique highlights.

Join us Tuesday through Thursday evenings for FREE MeetUps!

Sponsored by:     

MeetUps are **FREE** and **open to everyone**, even if you're not registered for ApacheCon. Come and join us to participate in your favorite project, or to discover something new.

MeetUps start at 20:00 Tuesday through Thursday. Check the ApacheCon site for updates. Currently planned MeetUps include:

Tuesday

- » OpenOffice.org
- » TrafficServer
- » BigTop (Incubating)

Wednesday

- » Rave
- » Lucene/Solr
- » Axis2
- » Callback (Incubating)
- » Branding and Trademarks

Thursday

- » Hadoop
- » Cassandra
- » Chemistry
- » Tomcat
- » HTTP Server

08:00 - 09:00 **Morning Coffee and Lite Nibbles** - Sponsored by STONECIRCLE PRODUCTIONS, LLC

08:00 - 18:00 **Registration Open**

09:00 - 09:30 **Welcome & State of the Feather | Bayshore Grand Ballroom A, B & C**

09:30 - 10:20 **Keynote | Building in Security and Innovation | Bayshore Grand Ballroom A, B & C**
David A. Wheeler

10:30 - 11:00 **Break** - Sponsored by hp

	Salon A Apache in Space! (OOTD)	Salon B Data Handling & Analytics (Lucene & Friends) Sponsored by lucid	Salon C Business Sponsored by A	Salon E Servers (Tomcat & Geronimo) Sponsored by spring source	Salon F Innovation & Emerging Technologies
11:00 - 11:50	A look into the Apache OODT ecosystem Chris Mattmann	Lucene 4.0 - next generation open source search Simon Willnauer	The Business of Open Source Panel Sally Khudairi , Debbie Moynihan , Ross Turk , Kevin Carson , Bertrand Delacretaz	Using Tomcat-Native with Comet/Asynch call in Servlet 3.0 Jean Frederic Clere	Web Technologies in Java EE6 : Servlet 3.0 and JSF 2.0 Whats New ? Rohit Kelapure

12:00 - 13:30 **Lunch** - Sponsored by Adobe

13:30 - 14:30	Introduction to the OODT Process Control System Operator's User Interface (OpsUI) Andrew Hart	Bet You Didn't Know Lucene Can... Grant Ingersoll	How to talk to your Boss about Open Source without creating a Monster Tim O'Brien	Apache TomEE: Tomcat with a Kick David Blevins	Wicket meets Scala and CouchDB Bruno Borges
14:30 - 15:20	Using OODT @ JPL data systems Emily Law	Apache Solr: Out Of The Box Chris Hostetter	Opening Sourcing Commercial Software - Apache Traffic Server Bryan Call	Developing Java EE and Enterprise OSGi Applications with Apache Geronimo Kevan Miller	Apache Rave: Enterprise Social Networking Out-Of-The-Box Ate Douma , Matt Franklin


15:30 - 16:00 **Break** - Sponsored by Adobe

16:00 - 16:50	Using OODT to Support Data-driven Clinical Decision Support Andrew Hart	Solr Flair Erik Hatcher	Open Source enables Open Innovation Bertrand Delacretaz	Apache Geronimo 3.0 Deep Dive Kevan Miller	Fast Feather Talks <i>various speakers</i>
17:00 - 17:50	Supercharging your Apache OODT deployments with the Process Control System Chris Mattmann	Solr Power FTW: Powering NoSQL the World Over Eric Pinkin	Can I depend on Software built By Volunteers? Ross Gardler	Building a scalable multi-tenant Application Server on the Cloud using Tomcat, Axis2 & Synapse Afkhm Azeez	Fast Feather Talks <i>various speakers</i>

18:00 - 19:30 **Welcome Reception and PGP Keysigning | Bayshore Grand Ballroom Foyer**
Sponsored by b: cloudera Hortonworks hp spring source

20:00 - 22:00 **Meet Ups** - Sponsored by Acunu Hortonworks lucid OPEN DIRECTIVE

A Full Conference Program Guide is available for download at <http://www.apachecon.com>
 Conference Wi-Fi | username: **apache** | password: **8734**
 Event Tag: **#ApacheCon**
 To evaluate a talk, login or create an account at <http://na11.apachecon.com/connect>
 The evaluation rating system goes live 10 minutes after each session begins.

08:30 - 09:30 **Morning Coffee and Lite Nibbles** - Sponsored by 
08:30 - 17:00 **Registration Open**


Salon A Enterprise Java Sponsored by 	Salon B Data Handling - Big Data	Salon C Community Sponsored by 	Salon E Servers (HTTPD)	Salon F Content Technologies
--	---	--	--	---

10:00 - 10:50 Apache ServiceMix future Jean-Baptiste Onofré	The Past, Present and Future of NOSQL Emil Eifrem	The Secret Life of Open Source Ted Husted	Apache httpd 2.4: The Web Server for the Cloud Jim Jagielski	If you have the content, Apache has the technology Nick Burch
---	---	---	--	---

11:00 - 11:30 **Break** - Sponsored by 

11:30 - 12:20 An architecture for enabling multi-tenancy for Apache Axis2 Afkham Azeez	Scaling Hadoop Applications Milind Bhandarkar	Talking people into creating patches Isabel Drost	mod_lua for beginners Eric Covener	Apache Tika: 1 point Oh! Chris Mattmann
--	---	---	--	---


12:20 - 13:30 **Lunch** - Sponsored by 

13:30 - 14:20 **Keynote | The Apache Way Done Right: The Success of Hadoop | Bayshore Grand Ballroom A, B & C**


14:30 - 15:20 An Interactive Example of Enterprise SOA, Apache Style Hadrian Zbarcea	The other Apache technologies your big data solution needs Nick Burch	Navigating the Apache Incubator Brett Porter	The Power of the mod_proxy Modules Paul Weinstein	ManifoldCF for Content Acquisition Karl Wright
--	---	--	---	--

15:30 - 16:00 **Break** - Sponsored by 


16:00 - 16:50 ActiveMQ In Action: Common Problems and Solutions Bruce Snyder	Realtime Big Data at Facebook with Hadoop and HBase Jonathan Gray	Life in Open Source communities Bertrand Delacretaz	Hardening Enterprise Apache Installations Against Attacks Sander Temme	Interoperability with CMIS and Apache Chemistry Florian Müller
17:00 - 17:50 Security Problems (and Solutions) for Service Oriented Applications Daniel Kulp	Dr. Mahout: Analyzing clinical data using scalable and distributed computing Shannon Quinn	Chefs with Feathers: The Sakai Project Carl Hall	Out and About with Apache Traffic Server Leif Hedstrom	Handling RDF data with Apache Jena Paolo Castagna

18:00 - 19:00 **Lightning Talks | Bayshore Grand Ballroom A, B & C** - Sponsored by 

20:00 - 22:00 **MeetUps** - Sponsored by    

08:00 - 09:00 **Morning Coffee and Lite Nibbles** - Sponsored by 


08:00 - 09:00 **Registration Open**

	Salon A Modular Java	Salon B Data handling & analytics (Hadoop & friends) <small>Sponsored by </small>	Salon C Infrastructure & Dev Ops	Salon E Innovation & Emerging Technologies	Salon F Content Technologies
09:00 - 09:50	OSGi for mere mortals Bertrand Delacretaz	State of the Elephant: Hadoop yesterday, today and tomorrow Owen O'Malley	Serving "Web" Over IPv6 Issac Goldstand	Rapid application development for dynamic cloud applications Bram de Kruijff , Marcel Offermans	Becoming a content-driven, modular application: A Case Study Brett Porter
10:00 - 10:50	Apache Celix - Universal OSGi? Alexander Broekhuis	Apache Mahout for intelligent data analysis Isabel Drost	Inside the Apache Infrastructure Tream Philip M. Gollucci	Whirr: Open Source Cloud Services Tom White	Bridging traditional Open Source Content Management and the Web of Data with the Apache Stanbol Semantic Engine Olivier Grisel

11:00 - 11:30 **Break** - Sponsored by 

Keynote | Watson, a Reasoning System: based on Apache Inside! | Bayshore Grand

11:30 - 12:20 **Ballroom A, B & C**
[David Boloker](#)

12:20 - 13:30 **Lunch** - Sponsored by 

13:30 - 14:20	Deployment With Apache Karaf and ACE Jean-Baptiste Onofré	Instrumenting Hadoop Jobs for Fun and Profit Shevek	Using JMeter For Testing A Data Center Siegfried Goeschl	Breaking Down Widget Silos with a friendly Wookiee Ross Gardler	Kafka - A distributed publish/subscribe messaging system Neha Narkhede
14:30 - 15:20	Provisioning distributed OSGi applications in a cloud Guillaume Nodet	Cassandra 1.0 and beyond Jake Luciani	From Dev to DevOps Carlos Sanchez	.NET @ Apache.org Ted Husted	NoSQL at work with JCR and Apache Jackrabbit Carsten Ziegeler
15:30 - 16:20	Using OSGi to Build Better Software: Lessons from a Telemedicine Software for Smartphones and Desktop Systems Doreen Seider	Hot HA for Hadoop NameNode Konstantin V. Shvachko	FreeBSD + ASF software & philosophy + ZFS == large \$\$ bonuses from your boss Philip M. Gollucci	Apache Commons Nabl: on the fly bytecode transformations for algorithmic differentiation Phil Steitz	Building a state of the art Content Repository with Apache Chemistry and CMIS Stephan Klevenz

16:30 - 17:00 **Closing Session with Coffee, Tea & Goodies | Bayshore Grand Ballroom A, B & C**
Sponsored by  

HP is proud to sponsor ApacheCon North America 2011

"Open Source Enterprise Solutions, Cloud Computing,
and Community Leadership"



Absolutely committed
to open source.

For over a decade more companies have chosen
HP servers to run Linux than any other vendor.



See what's possible at hp.com/go/linux

.NET @ Apache.org

Ted Husted

15:00-15:50 | Fri. 11/11

Like it or not, many open source developers are moving to the Microsoft .NET platform, and we're bringing our favorite tools with us! In this session, we look inside ASF projects that are creating software for .NET and Mono -- like ActiveMQ, Chemistry, Logging, Lucene, QPid, and Thrift -- and show how to create leading-edge ASP.NET applications with ASF open source libraries. We'll also look at integrating other .NET open source projects, like Spring.NET, NVelocity, and JayRock, into your C# application to create a complete open source .NET stack.

A Look Into the Apache OODT ecosystem

Chris Mattmann

11:00 - 11:50 | Wed. 11/9

Apache OODT is the first ever NASA project to be hosted at the Apache Software Foundation. After nearly a year of Incubation time, during which the OODT community translated from a set of collaborating NASA, university, and other government institutions into a set of collaborating individuals working together at the Apache Software Foundation, Apache OODT was made a top level project in November 2010.

One of the most frequent things we see in the OODT community however is the desire from our users and other individuals to know more about the overall ecosystem. As a project and community, OODT has existed for 10+ years, and spans

the areas of technology, research, and academics, including numerous book chapters, journal articles and peer reviewed conference publications documenting OODT's use in different science domains and across multiple projects.

This overview talk will give attendees insight into OODT's history: its community: its projects: and its ecosystem. Topics will include:

- overview of OODT and its history
- how we got it to Apache
- how other projects can follow our model
- existing successful deployments of OODT
- pointers to papers, and more information including case studies

ActiveMQ In Action: Common Problems and Solutions

Bruce Snyder

16:00-16:50 | Thurs. 11/10

Why does ActiveMQ just stop sending messages and hang? What's the best way to build a JMS consumer and producer? What configuration should I use for connection pooling? How can I query ActiveMQ for the message I need? Should I cluster ActiveMQ or group clients across brokers? If you use ActiveMQ, chances are you have run into some questions that are easily answered with a little knowledge. This session examines the top five questions from developers using ActiveMQ.

An architecture for enabling multi-tenancy for Apache Axis2

Afkham Azeez**11:30-12:20 | Thurs. 11/10**

Web services is the most popular technology for implementing an SOA. SOA & Cloud computing are a natural fit. Apache Axis2 is the most widely used Java Web services server in the industry. Deploying Web services on the Cloud would give all the benefits provided by Cloud Computing. However, deploying Web service applications on the Cloud requires that sharing of resources & infrastructure has to be optimally done without compromising security.

In this session, we will look at an architecture which will allow you to deploy Web services within a single Axis2 cluster in a multi-tenant fashion. We will also look at how to adopt this architecture for deploying Axis2 Web services as SaaS. We will demonstrate this in practice using the WSO2 Stratos Platform-as-a-Service (PaaS)

An Interactive Example of Enterprise SOA, Apache Style**Hadrian Zbarcea****14:30-15:20 | Thurs. 11/10**

Join us in this interactive session where Dan Kulp and Hadrian Zbarcea from Talend will demonstrate a practical approach to developing distributed applications using the most popular OSS SOA ecosystem – Apache CXF, Apache Camel, Apache Karaf and Apache ActiveMQ. Working with live feeds driven by attendee participation from the ApacheCon floor, you'll see firsthand how these components collaborate to fulfill complex service-oriented

requirements. We'll explore best practices of configuring services within a SOA environment using this fun, live demo.

Apache Celix - Universal OSGi?**Alexander Broekhuis****10:00-10:50:00 | Fri. 11/11**

Systems which require dynamics and interoperability need a good architecture and clear design principles. OSGi provides this for Java based systems, but for embedded/native systems currently no alternative is available.

To fill this gap, Apache Celix is created. Apache Celix is an implementation of the OSGi specification adapted to C, focused on embedded and distributed platforms. Celix is a new Apache project currently starting up in the Incubator.

The goal is to follow the OSGi specification as much as possible, and grow towards an implementation with most of the OSGi Core and Compendium implemented.

To be able to support distributed platforms, the Remote Services from the Enterprise Specification will be implemented. Using Remote Services also makes it possible to create interoperability with Java OSGi.

For distributed systems, deployment/distribution is an important aspect. For OSGi Apache Ace can be used to maintain and manage deployment to multiple targets. This makes Apache Ace a

perfect candidate for deployment of Celix Bundles and Artifacts.

This presentation shows how Celix solves the dynamic aspects of OSGi services in C. It will detail differences between the OSGi Specification and the Celix solution.

Apache Commons Nablax: On the Fly Bytecode Transformations for Algorithmic Differentiation

Phil Steitz

16:00-16:50 | Fri. 11/11

This talk presents an innovative use of the Java platform: using bytecode transformations to perform mathematical differentiation. This kind of operation is at the basis of numerous algorithms. It is straightforward for small equations but becomes a daunting task when applied to complex simulation models.

Nablax (named after the differentiation operator) attempts to do it directly on compiled code and on the fly at runtime.

The various issues related to these transformations are explained (class creation, instance creation, access to private parts, data sharing between differentiated and primitive instances ...).

Apache Geronimo 3.0 Deep Dive

Kevan Miller

16:00- 16:50 | Wed. 11/9

In its 3.0 release, Apache Geronimo 3.0 has been rebased on an OSGi runtime. Geronimo's traditional modularity is now expressed as

OSGi bundles. Server assembly and application-centered deployment are now done by installing sets of bundles into an OSGi core framework. JavaEE applications are deployed by transforming them into collections of bundles. In this talk, we will review this extensive restructuring of the Geronimo runtime and discuss the new capabilities that this new framework provides.

Apache httpd 2.4: The Web Server for the Cloud

Jim Jagielski

10:00-10:50 | Thurs. 11/10

The latest version of the venerable Apache web server, v2.4, includes numerous enhancements over all previous versions. But the biggest improvements have been made in creating a web server which is ideally suited for Cloud environments.

This session will go into the improvements within Apache httpd of interest to cloud developers and users.

Apache Mahout for Intelligent Data Analysis

Isabel Drost

10:00-10:50 | Fri. 11/11

"Searching the internet" has become a common pattern when looking for information. However with current tools finding the relevant piece of data often turns out to be similar to searching for the needle in a haystack of unstructured information: social networks, corporate content management systems, micro blogging platforms,

tend to generate an ever increasing flow of online data.

This talk gives an introduction to Apache Mahout - a framework of scalable implementations of algorithms for data mining and machine learning. After motivating the need for machine learning the talk gives an overview of the features of Apache Mahout. The talk shows how to integrate Mahout into your application. It shows the tremendous improvements that have been implemented in recent past - including the addition of several algorithms, performance improvements and better APIs for integration.

Apache Rave: Enterprise Social Networking Out-Of-The-Box

Ate Douma, Matt Franklin

14:30 -15:20 | Wed. 11/9

Apache Rave is a new Incubator project commissioned to develop an open standards-based web and social mash-up engine for the Enterprise. The goal for Apache Rave is to become a leading open-source, context-aware User Experience Platform capable of delivering a wide variety of interaction channels, using open standards technologies as OpenSocial and W3C widgets and REST based services. As Rave advances, it will add features like content integration, collaboration, personalization, and rich user application features.

Apache Rave was initiated in 2010 at the previous ApacheCon in Atlanta and through subsequent discussions during the first

European OpenSocial Event in Utrecht (The Netherlands), and thereafter created as a joint effort of several existing projects and individual participants. The Rave project thus provides an interesting case-study for bringing together multiple projects with already mature code bases. Rave members are also working together to reach out to other related Apache projects like Shindig and Wookie, to other potentially interested projects and developers both within and outside Apache, and most importantly to the many and diverse target communities.

During the presentation project initiators from Hippo, The MITRE Corporation and Indiana University will provide an overview of Apache Rave: the current status, its features and the goals, and talk about the process getting there.

Apache ServiceMix Future

Jean-Baptiste Onofré

10:00-10:50 | Thurs. 11/10

Apache ServiceMix is a flexible, open-source integration container that unifies the features and functionality of Apache ActiveMQ, Camel, CXF, ODE, Karaf into a powerful runtime platform you can use to build your own integrations solutions. It provides a complete, enterprise ready ESB exclusively powered by OSGi.

The main features are:

- reliable messaging with Apache ActiveMQ
- messaging, routing and Enterprise Integration Patterns with Apache Camel
- WS-* and RESTful web services

with Apache CXF

- loosely coupled integration between all the other components with Apache ServiceMix NMR including rich Event, Messaging and Audit API
 - complete WS-BPEL engine with Apache ODE
 - OSGi-based server runtime powered by Apache Karaf
- If ServiceMix3 was focus on JBI/Spring, JBI support is optional in ServiceMix4.

This session will introduce the discussion around ServiceMix5 and the roadmap of ServiceMix "NMR".

Apache Solr: Out Of The Box

Chris Hostetter

13:30-14:20 | Wed. 11/9

Apache Solr is the popular, blazing fast open source enterprise search platform from the Apache Lucene project. In this session we will see how quick and easy it can be to install and configure Solr to provide full-text searching of structured data without needing to write any custom code. We will demonstrate various built-in features such as: loading data from CSV files, tolerant parsing of user input, faceted searching, highlighting matched text in results, and retrieving search results in a variety of formats (XML, JSON, etc) We will also look at using Solr's Administrative interface to understand how different text analysis configuration options affect our results, and why various results score the way they do against different searches.

Apache Tika: 1 Point Oh!

Chris Mattmann

11:30-12:20 | Thurs. 11/10

Apache Tika, since April 2010 an ASF top level project, and a thriving Apache community has made tremendous strides over the past 4 years to grow and mature into a leading text extraction library, and content detection framework. Tika is used in a number of search projects, in a number of data management systems, and in a number of domains.

Those domains span from the technical industry to domains of science and within the federal government.

Tika has been used as a teaching platform for computer science graduate students, has been used to unlock information from NASA images, and from the National Cancer Institute, and has also been used to provide rich meaning and information representation of content captured in pervasive document repositories and warehouses. These are only some of Tika's broad applications.

In November, we hope to have released Tika 1.0. This will coincide with a number of other properties that demonstrate Tika has reached the point of a mature community, including:

1. Concrete, stable features, and core interfaces.
2. Tika's use in multiple programming languages and environments.
3. Our growth in Apache, and election of new committers and

- PMC members (and ASF members).
4. Developer articles appearing quite frequently on Tika.
 5. The culmination of a wealth of knowledge in the form of a book that will be published on Tika at the time of the ApacheCon meeting.

This talk will focus on how we got here, and what's next for this thriving Apache community.

Apache TomEE: Tomcat with a Kick

David Blevins

13:30-14:20 | Wed. 11/9

Apache TomEE, pronounced "Tommy", is a simple all-Apache stack aimed at Java EE 6 Web Profile certification where Tomcat is top dog. The desire to beef up Tomcat installations has persisted despite the existence of full-blown app servers, many of which include Tomcat in some truncated and stripped-down form. With today's embedded APIs, lighter-weight components and the Java EE 6 Web Profile, that can now be fully realized. Pioneered from the Apache OpenEJB project and including Apache OpenWebBeans, Apache MyFaces, Apache ActiveMQ, Apache OpenJPA and Apache CXF, the Apache TomEE integration of these technologies is simple, to-the-point, and focused on the singular task of delivering the Java EE 6 Web Profile in a minimalist fashion. Finally you can get a little more out of your lightweight Tomcat apps without having to give up anything in the process.

Becoming a Content-Driven, Modular Application: A Case Study

Brett Porter

09:00-09:50 | Fri. 11/11

Further building on the session from ApacheCon NA 2010, see how Apache Archiva revitalised itself by moving to being content-driven, and becoming more modular. The discussion includes moving from a database and ORM solution, how we defined the content model for the application and started migrating the data and the architecture, and the natural benefits we immediately found in the process. We also look at how the application was modularised more effectively, and the possibility at making it more dynamic with other component models and OSGi. We will discuss the other technologies such as Apache Jackrabbit and Apache Felix we used and evaluated and what we learned about them on the way.

Bet You Didn't Know Lucene Can...

Grant Ingersoll

14:30 -15:20 | Wed. 11/9

Lucene and Solr have always provided very capable text search, but did you know it is useful for many other things as well?

In this talk, we'll take a look at some of the myriad of ways that Lucene and Solr can be used to solve real world challenges ranging from classifying content, recommending movies all the way through to taking your unit testing to the next level.

Breaking Down Widget Silos with a Friendly Wookie

Ross Gardler

14:00-14:50 | Fri. 11/11

Widgets/gadgets are mini applications written in HTML + Javascript. They offer cool and dynamic content that can be placed on any page on the web and, in some cases, on your desktop or your mobile device. Unfortunately, there is not just one way to create and package widgets. We have Google Gadgets, W3C Widgets, OpenSocial Gadgets and Wave Gadgets to name just a few. Whilst widgets are an important part of web content delivery, particularly mobile web, the plethora of available widget/gadget standards could limit innovation by creating incompatible silos. This is where Apache Wookie (Incubating) comes in. Using Wookie we can harmonize all of these widgets/gadget standards behind the W3C Widget specification, thus freeing the user from concerns about implementation details.

Bridging traditional Open Source Content Management and the Web of Data with the Apache Stanbol Semantic Engine

Olivier Grisel

10:00-10:50 | Fri. 11/11

This talk will introduce the Stanbol project and showcase how it can be integrated in traditional Enterprise Content Management solutions.

Stanbol is an Open Source project under incubation at the Apache Software Foundation. Its goal is to provide Web and CMS developers with a set of HTTP / RESTful

services to help them integrate semantic technologies into their products and web sites.

The following Stanbol services are currently under active developments:

- Enhancement engines: use Natural Language Processing tools such as Apache OpenNLP to extract knowledge (topics, named entities, facts) from unstructured content and link it to unambiguous URIs from reference knowledge bases.

- Entity Hub: a Linked Data indexing cache built on top of Apache Solr, Clerezza and Jena that comes with precomputed indexes and live connectors to popular knowledge bases such as DBpedia, Geonames, YAGO...

- Content Hub: a faceted search engine based on Solr to search for content using the knowledge automatically extracted by the enhancement engines.

- CMS bridges to lift the structured content of document repositories using the JCR and CMIS access protocols (using Apache Chemistry) and store the result into a triple store suitable for SPARQL access;

- Rules engine based on Apache Jena for knowledge refactoring (e.g. convert extracted knowledge into the rich snippet vocabulary for SEO), integrity checks, merging rules, deductive inference...

The Semantic Web has made significant progress over the last years, and while it always gave a

lot of promises, it is now the time where it can concretely be used in Enterprise Solutions.

If you are curious about the web of data, and want to see how concretely it can be used and integrated today in enterprise solutions thanks to software like the Stanbol projects, this session is for you.

You should also attend if you are interested in emerging technologies and don't have knowledge about semantic technologies, this will provide a good insight on how they can disrupt the usual way to develop applications.

Building a Scalable Multi-Tenant Application Server on the Cloud using Tomcat, Axis2 & Synapse

Afkham Azeez
17:00 17:50 | Wed. 11/9

Apache Tomcat is one of most popular & widely used Application Servers, and Apache Axis2 is one of most widely used Java Web services servers. Apache Synapse is one of the popular, high performant ESBs widely used in the industry.

In this session, we will look at how we combine these great projects from the ASF into building a scalable, elastic, multi-tenanted Application Server, which allows you to deploy cloud-native webapps on the Cloud, and benefit from all the advantages that Cloud Computing brings in. We will also see how easy it is to deploy any standard webapp on the Cloud, and seamlessly integrate with the authentication, authorization

& management infrastructure provided by the underlying Platform-as-a-Service (PaaS).

Building a State of the Art Content Repository with Apache Chemistry and CMIS

Stephan Klevenz

16:00-16:50 | Fri. 11/11

This talk will provide a global overview of Content Repository technology and explain in more detail how implementing the CMIS standard (Content Management Interoperability Services) within the Apache Chemistry project significantly changed the landscape of content management technology and enabled developers to build more sophisticated content repositories.

This talk will be of interest to any developer or architect curious to discover how a Content Repository can be a great middleware to help develop faster and better applications dealing with structured and unstructured data. It is also recommended to anyone who is interested in standards and who would like to have a better understanding of the OASIS CMIS standard and how it is implemented within the Apache Chemistry project.

The talk will provide:

- A global understanding of what a Content Repository is from a functional standpoint: exploring all the services it offers, identifying the main standards and technologies integrated (such as CMIS, which is a key one), and understanding the main technical challenges to be

resolved, such as high scalability and high performance.

- An introduction and presentation of the Apache Chemistry project, which became an Apache top level project earlier this year.
- A retrospective on the evolution of this project and what it can bring compared to other technologies such as JCR.

Can I Depend on Software Built By Volunteers?

Ross Gardler

17:00 17:50 | Wed. 11/9

Apache projects are managed by volunteers, can you really build your products, infrastructure or services on software managed by volunteers? The answer to this commonly asked question is a definite yes, in this session Ross Gardler will explain why this is the case. At the core of the argument is the fact that whilst contributors are volunteers here at the ASF they will almost certainly be paid by someone for the time they spend working on our projects. This prompts a second common concern: can you really build your products, infrastructure or services on Apache software if you haven't got an army of staff to ensure the project is not "hijacked" by a third party? Once again, the answer to this question is a definite yes, in this session Ross will explain why this is the case. In this session we will examine the meritocratic governance model used in Apache projects and explain how it ensures that even the smallest of organizations can become an important, even critical, part of a project team whilst also ensuring

that no single organization can take control of a project by throwing resources at it.

Cassandra 1.0 and Beyond

Jake Luciani

15:00-15:50 | Fri. 11/11

The Cassandra distributed database has added many new features this year based on real-world needs of developers at Twitter, Netflix, Openwave, and others building massively scalable systems.

This talk will cover the motivation and use cases behind features such as secondary indexes, Hadoop integration, SQL support, bulk loading, and more.

Introduction

- * Project history and goals

Recap: Cassandra through 2010

- * Bulletproof reliability
- * Best-in-class support for multiple datacenters
- * High-performance storage engine based on Bigtable

New in Cassandra 1.0

- * Dynamic column indexes
- * Distributed counters for realtime analytics
- * CQL/SQL and JDBC support
- * Bulk loading
- * Off-heap allocation for GC performance
- * Hadoop support

Chefs with Feathers: The Sakai Project

Carl Hall

17:00-17:59 | Thurs. 11/10

The Sakai Project is not only a collaboration of learning institutions but also a marriage of many open source projects. It is this ecosystem of software systems that drives the success of Sakai's products. The interactivity with the various Apache communities that we build upon have strengthened our product and have given us the chance to contribute back. We would like to present our experiences building a large collaborative learning environment on top of open-source software from the ASF.

Deployment with Apache Karaf and ACE

Jean-Baptiste Onofré

14:00-14:50 | Fri. 11/11

Deployment of applications based on reusable components in a large network environment can become complicated very quickly. In this session, Jean-Baptiste Onofré from Talend will describe how to manage deployment tasks using Apache Karaf and Apache ACE.

Apache Karaf is a flexible, lightweight, enterprise-ready OSGi container that provides a runtime for a wide variety of components, including pure web applications and ESB-oriented services. Karaf's flexible tooling makes it suitable for large-scale deployment, and it supports multiple instances through high-availability and clustering.

Apache ACE is a software

distribution framework that provides centralized management of multi-node component deployments.

Jean-Baptiste will begin with an introduction to Karaf, covering the Karaf shell and the basics of multiple instance management. Then, he will demonstrate how to use ACE to provision applications running inside Karaf.

Developing Java EE and Enterprise OSGi Applications with Apache Geronimo

Kevan Miller

14:30 -15:20 | Wed. 11/9

Apache Geronimo 3.0 supports both Java EE 6 and Enterprise OSGi programming models. In this presentation, we will review Geronimo 3.0 features, discuss server administration, and demonstrate the development of Java EE and OSGi applications using Apache Geronimo.

Dr. Mahout: Analyzing Clinical Data using Scalable and Distributed Computing

Shannon Quinn

17:00-17:59 | Thurs. 11/10

Of the few realms cloud computing has not solidly taken root, one in which it has great potential is medicine. Clinicians generate massive amounts of data during the diagnostic process, the analysis of which, whether manual or computational, can take a great deal of time. For example, the rare genetic disease primary ciliary dyskinesia (PCD) affects the cilia on cells, causing them to behave

erratically and leading to breathing problems at best, necessitating lung transplants at worst. Cutting-edge diagnostic tools capture the ciliary motions with high-speed video and use automated methods to quantitatively describe the motion patterns. These methods, however, are computing-intensive and would benefit from parallelization. Here we propose using the Mahout framework to efficiently learn models that capture the motion patterns observed in the videos and aiding in objective diagnoses. Additionally, Hadoop's storage system will allow us to construct and preserve libraries of these motion models in the cloud for later comparison. The library will be in constant flux as new patterns are added and existing patterns are retrained, requiring a scalable and distributed architecture to handle the data and integrate it into the existing library. Ultimately this framework will be a boon for clinicians: they need only take biopsies, gather data as images or videos, upload them to a Mahout/Hadoop cluster, and wait for the results. Patient privacy is maintained by perpetuating only the low-dimensional motion models, computational time is reduced by parallelizing the model learning and comparison process, and models are available to clinicians everywhere through the cloud.

Fast Feather Talks

16:00- 16:50 | Wed. 11/9

For those of you who've yet to discover Fast Feather sessions, they are a series of fast-paced 20 minute talks covering new and exciting

projects and activities at and around Apache.

Fast Feather Talks

17:00 17:50 | Wed. 11/9

For those of you who've yet to discover Fast Feather sessions, they are a series of fast-paced 20 minute talks covering new and exciting projects and activities at and around Apache.

FreeBSD + ASF Software & Philosophy + ZFS == Large \$\$ Bonuses from your Boss

Philip M. Gollucci

16:00-16:50 | Fri. 11/11

A record holding TCP/IP stack, *the* cutting edge File System, and most respected Software collection..... What do you get? A much happier system admin and a much better work place. The cloud might be hot stuff, but its not for everyone or every task.

How can you Take your Existing Data Center Infrastructure and Make it Better?

From Dev to DevOps

Carlos Sanchez

15:00-15:50 | Fri. 11/11

The DevOps movement aims to improve communication between developers and operations teams to solve critical issues such as fear of change and risky deployments. But the same way that Agile development would likely fail without continuous integration tools, the DevOps principles need tools to make them real, and provide the automation required to actually be implemented. Most of the so called DevOps tools

focus on the operations side, and there should be more than that, the automation must cover the full process, Dev to QA to Ops and be as automated and agile as possible. Tools in each part of the workflow have evolved in their own silos, and with the support of their own target teams. But a true DevOps mentality requires a seamless process from the start of development to the end in production deployments and maintenance, and for a process to be successful there must be tools that take the burden out of humans. Apache Maven has arguably been the most successful tool for development, project standardization and automation introduced in the last years. On the operations side we have open source tools like Puppet or Chef that are becoming increasingly popular to automate infrastructure maintenance and server provisioning.

In this presentation we will introduce an end-to-end development-to-production process that will take advantage of Maven and Puppet, each of them at their strong points, and open source tools to automate the handover between them, automating continuous build and deployment, continuous delivery, from source code to any number of application servers managed with Puppet, running either in physical hardware or the cloud, handling new continuous integration builds and releases automatically through several stages and environments such as development, QA, and production.

Handling RDF Data with Apache Jena

Paolo Castagna

17:00-17:59 | Thurs. 11/10

Apache Jena, currently in incubation, is a Java framework for building semantic web applications. It provides developers with a library to handle RDF, RDFS, RDFa, OWL and SPARQL in line with the relevant W3C recommendations.

Jena has been developed by researchers at HP Labs, Bristol (UK) starting back in 2000. It has been an open source project since its beginning and it is extensively used within the semantic web community.

This talk introduces the fundamentals of the RDF data model and SPARQL query language as well as the basic usage patterns of Jena: how to parse and write data in RDF format, how to store it using TDB, Jena's native RDF database, querying with SPARQL using ARQ and how to integrate free text searches with SPARQL using Apache Lucene or Solr.

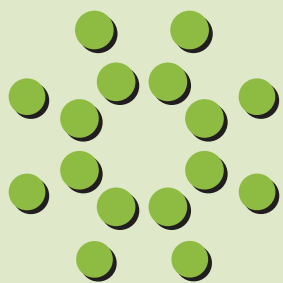
At Talis we use Apache Jena, in particular TDB, ARQ, LARQ and Apache Hadoop in our ingestion pipeline, as well as many other open source projects, to process RDF data, store it, implement our services and APIs.

Hardening Enterprise Apache Installations Against Attacks

Sander Temme

16:00-16:50 | Thurs. 11/10

Enterprise installations of Apache

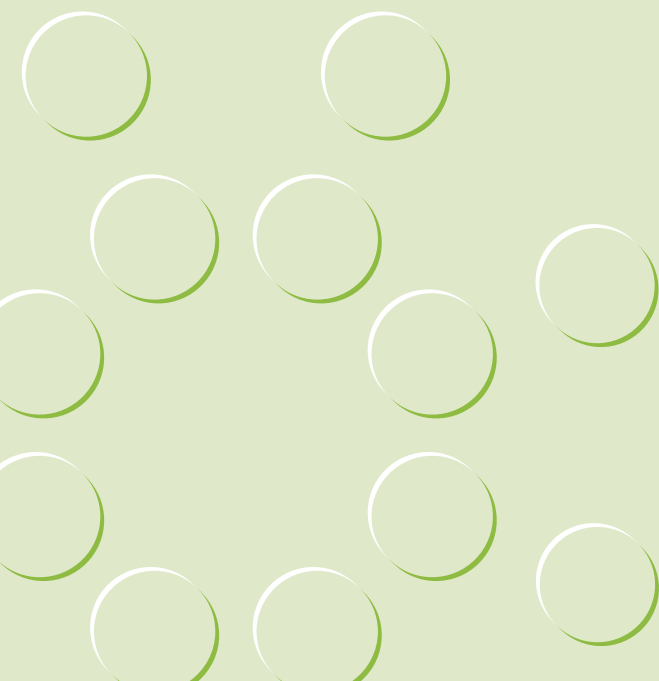


NING

is proud to sponsor
the Hackathon at

ApacheCon

North America 2011



are particularly attractive targets for malicious attacks including Denial of Service, defacement, theft of data or service and installation of zombies or viruses.

Hardening your deployment against such attacks calls for some special techniques and tactics.

Come to this session to learn about attack detection techniques, server protection, secure deployment of multiple servers, configuration of firewall "demilitarized zones" and judicious use of SSL encryption.

How do you deploy an off-the-shelf application that insists on writing to the file system?

And what steps do you take to securely deploy Apache on Windows or UNIX?

This presentation will explore solutions to these very real situations.

Hot HA for Hadoop NameNode

Konstantin Shvachko

16:00-16:50 | Fri. 11/11

Current HDFS design assumes that a single server, NameNode, dedicated to maintaining the file system metadata, controls the work of other cluster nodes, DataNodes, handling actual file data blocks. The system is designed to survive and recover in minutes from a loss of multiple DataNodes. But the NameNode failure makes the entire cluster unavailable, since there is no other place to obtain metadata information immediately. Although this design simplifies

overall architecture of HDFS, it also makes the NameNode a single point failure, which is considered a serious deficiency for production grade systems.

The primary goal of the proposed architecture is to build a highly available NameNode, which can failover to a Standby node in seconds, and which requires minimum changes to the existing code base.

The architecture introduces a StandbyNode, which is an evolutionary modification of BackupNode already existing in HDFS. This is the only major change required to the current Hadoop code base. The approach further utilizes standard HA software like LinuxHA, and existing functionality of load balancing hardware or software platforms. The system is prototyped on eBay clusters.

How to Talk to your Boss about Open Source without Creating a Monster

Tim O'Brien

13:30-14:20 | Wed. 11/9

As developers we often drive the adoption of open source technology in addition to being tasked with explaining the concepts of open source communities to management. While every company seems to be well into "leveraging open source models", there's still a great deal of confusion among management. For example, have you ever heard some C-level executive try to describe "open source"? Did it sound something like the following paragraph?

**** The past decade saw a lot of CTOs thinking "outside the box", attempting to "leverage" the collective "synergies" of open source to improve key performance indicators (KPI) all while preserving the necessary provenance and governance structures that allow the organization to really "fire on all cylinders". As they understand open source, the collaborative paradigm-shift of distributed "knowledge-workers" allows interested parties to benefit from external sources of innovation giving management the opportunity to achieve greater dominance of vertical markets without having to expend critical resources on issues irrelevant to other C-level stakeholders. In short, Open Source in the enterprise is a win-win, and everybody likes a win-win, right?

If that paragraph made you want to cry, this presentation is for you. We're going to take a look at some of the issues that arise when your boss doesn't fully understand how open source works, and we're also going to explore some ways in which you can help guide management toward a fuller, more accurate appreciation of the culture.

A humorous analysis of the various ways in which "open-source" is misapplied and misunderstood in the corporation will be presented. Along the way we'll use these misconceptions to make some strong recommendations for how not to bring Apache to your organization.

The following topics will be

explored:

- * Consensus-driven Development in a Business Settings
- * Self-organizing Communities in the Corporation
- * How not to confuse your boss about "Open Source", answers to the following questions and statements:
 - * "Right so Open Source is really just like Agile?" No.
 - * "Why aren't these open source developers doing what we need them to do?" Errr.
 - * "I don't want to spend any money developing open source. That's not what I signed up for."
 - * Lastly, what if a corporation really was run as a pure Meritocracy? what would it look like?

If you have the Content, Apache has the Technology

Nick Burch

10:00-10:50 | Thurs. 11/10

Within the ASF, there are a wide variety of projects with technologies to help you store, retrieve, host, transform and generate content. This talk will review the landscape of Apache content technologies, provide a quick introduction to the more common and more interesting projects, and flag up new and innovative features within them. It'll also highlight talks from the rest of the week on many of the projects covered, so that you'll know where and when to go to learn more about those projects and technologies which catch your eye!

Inside the Apache Infrastructure Team

Philip M. Gollucci

10:00-10:50 | Fri. 11/11

Have you ever wondered how a team of barely 10 rotating volunteers and 2 paid staff can manage a 24x7x365 infrastructure that spans 3 continents and is used by millions of people across the globe? All without any offices and the bare minimum of paper work? Well now's your chance! Hear it directly from the camels mouth, the VP of Apache Infrastructure. We'll tell you all the technologies, how staff is managed, how responsibilities are delegated out, how we deal with 3rd party vendors, and best of all how we leverage the Apache Way to accomplish our goals.

Instrumenting Hadoop Jobs for Fun and Profit

Shevek

14:00-14:50 | Fri. 11/11

Instrumentation is a general purpose technique to automatically gather detailed information about the execution of a process.

The distributed nature of a Hadoop job makes both the engineering of the instrumentation and the presentation of the output harder.

However, instrumentation can also take advantage of a detailed knowledge of the code paths within Hadoop to build a much deeper insight into the behaviour of the user code.

We will present our approach to general purpose instrumentation for Hadoop, which uses Hadoop-

specific insights to profile, debug and diagnose faults in a job.

We will describe techniques using attempt success/failure, internal exception rates and differential analysis, amongst others, to help us localize badly performing code or malformed input data without user intervention.

Interoperability with CMIS and Apache Chemistry

Florian Müller

16:00-16:50 | Thurs. 11/10

Content Management

Interoperability Services (CMIS)

is a specification for improving interoperability between Enterprise Content Management systems. The standard has been ratified in May 2010 and is now supported by many ECM vendors.

The number of applications using CMIS to access and manage documents of all kinds is steadily growing. Many applications use libraries provided by Apache Chemistry, which provides four implementations of CMIS: OpenCMIS (Java), cmislib (Python), phpclient (PHP) and DotCMIS (.NET).

This presentation will explain the standard, its relevance and acceptance in the ECM industry and, of course, it will present Apache Chemistry and its role in the increasing use of CMIS.

Introduction to the OODT Process Control System Operator's User Interface (OpsUI)

Andrew Hart

13:30-14:20 | Wed. 11/9

One of OODT's core strengths is the loosely connected nature of its components. This architecture allows data management infrastructures to be composed by linking the core building blocks together in various ways to support a broad spectrum of data system requirements.

Three of the core OODT components, the File Manager, Workflow Manager, and Resource Manager, along with components for crawling data repositories and extracting metadata, together represent the fundamental ingredients of many large-scale data system projects and are often referred to as a Process Control System or PCS. This session will discuss a new, unified "operator" interface for monitoring and interacting with the Process Control System.

The OODT PCS Operator User Interface, or OpsUI, provides high-level and detailed information about the status of the underlying components in an intuitive, browser-based interface. It takes advantage of a suite of RESTful web services developed in OODT 0.3 that expose component level information as JSON.

This session will provide an overview of the OpsUI, discuss the services it provides, and explain how the interface can be used to

monitor and manage a full-scale data system in real time.

Kafka - A Distributed Publish/Subscribe Messaging System

Neha Narkhede

14:00-14:50 | Fri. 11/11

Kafka is a distributed publish-subscribe messaging system aimed at providing a scalable, high-throughput, low latency solution for log aggregation and activity stream processing for LinkedIn. Built on Apache Zookeeper in Scala, Kafka aims at providing a unified stream for both real-time and offline consumption. We provide a mechanism for parallel data load into Hadoop as well as the ability to partition real-time consumption over a cluster of machines. Kafka combines the benefits of traditional log aggregators and messaging systems and has been used successfully in production for 8 months. It provides API similar to that of a messaging system and allows applications to consume log events in real-time. Written by the SNA team at LinkedIn, Kafka is open sourced under the Apache 2.0 License and preparing to be submitted as an Apache incubator project. In this presentation, we will highlight the core design principles for this system, and how this system fits into LinkedIn's data ecosystem as well as some of the products and monitoring applications it supports in our usage.



CLOUD FOUNDRY™

You Bet Your PaaS, It's Open.

THE INDUSTRY'S FIRST OPEN PLATFORM AS A SERVICE

www.cloudfoundry.com

Life in Open Source communities

Bertrand Delacretaz

16:00-16:50 | Thurs. 11/10

Open Source communities often seem to have their own unwritten rules of operation and communication, their own jargon and their own etiquette, which sometimes make them appear obscure and closed to outsiders. In this talk, we'll provide recommendations on how to get touch with, and how to join, Open Source communities. Based on ten years of experience in various Open Source projects, we will provide practical information on how to communicate effectively on mailing lists, how to formulate questions in an effective way, how to contribute in ways that add value to the project, and generally how to interact with Open Source communities in ways that are mutually beneficial. This talk will help Open Source beginners get closer to the communities that matter to them, and help more experienced community members understand how to welcome and guide newcomers.

Lucene 4.0 - Next Generation Open Source Search

Simon Willnauer

11:00 - 11:50 | Wed. 11/9

Lucene 4.0 is the next intentionally backwards incompatible release of Apache Lucene bringing a large set of fundamental API changes, performance enhancements, new features and revised algorithm. Motivated by state-of-the-art information retrieval research Lucene 4.0 exploits an entire new

low-level Codec-Layer, Automaton based inexact search, low-latency realtime-search, Column-Stride Fields and new highly-concurrent indexing capabilities. This talk will introduce Lucene's new major features, briefly explains their implementation, introduces their capabilities and several performance improvements up to 20000% compared to previous versions of Lucene.

ManifoldCF for Content Acquisition

Karl Wright

14:30-15:20 | Thurs. 11/10

I'll introduce ManifoldCF, and describe the general enterprise content acquisition and indexing problem which led to its development. I will discuss accessing multiple repositories, enforcing repository security, and incrementally keeping indexes up to date. I'll give an overview of its architecture, and demonstrate simple crawls and a secure integration with Apache Solr.

mod_lua for beginners

Eric Covener

11:30-12:20 | Thurs. 11/10

I will present a beginners guide to using the Lua scripting language inside of Apache HTTP Server, a new mod_perl-like module that will accompany the release of Apache HTTP Server 2.4.

Navigating the Apache Incubator

Brett Porter

14:30-15:20 | Thurs. 11/10

Looking to bring an open source project to the Apache Software Foundation? Already a member of a podling? Looking to get involved?

Inspired by the popular Q&A session at BarCampApache Sydney, this session will walk through all aspects of navigating the Apache Incubator, including:

- * bringing a project to Apache
- * is Apache the right home for a project?
- * the Incubator's procedures and requirements
- * when and how to graduate
- * what makes a successful Apache project
- * examples of successful and less successful podlings

NoSQL at Work with JCR and Apache Jackrabbit

Carsten Ziegeler

15:00-15:50 | Fri. 11/11

A Java content repository avoids content silos but also enables persisting of all application data ranging from small objects to audio or movie files. In contrast to relational databases the data is stored in a hierarchical way. This session enables a quickstart into JCR and demonstrates content modeling and handling content by developing a sample application.

Open Source Enables Open Innovation

Bertrand Delacretaz

16:00- 16:50 | Wed. 11/9

Open source is more than just a licence, it is also a software development methodology that allows companies to share resources and collaborate on critical parts of their software/service offerings.

Open innovation means combining internal and external ideas, and internal and external paths to market, to advance a company's technology.

The parallels should be obvious, yet people don't always think as open source as an enabler for open innovation. Open source, if done right, brings many external eyeballs and fast feedback to the software development process.

We will show how those eyeballs and feedback can make a huge difference in a company's potential for innovation, and as a result provide compelling arguments for moving large parts of your software development efforts to open source, as Day Software (now part of Adobe) started doing a few years ago.

Opening Sourcing Commercial Software - Apache Traffic Server

Bryan Call

14:30 -15:20 | Wed. 11/9

When a company decides to open source software there are a lot of details to consider. There are security concerns, software license

issues, patent ownership issues, trade mark ownership, existing software contracts, and possibility meeting the requirements of organizations like the Apache Foundation.

I will talk about how we handled each of these issues when open sourcing Traffic Server and why we chose to open source our code with the Apache Foundation. I will talk about the pros and cons of open sourcing and as a company what we did well and didn't do well.

Open sourcing was beneficial to the Traffic Server project. As a company it is important to know how to help a foster a community and benefit from the open source development process.

OSGi for Mere Mortals

Bertrand Delacretaz

09:00-09:50 | Fri. 11/11

In the last few years, OSGi has become "the" module system for Java - but is OSGi just for gurus, or are mere mortals actually able to use it?

The simple tutorial RESTful server application presented in this talk aims to demonstrate that the average Java developer can actually use OSGi, and greatly benefit from it. It is built from scratch based on a set of standard and custom OSGi services, in a simple and understandable way.

Taking advantage of a number of build plugins and runtime tools provided by the Apache Felix and Apache Sling projects allows us to write little code in our example application, while exposing the

advantages of an OSGi-based architecture in a simple and convincing way. Our walkthrough of the example application will give developers a way to get started with OSGi, without getting bogged in unnecessary details.

Out and About with Apache Traffic Server

Leif Hedstrom

17:00-17:59 | Thurs. 11/10

Apache Traffic Server is an ASF Open Source project implementing a fast, scalable and feature rich HTTP proxy and caching server. We will examine the technical details behind TS, what it is good for, and how you can configure it to accelerate your web traffic, and make complex problems easier to solve. Traffic Server was originally a commercial product from Inktomi corporation, and has been actively used inside Yahoo! for many years, as well as by many other large web sites. As of 2009, Traffic Server is an Open Source project under the Apache umbrella, and is rapidly being developed and improved on by an active community. The community is vibrant, with well over 150 active users, contributors and committers.

This talk will explain the details behind the Traffic Server technology; What makes it fast? Why is it scalable? And how is it different compared to other HTTP proxy servers? We will discuss several use cases, and show how to configure and operate TS for common tasks. Being an HTTP proxy server and cache, there are many use cases, in the areas of

forward, reverse and transparent proxying.

Traffic Server is designed using a hybrid processing model, combining an event driven engine (state machine), with a multi-threaded process approach. This allows Traffic Server to scale on modern multi-core systems, taking advantage of available CPUs. From our perspective, we've combined the best features traditionally used, solving many difficult problems and at the same time we avoid running into some of the pitfalls associated with existing solutions. This approach gives us

- * Scalability on SMP
- * Predictable and low latency characteristics
- * Lightweight on system resources (few threads, little memory wasted)
- * Efficient and reliable disk I/O

After introducing the technical details behind TS, we will discuss the common applications of a proxy and cache, when and why they would be applicable, and how to configure and use Apache Traffic Server effectively. Focusing on how to use Traffic Server in a production environment, we'll walk the audience through

- * Installation process
- * Configuration files
- * Operations and monitoring

The goal is to give a solid foundation of web proxying and caching, and why Apache Traffic Server is a contender in this space. No previous experience with Apache Traffic Server is necessary,

but familiarity in the general areas of HTTP and HTTP servers is beneficial to follow the presentation.

Provisioning Distributed OSGi Applications in a Cloud

Guillaume Nodet

15:00-15:50 | Fri. 11/11

OSGi has become a key technology for modularity and this very dynamic platform is the best choice for creating containers as proven by the migration of all JEE servers toward OSGi. However, when it comes to deploying and managing huge deployments of OSGi based applications, the tools available are quite limited. This presentation will give you an overview of a solution for provisioning and configuring distributed OSGi based applications in such environments using Apache Karaf and Apache ZooKeeper.

Rapid Application Development for Dynamic Cloud Applications

Bram de Kruijff, Marcel Offermans

09:00-09:50 | Fri. 11/11

In this presentation we would like to introduce and demonstrate Amdatu, a new open source community project that is building an application platform for dynamic composite service oriented applications in the cloud.

The Amdatu Platform consists of a set of enabling services and application layer services and can run on public, private or hybrid (cloud-burst) environments. These services provide the foundation for any application built on the platform and consist of an OSGi framework that is cluster-ready, multi-tenant

aware and can be accessed via a comprehensive set of REST based services. Applications can be dynamically assembled, monitored and managed.

The application layer services are available to any application running on the platform and include functionality like software and configuration provisioning, semantic indexing, search and recommendation, authentication and authorization, an open social implementation, extensive user and social network profiling and big data storage.

Amdatu leverages many Apache projects and hopes to join the Apache Incubator in the near future.

Realtime Big Data at Facebook with Hadoop and Hbase

Jonathan Gray

16:00-16:50 | Thurs. 11/10

Facebook has one of the largest Apache Hadoop data warehouses in the world, primarily queried through Apache Hive for offline data processing and analytics. However, the need for realtime analytics and end-user access has led to the development of several new systems built using Apache HBase. This talk will cover specific use cases and the work done at Facebook around building large scale, low latency and high throughput realtime services with Hadoop and HBase. This includes several significant contributions to existing projects as well as the release of new open source projects.

Scaling Hadoop Applications

Milind Bhandarkar

11:30-12:20 | Thurs. 11/10

Apache Hadoop makes it extremely easy to develop parallel programs based on MapReduce programming paradigm by taking care of work decomposition, distribution, assignment, communication, monitoring, and handling intermittent failures. However, developing Hadoop applications that linearly scale to hundreds, or even thousands of nodes, requires extensive understanding of Hadoop architecture and internals, in addition to hundreds of tunable configuration parameters. In this talk, I illustrate common techniques for building scalable Hadoop applications, and pitfalls to avoid. I will explain the seven major causes of sublinear scalability of parallel programs in the context of Hadoop, with real-world examples based on my experiences with hundreds of production applications at Yahoo! and elsewhere. I will conclude with a scalability checklist for Hadoop applications, and a methodical approach to identify and eliminate scalability bottlenecks.

Security Problems (and Solutions) for Service Oriented Applications

Daniel Kulp

17:00-17:59 | Thurs. 11/10

One important aspect of Service Oriented Architecture that is often challenging to address and implement is Security. Providing robust and scalable security solutions throughout highly distributed applications is a difficult problem to solve. For

traditional WebServices, standards like WS-Security, WS-Trust, WS-SecureConversation, and WS-SecurityPolicy have emerged to ease some of those problems. However, those standards don't always provide the best solutions for modern distributed applications that may include REST based services in addition to traditional SOAP based applications.

This session will cover various options for providing security to your services and will demonstrate how Apache projects such as CXF, WSS4J, Camel, and Karaf can work together to provide a complete security solution.

Serving "Web" Over IPv6

Issac Goldstand

09:00-09:50 | Fri. 11/11

Over the course of 2011, pressure has been steadily rising to move over from IPv4 to IPv6, but many developers, IT engineers and even users don't have a clue of how it all works.

This presentation will introduce IPv6 to newcomers, and explain basic setup for popular Apache technologies, including the Apache Web Server and Apache Tomcat.

Solr Flair

Erik Hatcher

16:00- 16:50 | Wed. 11/9

Show off the power of Apache Solr with state of the art user interfaces and interactions. Solr Flair demonstrates live systems leveraging Ajax suggest, "instant" search and preview, did you mean?,

spell checking, faceting, filtering, grouping, and clustering. We'll see how to generate charts, maps, and timelines from Solr indexed data. Each example will be presented with the complete code, configuration, and user interface elements.

Solr Power FTW: Powering NoSQL the World Over

Eric Pinkin

17:00 17:50 | Wed. 11/9

Solr is an open source, Lucene based search platform originally developed by CNET and used by the likes of Netflix, Yelp, and StubHub which has been rapidly growing in popularity and features during the last few years. Learn how Solr can be used as a Not Only SQL (NoSQL) database along the lines of Cassandra, Memcached, and Redis. NoSQL data stores are regularly described as non-relational, distributed, internet-scalable and are used at both Facebook and Digg. This presentation will quickly cover the fundamentals of NoSQL data stores, the basics of Lucene, and what Solr brings to the table. Following that we will dive into the technical details of making Solr your primary query engine on large scale web applications, thus relegating your traditional relational database to little more than a simple key store. Real solutions to problems like handling four billion requests per month will be presented. We'll talk about sizing and configuring the Solr instances to maintain rapid response times under heavy load. We'll show you how to change the schema on a live system with tens of millions of documents indexed while supporting real-time results.

And finally, we'll answer your questions about ways to work around the lack of transactions in Solr and how you can do all of this in a highly available solution.

Questions to be answered:

1. Why should I use Solr to relieve load from my relational database?
2. How is Solr better than the alternative NoSQL solutions already in place?
3. How do I address the pitfalls of working with Solr in large scale applications?
4. What things would be more difficult in Solr than if I had stuck with my relational database?
5. Is Solr a complete and competitive NoSQL datastore?

State of the Elephant: Hadoop Yesterday, Today and Tomorrow

Owen O'Malley

09:00-09:50 | Fri. 11/11

Apache Hadoop is rapidly gaining usage across the enterprise market and has become the primary framework for processing large datasets. It helps companies derive more value from the data that they already have and enables them to collect and analyze more data. Spreading for the early adopters in the internet sites (Yahoo, Facebook, Amazon, LinkedIn and Ebay) to a much wider audience, Hadoop is disrupting the business of analyzing data. The presentation will describe the current state of the project, lessons learned by deploying it at scale, and the roadmap for the future of the project.

Supercharging your Apache OODT Deployments with the Process Control System

Chris Mattmann

17:00 17:50 | Wed. 11/9

With the advent of OODT-139 and with recent work to build RESTful services for the Apache OODT Catalog and Archive Service (CAS) components, users now have the capability to rapidly monitor their OODT deployments and their collective health, along with perform multi-component downstream analysis, including file pedigree tracking and provenance.

This discussion will outline the importance and unique user capabilities afforded by the Process Control System (PCS) core package, along with the PCS REST services work undertaken in OODT-139. Real world Apache OODT deployments in the area of Regional Climate Modeling, Decision Support for Snow Data, Radioastronomy and in Biomedicine will be discussed, and the use of PCS will be described in an end-to-end example.

This session should be applicable for folks with a basic understand of the Apache OODT CAS components and should show how to rapidly take advantage of the exciting new PCS features making their way into Apache OODT 0.3.

Talking People into Creating Patches

Isabel Drost

11:30-12:20 | Thurs. 11/10

"Contributing to open source projects is trivial: Make a change,

create a patch, review and revise, have it accepted." When heavily involved with open source projects it's easy to forget what developers interested in contributing have to learn before even making the smallest first change.

The talk summarises some of the issues and questions students, long time developers, researchers have when faced with free software development. The talk mainly focuses on the technical issues, touching only briefly the (at least) equally large space of cultural differences of open development communities vs. corporate or even research environments.

Instead of providing pre-baked solutions to filling this gap the goal of the talk is to initiate a discussion on how to best talk your friends and colleagues into creating patches: Which strategies did work for you, which failed? Which resources do you generally use when mentoring interested peers? Where do you see most problems?

The Business of Open Source - Panel

Sally Khudairi, Debbie Moynihan, Ross Turk, Kevin Carson
11:00 - 11:50 | Wed. 11/9

The ApacheCon Business Track has expanded over the past nine conferences, addressing an array of key business, marketing and legal/licensing issues in Open Source. Our panel of influencers will answer your questions on The Business of Open Source, that includes customer requirements, application opportunities, deployment

challenges, best practices, product development, standards compliance, business model disruptors, and more. Moderator Sally Khudairi will create a lively, interactive dialogue by inviting comments from the audience throughout the session.

PANELISTS INCLUDE Debbie Moynihan of FuseSource, Ross Turk of Talend, Kevin Carson of Hewlett-Packard, and Bertrand Delacretaz of Adobe Systems.

The Other Apache Technologies your Big Data Solution Needs

Nick Burch

14:30-15:20 | Thurs. 11/10

You've gone to the talks on Hadoop / SOLR / NoSQL / etc, and now you're ready to start building your own solution on top of that! What you might not realise is that you may end up reinventing some bits of the wheel whilst building your system..

In this talk we'll take a whistle-stop tour through a number of the projects from the Apache Software Foundation, that aren't "big data" projects, but which could prove very helpful to you in building your big data solution. Knowing about these projects should help you build a better big data solution, and build it faster!

The Past, Present and Future of NOSQL

Emil Eifrem

10:00-10:50 | Thurs. 11/10

Ok, it seems like everywhere we look there's talk about next-gen

databases. Under the rallying cry of "NoSQL," several close-kept secret databases have gone open-source recently, giving everyone many exciting and confusing new opportunities. Some perform like crazy, some scale out to football field size, some just keep going no matter what, and none of them look like the SQL database we all grew up with.

In this session, Emil Eifrem will take you on a whirlwind tour of the current NoSQL landscape. He'll give a crash course overview of the four main categories of NoSQL databases, discuss what's currently lacking in NoSQL today and where it needs to go to gain real mainstream adoption.

The Power of the mod_proxy Modules

Paul Weinstein

14:30-15:20 | Thurs. 11/10

This presentation reviews the concepts of web proxies and load balancing, covers the creation and maintenance of proxies (forward and reverse) for HTTP, HTTPS and FTP using Apache and mod_proxy and how mod_proxy_balancer can be used to provide a basic load balancing solution. Configuration Examples of implementing proxies and load balancer will be discussed including; how and when mod_proxy modules can help, configuring mod_proxy for forward or reverse proxy and configuring mod_proxy_balancer for one or more backend web servers.

The Secret Life of Open Source

Ted Husted

10:00-10:50 | Thurs. 11/10

Apache, GNU, Mozilla, Ubuntu, PHP, LibreOffice, Wikipedia -- Today, there are hundreds of open source groups, each with its own culture, methodology, and governance model.

- * How are these groups alike?
- * How are they different?
- * Is there one true path to open source enlightenment, or do many paths converge around a common singularity?

Join open source insider Ted Husted as we look behind the curtain to see who's pulling strings that steer your favorite open source projects.

Turning the Knobs - Real world Java, server and application performance tuning

Filip Hanik

14:00-14:50 | Fri. 11/11

Where do you start troubleshooting a performance problem? At the network level, database queries, slow code?

Join Apache Tomcat committer Filip Hanik, Staff Systems Engineer at VMWare, in a informative session on how to tackle and resolve performance problems using a systematic approach.

Performance problems can manifest themselves at all layers of an application and infrastructure. Troubleshooting and tuning efforts are often focused on the components that an administrator or developer feel a core competency in, resulting in less than

optimal results. In this session, we will focus our efforts in detecting a problem, identifying the source of the problem and correctly solve the problem with predictable results.

After this session, you will have a better understanding of the process of identifying problems with hardware, network topologies, load balancers, web servers, Java applications and database technologies using tools available in various parts of the system as well as correctly identify the knobs to turn and how to interpret the results of various performance tuning metrics.

Using OODT @ JPL data systems

Emily Law

14:30 -15:20 | Wed. 11/9

Apache™ OODT (Object-Oriented Data Technology) is being used to support multiple science data systems at NASA's Jet Propulsion Laboratory (JPL). In the scientific data systems domain, one of the critical functions that data systems must provide is to enable users to efficiently discover, find, access, and readily utilize useful science content from NASA's increasingly large volumes of multi-mission, multi-instrument science data heterogeneously distributed among NASA's data sources at various geographic locations.

In the past, when users want to analyze or retrieve these science data, they had to leverage various custom-built tools. Using OODT that provides transparent access to distributed resources, functionality

for data discovery and query optimization, as well as distributed processing and virtual archives; JPL data systems can handle data from various sources in a uniform way.

In this talk, we will describe our experience over the past several years in leveraging and deploying OODT in our science data systems. We will outline the software engineering challenges that we faced, addressed, and along the way using OODT. We will describe several large-scale deployments of OODT, and the manner in which OODT helped us to address the data discovery and access challenges. We will also relate the lessons we have learned drawing from our experience.

Using OODT to Support Data-driven Clinical Decision Support

Andrew Hart

16:00- 16:50 | Wed. 11/9

OODT traces its roots to planetary and Earth science data systems to support research at the NASA Jet Propulsion Laboratory. Today, as an open-source project at Apache, OODT has a thriving and diverse community that includes projects in Cancer Research, Radio Astronomy, Pediatric Care, Computer Modeling, and Visualization. This session will discuss how many of the components of OODT have been used in concert to develop end-to-end data analysis infrastructures for examining large volumes of complex medical data. Motivated by, and illustrated with, concrete examples from our experience developing an infrastructure for data-driven decision support at the

Whittier Virtual Pediatric Intensive Care Unit at Childrens Hospital Los Angeles.

Topics include:

- How the OODT Product Server component can be used to bring together data from multiple upstream sources
- How the OODT File Manager component can be used to manage scalable collections of data + metadata
- How the OODT Workflow Manager component helps build composable workflows for de-identification, aggregation, and analysis
- How the OODT Web-Grid and web services can help to distribute research data to both users and applications
- How OODT Balance can help in developing purpose-built interfaces for data discovery and visualization

Using OSGi to Build Better Software: Lessons from a Telemedicine Software for Smartphones and Desktop Systems

Doreen Seider

16:00-16:50 | Fri. 11/11

OSGi brings benefits to Java applications and can even enable software to meet specified requirements. The talk will show how we applied OSGi to develop a telemedicine software for smartphones and desktop systems. This software captures vital signs from medical devices of patients via Bluetooth and sends them to medical expert centers. With this real-world example the talk will illustrate how we used dependency

injection with OSGi Declarative Services (DS) to build an easy-to-use plugin and registry mechanism. It will demonstrate how we used the modularity of OSGi to have different deployments for different platforms without rewriting all of the code or how we used loose coupling between components via services to abstract hardware layers like Bluetooth.

In this context the talk will also introduce in general the development of OSGi applications for smartphones using an OSGi stack for mobile systems and the appropriate development environment called mBS mobile. Problems we encountered during development on the different mobile systems will be described. In this context the talk will also introduce in general the development of OSGi applications for smartphones using an OSGi stack for mobile systems and the appropriate development environment called mBS mobile. Problems we encountered during development on the different mobile systems will also be described.

Using Tomcat-Native with Comet/Asynch call in Servlet 3.0

Jean Frederic Clere

11:00 - 11:50 | Wed. 11/9

Browsers and web servers are standards and the need for instantaneous data exchange has grown. AJAX for example allows web clients to communicate "asynchronously" with remote web servers.

Comet is a Tomcat feature that goes beyond AJAX and allow real asynchronous unidirectional and bi-directional connections between client and server using the HTTP protocol and Servlets.

Servlet 3.0 SPEC's are also providing asynchronous calls see what is possible do with them.

Tomcat-Native is a Tomcat sub-project that provides a non-blocking and very efficient SSL connections. Tomcat-Native relies on the APR (Apache Portable Runtime) for Socket input/output and use OpenSSL to make the cryptographic layers.

Comparison of the performance of Tomcat, Tomcat + APR and httpd.

Web Technologies in Java EE6 : Servlet 3.0 and JSF 2.0 Whats New ?

Rohit Kelapure

11:00 - 11:50 | Wed. 11/9

Servlet 3.0 is a new specification that is part of the Java EE 6 technologies. This session will introduce you to the new features of Servlet 3.0 and explain how you can leverage them in your applications. The session will focus on two major themes: ease of development, and improving application scalability. This session is intended for Java EE developers, administrators, and architects.

JSF 2.0 is a new specification that is part of the Java EE 6 technologies. This session will introduce you to the new features of JSF 2.0 and explain how you can leverage

JSF in your Java EE and Portlet applications. The session will focus on three major themes: ease of development, performance improvements, and open source technology adoption. You will learn how the tools for the Application Developer make developing UI easier, as well as the JSF-Dojo component library. This session is intended for Java EE developers, administrators, and architects.

Whirr: Open Source Cloud Services

Tom White

10:00-10:50 | Fri. 11/11

Apache Whirr is an Incubator project which provides a way to run distributed systems - such as Hadoop, HBase, Cassandra, and ZooKeeper - in the cloud. Whirr provides a simple API for starting and stopping clusters for evaluation, test, or production purposes. Whirr is cloud neutral, so services may be run on a wide variety of cloud providers (such as Amazon EC2 or Rackspace), simply by changing a configuration property. This talk explains Whirr's architecture and shows how to use it.

Wicket meets Scala and CouchDB

Bruno Borges

13:30-14:20 | Wed. 11/9

cloudera

is proud to sponsor



November 7-11 Vancouver, BC

```

      C
      JO.
      A010
      10111
      0      0110111      A
      010_ 110010001 JO1C
      )10111000001100001(
      11011110110111001
      00100011001010111
      1      0000110000101100      0
      J11(      011011100100001      J11(      A
101010110010 10010101110010C 000100100000
0010111000001. 00010110001101: A000011001010?
0110111101101110010000110110110001101111011
100100011001010111001001100001001000000100C
0000011000010110001101101000011001010100001
1111011011100100001101101100011011101110101011
^011001010111001001100001001000000100000101J
 00010110001101101000011001010100001101?
^100001101101100011011110111010101?
 01001100001001000000100000101?
      011010000110010101000011011
      0110001101111011101010110
      0010010000001000001011100
      00110010???00??01101111^
0
      )1
      )1
      01
      10
      )0(
      11(
      011
```

Afkham Azeez

Azeez is a Member of the Apache Software Foundation (ASF) and also a Project Management Committee (PMC) member, committer & long time contributor on a number of projects at ASF. He is the Director of Architecture at WSO2 Inc. His areas of interest include cloud computing, distributed computing, SOA, Highly Available & Highly Scalable applications & J2EE technologies. He is also the author of Clustering implementation for Apache Axis2. Azeez is also a co-author of the Axis2 Web Services book.

Alexander Broekhuis

Alexander works for Luminis and is a Java engineer since 2004 with a clear focus on OSGi. In the last years he has worked for embedded product companies on distributed middleware in which Java and C is used. His work includes creating meta-models and transformations from which the code for the middleware is generated (MDA, MDD), and more recently updating existing middleware to support dynamic services, based on OSGi. Alexander is initiator and committer of the Apache Celix project, a platform for distributed systems in C, with a focus on interoperability with Java (OSGi).

Alex Pinkin

Alex is a contributor to a number of open source projects including JBoss Application Server. He currently works as a lead developer on the Data Infrastructure team at Bazaarvoice. His team focuses on big data problems, analytics,

and search. Bazaarvoice is utilizing Apache Lucene and SOLR to power features which are exposed to hundreds of millions of users on thousands of web sites. Alex lives in Austin, TX where he enjoys sunny weather and outdoor life when he is not glued to a computer.

Andrew Hart

Andrew is a staff software engineer at the NASA Jet Propulsion Laboratory (JPL) where he works in the Data Management Systems and Technologies group. He brings his computer science background and creative talents to the design and development of data management components and web interfaces for a wide range of projects including an informatics infrastructure for the National Cancer Institute's Early Detection Research Network (EDRN), data modeling and architectural support for the Laura P. and Leland K. Whittier Virtual Pediatric Intensive Care Unit (Whittier VPICU) at Children's Hospital Los Angeles, and web services for JPL's Regional Climate Model Evaluation System. He is also an active contributor to the Object Oriented Data Technology (OODT) project which last year became a TLP at Apache.

Ate Douma

Ate is Chief Architect at Open Source WCM vendor Hippo in The Netherlands. He is a member of the Apache Software Foundation and PMC member and committer for Apache Portals, Apache Wicket and the Apache Incubator. He is also mentoring the Wookie and Rave

think

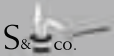
“...before you send that flame.”

“...about Sponsoring Apache.”

“...on how community should work.”

“...about the other person’s shoes.”

“Those who would give up essential Liberty, to purchase a little temporary Safety, deserve neither Liberty nor Safety.” — Benjamin Franklin, 1755



podlings through the Incubator process, and one of the initiators and active participants of the Apache Rave project. Ate served as Expert Group member of the Java Community Process (JCP) for JSR-286: Portlet Specification 2.0 and JSR 301: Portlet Bridge Specification for JavaServer Faces.

Bertrand Delacretaz

Bertrand works as a Senior Developer in the Enterprise R&D team of Adobe Systems Inc (www.day.com), using open source tools to create world-class content management systems and frameworks. Bertrand is an active member and current director of the Apache Software Foundation, involved in a number of Apache projects as a committer, PMC member and incubation mentor.

Bertrand Delacretaz

Bertrand works as a Senior Developer in the Enterprise R&D team of Adobe Systems Inc (www.day.com), using open source tools to create world-class content management systems and frameworks. Bertrand is an active member and current director of the Apache Software Foundation, involved in a number of Apache projects as a committer, PMC member and incubation mentor.

Bram de Kruijff

Active on the web since 1995, Bram de Kruijff is an open source enthusiast responsible for Research and Development at GX Software inc. As a software architect he has a broad experience with server side web application architectures, frameworks and technologies for enterprise grade applications. Since 2005 he has been actively involved in several commercial products and projects that successfully adopted OSGi as a server side application architecture, leveraging open source software including many several ASF projects. Bram is a PMC member and committer of the Amdatu project.

Brett Porter

Brett is an open source software developer from Sydney, Australia with a passion for development tooling, and automation. He discovered an early beta of Maven 1.0 in 2003, and has been heavily involved in the development of the project since. A member of the Apache Maven Project Management Committee, he has conducted

presentations and training on Maven and related tooling at several conferences and events. He founded what became the Archiva project in 2005, and is the co-author of *Apache Maven 2: Effective Implementation* (2009) and *Better Builds with Maven* (2005).

Bruce Snyder

Bruce is a veteran of enterprise software development and open source software. With over 15 years of experience, Bruce has fulfilled roles around not only software research and development but also software consulting using a wide range of technologies including Java EE, enterprise messaging and integration, service-oriented architecture and event-driven architecture. Bruce is a senior software engineer at SpringSource/VMware, a member of the Apache Software Foundation and a committer on numerous Apache projects. He is not only a published author of books on ActiveMQ, the Spring Framework, Maven and Geronimo, he is also a recognized international speaker at software conferences around the world and has helped to build communities around open source software. Bruce lives in beautiful Boulder, Colorado with his family where he enjoys cycling, hiking, running and anything outdoors.

Bruno Borges

Bryan Call

Bryan has been writing code and working with on large scale solutions for 14 years. He has

experience optimizing and profiling projects, including Apache Traffic Server and many internal projects at Yahoo!. He came to Yahoo! through an acquisition of a startup and have been working at Yahoo! for the last 12 years. He has worked on various products and teams, such as WebRing, GeoCities, People Search, Yahoo! Personal, Tiger Team (internal consulting team), Architect in the Platform Group, and is now an Architect in the Edge Platform group. Bryan is also a committer on the Apache Traffic Server project and instrumental in bring Traffic Server to the Apache Foundation.

Carl Hall

Carl is a co-owner of Hallway Technologies; a software consulting firm focused on developing and supporting the Sakai Open Academic Environment. He spends most of his waking hours writing software with the occasional cycling jaunt. He is a committer on Sakai and Apache Sling.

Carlos Sanchez

Carlos is specialized in solving business challenges in a wide variety of industries, focusing on automation and quality of software development, QA and operations processes. Involved in Open Source for nearly ten years, he is a member of the Apache Software Foundation and project leader at the Eclipse Foundation amongst other open source groups. Carlos contributes to several projects, including Apache Maven, Continuum and Archiva, and is a co-author of "Better Builds with Maven," the first book about Maven

2. Currently works as Architect at MaestroDev, a company focusing on development and DevOps tools, from his home in Spain.

Carsten Ziegeler

Carsten is senior developer for JEE and portal applications at Adobe. He is a member of the Apache Software Foundation and has participated for more than twenty years in several open source projects. Carsten is a member of several Apache communities and Project Management Committees including Sling, Felix, and Portals.

Chris Hostetter

Chris is a Member of the Apache Software Foundation, and serves on the Lucene Project Management Committee. Prior to joining Lucid Imagination in 2010 to work full time on Solr development, he spent 11 years as a Principal Software Engineer for CNET Networks thinking about searching "structured data" that was never as structured as it should have been.

Chris Mattmann

Chris is a senior computer scientist at the NASA Jet Propulsion Laboratory, working on instrument and science data systems on earth science missions and informatics tasks. He's also an adjunct assistant professor in the Computer Science Department at the University of Southern California. His research interests are primarily software architecture and large-scale data-intensive systems. Mattmann received his PhD in computer

science from the University of Southern California. He's a senior member of IEEE. Contact him at mattmann@apache.org.

Daniel Kulp

Daniel is the PMC Chair of the Apache CXF project as well as a PMC Member of Apache Aries, Apache Maven, and Apache WebServices. He attended Northeastern University in Boston where he received degrees in Chemical Engineering and Computer Science. As the VP of Open Source Development for the Application Integration Division at Talend, Dan gets to practice his passion for coding open source at work, and still has time to dedicate to his loving family.

David Blevins

David is co-founder of the Apache OpenEJB project, a founder of Apache Geronimo, contributor to Apache OpenWebBeans and many other Open Source Java EE related projects for over 10 years. David was an active member of the EJB 3.0 (JSR 220), EJB 3.1 (JSR 318) and Java EE 6 (JSR 316) Expert Groups, and contributing author to Component-Based Software Engineering: Putting the Pieces Together from Addison Wesley. He can be found speaking on these topics at JavaOne, ApacheCon, O'Reilly Open Source Convention, JAX London and other conferences.

Debbie Moynihan

Debbie, who has been with the team since it was part of IONA,

was instrumental in the LogicBlaze purchase and the creation of what is now FuseSource Corporation. In addition to traditional marketing responsibilities, Debbie is also responsible for community development and is a frequent participant in open source events. Her tenure at FuseSource follows a long and successful career at IBM where she led global product management and product marketing teams for WebSphere Business Integration and WebSphere Commerce. Prior to IBM, Debbie was the Director of Marketing and Technical Sales for OpenOrders, a start-up company that was acquired by IBM.

Doreen Seider

Doreen is the deputy head of the Department for Distributed Systems and Component Software of the German Aerospace Center's (DLR) Simulation and Software Technology division (<http://www.dlr.de/sc>). She leads the development of the OSGi-based open source software integration framework RCE (Remote Component Environment). In 2010 she worked several month at the Argonne National Laboratory on the Web services platform Crux. In the last years she talked at different conferences like Super Computing, JAX London, ApacheCon about topics concerning modularization with OSGi or Open Source software in aerospace fields.

Emil Eifrem

Emil is the founder of the Neo4j open source graph database project, which is the most widely

deployed graph database in the world. As a life-long compulsive programmer who started his first free software project in 1994, Emil has with horror witnessed his recent degradation into a VC-backed powerpoint engineer. As the CEO of Neo4j's commercial sponsor Neo Technology Emil is now mainly focused on spreading the word about the powers of graphs and preaching the demise of tabular solutions everywhere. Emil presents regularly at conferences such as JAOO, JavaOne, QCon, and OSCON.

Emily Law

Emily has over twenty years of experience in the research, development and management of complex information systems. Ms. Law has worked for both large and small organizations having served on a diverse set of projects during her career. Since 1996, Ms. Law has been working at NASA's Jet Propulsion Laboratory where she has provided leadership and management in the architecture, development and operations of highly distributed ground data systems for planetary exploration and earth science. She currently serves as a deputy program manager and development manager to two separate directorates covering data systems in solar system research and earth science. In 2005, she was the recipient of the NASA Exceptional Achievement Medal in recognition of outstanding contributions to the NASA Deep Space Network. She currently leads operations for NASA's Planetary Data System, NASA's official archive to manage results from solar system

research; and leads development of the Lunar Mapping and Modeling Project data management infrastructure and portal in support of Lunar Exploration activities. Ms. Law has B.S. degrees in Math and Computer Science from Cal State Long Beach and is a graduate of University of Southern California where she holds a M.S. degree in Computer Engineering.

Eric Covener

Eric is an Apache HTTP Server committer, Apache Software Foundation member, and frequent IRC and users list supporter.

Erik Hatcher

Erik is an Apache Software Foundation member, and an active committer on the Lucene and Solr projects. Erik has co-authored the award-winning book *Java Development with Ant* (Manning) and the well reviewed *Lucene in Action* (Manning). Erik has spoken frequently at industry conferences, including JavaOne, ApacheCon, OSCON, and the No Fluff, Just Stuff symposium circuit. Erik co-founded Lucid Imagination, a full-service Lucene/Solr support company.

Filip Hanik

Filip is a Senior Software Engineer for SpringSource, a division of VMware, and a key participant in SpringSource's Apache Tomcat initiatives. Filip brings 15 years of extensive experience in architecture, design and development of distributed application frameworks and containers and is recognized for

his top-quality system development skills and continuous participation of Open Source development projects. Filip is a committer to the Apache Tomcat project where he is a leading authority on Tomcat clustering and a key contributor to the core of the platform. He also authored the online book *Hitch-Hiker's Guide to Tomcat*. Prior to joining Covalent, Filip worked as a Senior Software Engineer/Architect for La Quinta Corporation. Previously, Filip has made contributions to software initiatives for Walmart.com, Sony Music, France Telecom and has held a variety of senior software engineering positions with technology companies in both the United States and Sweden. He received his education at Chalmers University of Technology in Gothenburg, Sweden where he majored in Computer Science and Computer Engineering.

Florian Müller

Software Architect at Alfresco
- Member of the OASIS CMIS Technical Committee and CMIS specification editor - Vice President Apache Chemistry

Grant Ingersoll

Grant is a long time Lucene and Solr committer as well as the co-founder of the Apache Mahout project and of Lucid Imagination, a company dedicated to supporting Lucene and Solr technologies. Grant has spent the better part of his career working on search and natural language processing problems.

talend*

*open integration solutions

**Proud to
Support The
Apache Software
Foundation**

www.talend.com

Guillaume Nodet

Guillaume is Software Fellow at FuseSource, a Progress Software subsidiary offering Service Oriented Architecture (SOA) products based on leading open source projects hosted by the Apache Software Foundation. Guillaume has extensive experience within the open source community and is a highly regarded committer across several projects, including Apache ServiceMix, Apache ActiveMQ, Apache Camel. He is a member and VP Apache Karaf at the Apache Software Foundation. Guillaume is also involved in the OSGi Alliance expert groups. Prior to joining Progress Guillaume was Senior Architect at LogicBlaze, a leading provider of open source solutions for SOA and business integration which was acquired by IONA in April 2007, which was in turn acquired by Progress in September 2008. At LogicBlaze and IONA Guillaume was responsible for leading key development contributions in particular he was lead developer of the Apache ServiceMix project. Guillaume has also held Java Architect and software developer roles at Delia Systems, headquartered in Paris, France and European consulting organization, Altran Technologies. Based in Normandy, France, Guillaume holds an Engineering degree in Computer Science from Ecole supérieure d'Electricité.

Hadrian Zbarcea

Hadrian is the PMC Chair of the Apache Camel project and a PMC Member of Apache ActiveMQ.

With over 15 years of experience in system integration, he has implemented countless enterprise scale deployments. Hadrian is currently a Principal Architect at Talend devoting most of his time to open source. He enjoys quality time with his family, training in taekwon-do and competing in chess tournaments.

Isabel Drost

Isabel co-founded the Lucene sub-project Apache Mahout. She is employed at Nokia Berlin. After an internship at Google in 2005/06 Isabel joined Neofonie GmbH. She worked as research assistant in Berlin. Isabel holds a master's degree in computer science from the University of Applied Sciences Mittweida.

Issac Goldstand

Issac has been involved in the Web community for over 13 years. Having been fascinated by the growing web since the early 90s, he decided to learn how it all worked. Starting with basic HTML in 1997, he moved to working on some HTTP-related IETF standards in 1999-2000. Around then he started to pick up Perl to use in basic CGI scripts. From there, he quickly advanced to mod_perl, hacking on the Apache HTTP server and writing finely-tuned web applications. Today, he is 30, happily married with 2 boys, and somehow found himself caught in the crux between online marketing and internet technologies.

Jake Luciani

Jake is a lead engineer at DataStax focusing on Cassandra development its integrations with other popular apache software such as Hadoop, Hive and Solr. He is a committer on the Cassandra and Thrift projects.

Jean Frederic Clere

Jean-Frederic has spent more than 20 years writing client/server software. He is committer in APR, Jakarta, Httpd and Tomcat and he likes complex projects where different languages and machines are involved. Borne in France, Jean-Frederic lived in Barcelona (Spain) for 14 years. Since May 2006 he lives in Neuchatel (Switzerland) where he works for RedHat in the JBoss division.

Jean-Baptiste Onofré

Jean-Baptiste is PMC on Apache ServiceMix and Apache Karaf. He's also commiter on Apache ACE and contributes to several projects such as Apache Camel, Apache CXF, Apache ActiveMQ. Apache ServiceMix PMC. Apache Karaf PMC. Apache ACE Committer.

Jim Jagielski

Web/Open Source visionary and advocate Jim Jagielski currently serves as Sr Consulting Software Engineer for Red Hat, Inc.. It was his association with Apple's first effort with Unix, A/UX, that started his guru reputation on the 'Net. Since then, Jagielski has made developer and porting contributions to numerous Open Source projects, including Perl, sendmail, BIND/

named, PHP, xntpd, GNU, cgiwrap and FreeBSD. He's also written a lot of independent code but is likely best known as a long-time core developer of Apache httpd and other ASF projects. Jagielski also serves on the Board of Directors of the Apache Software Foundation (ASF), which he co-founded and holds the position of President.

Jonathan Gray

Jonathan is a long-time Apache HBase committer and user of Apache Hadoop and other related projects. As a software engineer and open source advocate at Facebook, Jonathan works across product teams on data infrastructure and the usage, contribution, and release of open source software.

Karl Wright

I am the original developer and a principal committer of the Apache ManifoldCF project, a project undergoing incubation at ASF. I have extensive experience in content management, speech recognition, and compiler development, and I am the author of Borland's Turbo Assembler. I am also the author of the book ManifoldCF in Action, available from Manning Publishing. I hold computer science degrees from both M.I.T. and Stanford University, and I currently work for Nokia, Inc.

Kevin Carson

Kevin is a Solution Architect for HP's pre-sales support organization, the Americas lead for the HP internal Open Source and Linux Profession,

and an HP sales organization Linux Advocate. He continues to reside in the greater Vancouver area after graduating from the University of British Columbia with a BSc in Computer Science. Kevin's 24 years as a computing professional working on various military prototypes, weather forecasting prototypes, and pre-sales support has included 19 years of working with Linux.

Kevan Miller

Kevan is a member of the ASF and current PMC chair of the Apache Geronimo project. In addition to Geronimo, Kevan is also a participant in the OpenEJB, OpenWebBeans, and multiple Incubator Podling communities.

Konstantin Shvachko

Konstantin is a veteran Hadoop developer. He is a principal Hadoop architect at eBay. Konstantin specializes in efficient data structures and algorithms for large-scale distributed storage systems. He is a member of the Apache Hadoop PMC. Ari Flink Ari Flink is an Operations Infrastructure Architect at eBay Timothy Coulter Tim Coulter is a specialist consultant in the areas of Enterprise Networking, Operations, and Infrastructure.

Leif Hedstrom

Leif is a software architect at Akamai Technologies, where among other things, he implements applications using the Apache Traffic Server software. Before joining Akamai, Leif designed and

implemented several CDN services for Yahoo!, used by sites such as Y! News and Flickr. Currently the chair person of the Apache Traffic Server program committee, he's actively involved with development and evangelism of the project. Prior to Yahoo!, Leif has worked at several Internet companies such as Netscape Communications and Infoseek Corporation. Leif is an avid scuba diver, dirt biker, alpine skier, dog person and family man. And of course, he's a huge computer nerd! I'm a long time Unix / C / C++ / Network software architect and developer. I'm currently the V.P. of Apache Traffic Server, and I'm also a Member of the Apache Software Foundation.

Marcel Offermans

Marcel is a fellow at Luminis and leads the competence center for OSGi at Luminis Technologies. As a software architect he has a broad experience in different operating systems, languages and applications, ranging from enterprise applications to mobile, embedded and distributed systems. He has been involved in lots of commercial OSGi projects for over eight years and is responsible for the architecture, development and support of an OSGi based software provisioning product. He is also a member of the Apache Software Foundation and involved as a committer and PMC member in the Apache Felix, Apache ACE and Apache Celix projects.

Milind Bhandarkar

Dr. Bhandarkar was the founding

member of the team at Yahoo! that took Apache Hadoop from 20-node prototype to datacenter-scale production system, and has been contributing and working with Hadoop since version 0.1.0. He started the Yahoo! Grid solutions team focused on training, consulting, and supporting hundreds of new migrants to Hadoop. Parallel programming languages and paradigms has been his area of focus for over 20 years. He worked at the Center for Development of Advanced Computing (C-DAC), National Center for Supercomputing Applications (NCSA), Center for Simulation of Advanced Rockets, Siebel Systems, Pathscale Inc. (acquired by QLogic), and Yahoo!. Currently, he works on distributed data systems at LinkedIn Corp. Parallel programmer, data-intensive supercomputing.

Neha Narkhede

Neha is a Senior Software Engineer in the Search, Network and Analytics Team at LinkedIn, working on Project Kafka. In the past she's worked on search systems in large scale databases and has been an active contributor to several projects LinkedIn has open sourced, including Bobo and Zoie. Search, Distributed Systems, Stream data processing, LinkedIn

Nick Burch

Nick is the PMC chair for the POI project, as well as a committer on Tika. He works as a senior developer at Alfresco, a leading open source Enterprise Content Management vendor. He is a big

fan of many aspects of agile and extreme development, and rapid web development frameworks such as Django. Nick is an ASF member, Current interests include agile and extreme development, Django and python, mobile development and open geodata.

Olivier Grisel

R&D Software Engineer at Nuxeo with a background in Machine Learning and Natural Language Processing. Works on text analytics and semantic knowledge extraction using OpenNLP, Mahout and Pig and contribute to the Stanbol project.

Owen O'Malley

Owen is a Software Architect in Yahoo's Hadoop development team and is an Apache member and member and original chair of Hadoop's Program Management Committee. He has been a Hadoop committer since March of 2006 and more than 200 of his patches have been committed to Hadoop. He was the technical lead on adding security in to Hadoop. Before working on Hadoop, he worked on Yahoo Search's Webmap that builds and analyzes the graph of the World Wide Web. Prior to Yahoo, he worked at NASA Ames Research Center on software model checking and at Sun working on a distributed version control system. He received his PhD in Software Engineering from University of California, Irvine.

Paolo Castagna

I am a committer for the Apache Jena project, currently in incubation.

I am a software engineer at Talis Systems Ltd. Previously, I worked at HP Labs where I begun my sorcerer's apprenticeship on Jena. At Talis I work in the team who develops the Talis Platform to host and publish RDF data via REST APIs, as Linked Data and SPARQL endpoints.

Paul Weinstein

With the popularization of the Internet, Paul has devoted his energies to developing and integrating web-based systems. In doing so he has become knowledgeable in the workings of many tools of the trade; Linux, FreeBSD, Apache, Perl, PHP, PostgreSQL and MySQL among others. Currently Paul resides in Chicago and works at Orbit Media Studios, Inc. as a Web Programmer Analyst.

Phil Steitz

Phil is an independent technology and management consultant. He was formerly CIO at Barclaycard US and CTO at American Express. Phil is an ASF Member who works on Apache Commons.

Philip M. Gollucci

VP Apache Infrastructure, root@mod_perl, httpd, apr committer FreeBSD ports committer I spend my weekends rebuilding my flood damaged house after which I'll be getting a labrador retriever.

Rohit Kelapure

I am responsible for the architecture

and delivery of caching and data replication function in WebSphere Application Server. I advise customers on caching and replication strategy for their deployments. I am a member of key feature teams delivering JEE6 function in WebSphere Application Server. As a committer on the Apache Open Web Beans project I am particularly involved in the the new JSR299 - CDI, Java Contexts and Dependency Injection JEE6 specification and its peer technologies like Weld, Seam & Spring. I am often called upon to resolve critical situations in production at WebSphere Application Server and our stack products WebSphere Portal, WebSphere Commerce & WebSphere eXtreme Scale. I have an avid interest in java debugging tools and techniques that make it easier to understand complex problems. Customer advocacy for large banks and retailers as well as WebSphere User Group presentations. Specialties: CDI, Seam, Spring, Weld, Apache Open Web Beans, Apache Open JPA, JPA L2 Caching, Dynamic and Static Caching, WebSphere Application Server, WebSphere Portal Server, WebSphere Commerce Server, WebSphere eXtreme Scale, JEE6 specification, Application Server debugging tools like Eclipse Memory Analyzer, Heap & Thread Analyzer, GC analysis tools.

Ross Gardler

I'm just starting a new chapter in my working life. By the time ApacheCon comes round it should be well underway. Looking

forward to then next ten years of my personal involvement with the ASF. Ross Gardler is a committer and PMC member on a number of Apache projects, a champion and mentor on incubating projects and a PMC member of the Community Development project. Ross is a consultant and advisor to a large number of open source software projects, companies and users. With a foot in both the academic research and commercial sector Ross specialises in identifying research outputs that are almost ready for commercial exploitation. Once identified he attempts to pair the academic team with appropriate commercial teams to bring the technologies to market. Ross is chair of the TransferSummit conference, which seeks to link the academic research sector with the commercial sector, and speaks at a great many open source related events.

Ross Turk

Ross is the Senior Director of Communities for Talend, responsible for building a strategic relationship with users, contributors, and open source developers. Ross brings more than 15 years of experience creating software, managing complex IT systems, and helping companies understand and serve developers. Before joining Talend, Ross managed developer outreach for Alcatel-Lucent and SourceForge.net, the net's largest open source community.

Sally Khudairi

Active in the Web since 1993, Sally Khudairi is the publicist

behind some of the industry's most prominent standards and organizations. The former deputy to Sir Tim Berners-Lee and long-time champion of collaborative innovation, she became involved with The Apache Software Foundation (ASF) in 1999, and was elected its first female and non-technical member. Sally is chief executive of luxury brand communications consultancy HALO Worldwide, vice president of ASF marketing and publicity, co-chair of ApacheCon, and advisor/co-producer of TransferSummit.

Sander Temme

Sander Temme: Sander Temme is a Sales Engineer for Thales E-Security, a security and cryptography company whose clients include Fortune 500 companies, financial services companies and government agencies. He is a member of the Apache Software Foundation and is active in the httpd, Infrastructure and Gump projects. Sander has spoken at numerous conferences like ApacheCon, CodeCon, LinuxWorld, and chapter meetings and conferences of the Information Systems Security Association (ISSA). Sander is owned by Murphy, the wonder cat.

Shannon Quinn

Shannon Quinn: Shannon Quinn received his BS in computer science from the Georgia Institute of Technology in 2008, and his MS degree in computational biology from Carnegie Mellon in 2010. He is currently finishing his first year as a PhD student in the Joint Carnegie

Mellon-University of Pittsburgh PhD program for computational biology. His research focus involves bringing bioimage informatics and distributed computing to analyze massive datas in a clinical setting.

Shevek

Shevek is an expert Java programmer, and one part of the creative team behind Karmasphere, a San Francisco based big data analytics company. He has worked on cutting edge research in compilers and language design, algorithmic optimization, systems and security. He received a Doctorate in Computing from the University of Bath, England. He also holds a Masters in Pure Mathematics and an epee.

Simon Willnauer

Simon is a Lucene core committer and PMC member. During the last couple of years he worked on design and implementation of scalable software systems and search infrastructure. He studied Computer Science at the University of Applied Sciene Berlin. Currently, he works as a consultant for Apache Solr, Lucene Java and Hadoop and is a co-organizer of the "BerlinBuzzwords" conference on Scalability in Berlin.

Stephan Klevenz

Stephan is software architect at SAP and has more than 10 years of experience in large scale software solutions in the area of enterprise content management. He is a member of the OASIS CMIS

Technical Committee and active committer to the Apache Chemistry project.

Ted Husted

Ted is an author, consultant, and speaker. His books include Google Wave (Preview) Explained, JUnit in Action, Struts in Action, and Professional JSP Site Design. Ted has consulted with teams throughout the United States, including CitiGroup, Nationwide Insurance, and Pepsi Bottling Group, and he is a regular speaker at ApacheCon and the Ajax Experience. Ted is also a former member of the Apache Struts project and co-founder of the Apache (Jakarta) Commons. You can follow him on Twitter and on several LinkedIn Groups.

Tim O'Brien

Tim has been writing documentation for Apache-related projects for a greater part of a decade. He started with the Jakarta Commons (now Apache Commons) project and wrote the Jakarta Commons Cookbook for O'Reilly. His focus over the past few years has been to support the community outreach, Maven, and Nexus documentation efforts at Sonatype. Tim used to contribute heavily to Commons back in the days of Jakarta, and he desperately wants to get back to coding and contributing to ASF projects. He's continuously trying to convince people that he's a coder, not a technical documentation expert. In addition to open source work and documentation efforts, Tim has implemented systems for

Forbes, Inc., Consumer Reports, and TheStreet.com over the past decade and he has occasionally pretended to be a journalist for O'Reilly Media covering science, government, and technology.

Tom White

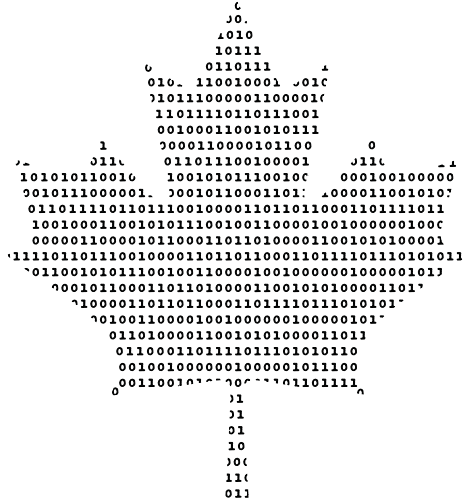
Tom has been an Apache Hadoop committer since February 2007, and is a member of the Apache Software Foundation. He started the Apache Whirr project in May 2010. Tom is the author of "Hadoop: The Definitive Guide", published by O'Reilly Media. He works for Cloudera, a company set up to offer Hadoop support and training. Previously he was an independent Hadoop consultant, working with companies to set up, use, and extend Hadoop. Tom has a Bachelor's degree in Mathematics from the University of Cambridge and a Master's in Philosophy of Science from the University of Leeds, UK.

ApacheCon 2011



ON **ALL BOOKS AT MANNING.COM**

USE PROMOTIONAL CODE:
apache3911



Thank You to Our Track Sponsors

Data Handling & Analytics:



Buisness & Community:



Servers:



Enterprise Java:



Open Source
Enterprise

Custom **E-commerce** &
ERP Platform Consulting



Built on Apache Open For Business

Bring open source front-and-center in your enterprise by partnering with the global leader in Apache Open For Business. From business analysis through design and implementation, HotWax Media is with you every step of the way.

HotWax Media

(877) 736-4080 www.hotwaxmedia.com

Since 1997, delivering leading edge e-commerce, creative services and enterprise automation.