STAC™ BENCHMARK COUNCIL

STAC® Summit

October 30, 2014
Doors open: 9:30am
Meeting starts: 10:00am

Hyatt Andaz Hotel
Great Eastern Room
40 Liverpool Street
London

Platinum Sponsor:

Gold Sponsors:

AGENDA

Note: Times are approximate.

Morning welcome [slides]

- *Peter Lankford, Founder & Director, STAC*

The role of innovation in financial IT today

- *Neil Horlock, Director, Credit Suisse*
- *Barry Childe, Head of GM Research & Innovation, HSBC*
- *Rashmi Tank, MD, eFICC, Barclays*

Banks today need to deliver new functionality more quickly despite continued cost constraints. This runs counter to the old saying: 'Good, cheap, and fast: pick any two.' But the tradeoff curves between functionality, cost, and time to market are not static. Shifting those curves requires innovation. In this panel, we'll ask senior bank technologists what innovation means in the context of their businesses. What kinds of technology have disruptive potential today? What kinds of innovation do firms seek from vendors? What kinds of innovation do they expect from technologists within the bank? How can financial firms both promote innovation internally and project that image externally to attract new talent?

STAC update on STAC-A2 (risk computation) [slides]

- *Peter Lankford, Founder & Director, STAC*

The STAC-A2 Benchmark suite is the industry standard for testing technology stacks used for compute-intensive analytic workloads involved in pricing and risk management. Peter will provide a brief update on the latest learnings and the latest activities of the STAC-A2 Working Group.

Key decisions when going parallel [slides]

- *Jim Cownie, Principal Engineer, Intel*

By now, most financial firms know that the way to harness Moore's Law (which continues, by the way) is to parallelize their applications. But that's easier said than done. Huge amounts of existing code are still serial or not parallelized well. What are some of the strategies a firm can take to deciding which applications and libraries to parallelize first? Once the decision is made, how can a developer decide what approach to take to parallelize a given piece of code, such as the programming model, compiler, and tools? Jim will propose answers to these questions.

rCUDA: Virtualizing GPUs to reduce cost and improve performance [slides]

- *Federico Silla, Associate Professor, Technical University of Valencia*

The use of GPUs for quantitative finance is well established, but GPU-enabling an additional compute node requires justifying the additional cost (power/cooling, space, maintenance, and acquisition costs). Power is especially important, as GPUs consume significant energy even when idle. According to Federico, such cost justification suffers from the fact that GPUs often have a relatively low utilization rate, given the level of data concurrency in most applications. In his view, enabling GPU-equipped nodes to provide GPU services to all the nodes in the cluster--i.e., virtualizing the GPUs--increases utilization and improves the cost-benefit tradeoffs significantly. In this talk, Federico will introduce rCUDA, a remote GPU virtualization framework that enables existing CUDA applications to use virtualized pools of GPUs. According to Federico, rCUDA not only increases jobs executed per time in a cluster, but it can also improve per-job performance by enabling a single application to exploit all of the GPUs in the cluster (up to 64 have been tested).

Networking Luncheon

STAC update on low-latency research [slides]

- *Peter Lankford, Founder & Director, STAC*

Peter will review the latest STAC activities related to low-latency workloads.

Innovation Roundup Presentations

"The Business of Scaling" [slides]	*Srini Srinivasan, Founder & VP Engineering & Ops, Aerospike*
"Experiences at Layer 1" [slides]	*Dave Snowdon, Founder, co-CTO, Metamako*
"Risk Management Building-in Technique" [slides]	*Danil Baburin, Head of QUIK Development, ARQA Technologies*
"Introducing the ExaLINK Fusion: a new ultra-low-latency switch and application platform" [slides]	*Dr Matthew Chapman, CTO, Exablaze*
"Low Latency applications for FPGA" [slides]	*Steinn Gustafsson, Director, Chevin Technologies*

Rethinking network capture

- *Matthew Knight, Marketing Director Financial Services, Solarflare* **[slides]**
- *Glenn Wright, Systems Architect, DataDirect Networks* **[slides]**

Financial firms today rely on network captures for a range of crucial analytics. But capturing data remains a fairly expensive and rigid process, handled by specialty cards and appliances. How well does this model fit today's customer needs? Would a more open-systems approach be superior--or just more complicated? If a firm takes over the responsibility to engineer its own capture solutions, how should it handle some of the major considerations (e.g., constructing a storage architecture with sufficient capacity and the ability to capture during traffic bursts, what form to capture data in, how to accommodate realtime analytic needs, how to satisfy compliance requirements, etc.)?

Achieving predictable low latency in shared networks [slides]

- *Matthew Grosvenor, Doctoral Researcher, Cambridge University*

For years, trading organizations have had to create and manage separate networks for latency-sensitive workflows. That is because in a shared environment, interference from throughput-intensive applications can cause queuing that delays traffic between latency-sensitive applications. Organizations could improve agility and decrease costs if they could accommodate low-latency workflows on networks that carry other data center traffic. Matthew and his colleagues at Cambridge have developed a technology that they claim makes this possible. It is a new Linux Traffic Control module that is deployed on latency-sensitive nodes without requiring modification of network switches. According to Matthew, this approach bounds latency and radically reduces in-network interference, outperforming Ethernet Flow Control (802.3x), ECN (WRED) and DCTCP while improving average flow completion times to levels rivaling (or beating) DCTCP and pFabric. In this talk, Matthew will explain how the technique works and how it can be deployed today.

From art to science: Accelerating front, middle, and back office apps via thread pinning [slides]

- *Leo Martins, Managing Director, Pontus Networks*

Deploying multi-threaded applications (and/or multiple processes) on systems with lots of cores and several interfaces presents a challenge: performance can be very sensitive to whether and how threads and processes are pinned to cores. Engineering groups tend either to invest a great deal of effort to determine the optimal thread pinning for an app, or they do very little pinning at all. Taking a variety of application types as examples, Leo will explain what thread pinning is, how it affects OS schedulers, and how it can increase performance, reduce hardware costs, and reduce operational risk. He will discuss the pros and cons of traditional approaches to pinning and provide a point of view on better approaches.

Coffee Break

STAC update on tick analytics and backtesting [slides]

- *Peter Lankford, Founder & Director, STAC*

Peter will summarize the latest activities in areas relating to backtesting and tick data.

Innovation Roundup Presentations

"Solving the latest challenges in Regulatory Reporting" [slides]	*Daniel Roberts, Principal Sales Engineer, MarkLogic*
"Capitalizing on Effective Data Mining: Methods for Big Data and Complex Problems" [slides]	*Steve Howard, Sales Engineer, Rogue Wave Software, Inc*
"Can SQL handle tick data with speed?" [slides]	*Ian Hillier-Brook, CEO, MCO Europe Ltd.*

It's About Time: Better ways to manage financial time series data [slides]

- *Patrick Callaghan, Solutions Architect, DataStax*

Statistics and modeling on time series data are essential components of mathematical finance for trading and risk management. But in Patrick's view, traditional approaches to storing and accessing time series fall short of today's requirements. In this talk, Patrick will lay out the principles of a successful time series architecture and how some trading houses are using those principles to enable innovative data management strategies that provide a competitive advantage. Along the way, he will discuss strategies for optimising real-time distribution of time series across multiple data centers as well as techniques for short- and long-term data retention.

Apache Samza & Kafka - LinkedIn's stream processing tools [slides]

- *Martin Kleppmann, Apache Samza committer, former LinkedIn Engineer*

Only a handful of industries used to be concerned with event-stream processing, such as defense, sensor-driven manufacturing, and of course capital markets. Today, stream processing is a topic in many more industries, from retailing to utilities to social media. And as with so many data-intensive problems today, web companies are creating and open sourcing a large amount of code to handle them. Retail banking and wealth management technologists are considering these open source tools to deal with whole new classes of problems, while some trading firms are starting to use them for old classes of problems. Apache Kafka and Apache Samza are two big data technologies open sourced by LinkedIn. Kafka is a publish-subscribe message bus designed for high throughput and reliability. Samza builds on Kafka and Hadoop to provide high-throughput, stateful stream processing across a cluster (joining, filtering, transforming, etc.). In this talk, Martin will introduce the motivation and architecture of these systems and explore their uses and limitations.

New approaches to realtime analytics

- *Martin Kleppmann (see above)*
- *Herman Fick, CEO, fraXses* **[slides]**
- *Patrick Callaghan, Solutions Architect, DataStax*
- *James Blackburn, Trading Systems Developer, Man Group (AHL)*

The previous presentation was just one example of the focus that parts of Silicon Valley have on analyzing data in real time and sharing the enabling software with the rest of the world. This phenomenon raises a number of questions. Are the new open source products enterprise ready? How do they compare to proprietary products already in the market, in terms of functionality and performance? How does development using these open source frameworks compare to development using proprietary products? On the flip side, what are proprietary products doing to integrate with open source tools and respond to some of the motivations that led to their creation?

Networking Reception

Speaker Biographies – Feature Sessions

James Blackburn, Trading Systems Developer, Man Group (AHL). James is a technical lead responsible for AHL's market data platform. He's responsible for the data access APIs, data storage and retrieval. He has rebuilt AHL's data storage platform, which is used for tick capture and storage, downsampling to minute bars (and lower frequencies), and data access through a Python and Java API.

Patrick Callaghan, Solutions Architect, DataStax. Patrick is an experienced developer/ architect and has been working in the finance industry in London for the last 12 years. He has worked with a long list of investment banks such as RBS, UBS and Commerzbank. Patrick now works as a Solutions Architect for DataStax in London helping financial companies to develop mission critical systems using Apache Cassandra and DataStax Enterprise.

Barry Childe, Head of Research & Development - HSBC HTS iLABS, HSBC. A specialist in Investment Banking technology, Barry serves as a technology Intrapreneur in HSBC's strategic innovation and investments team. Barry holds a PhD in computer science and has over 30+ years experience within Financial Services technology. Barry represents HSBC as a technology ambassador at a number of financial services technology councils and at academic and industry accelerators and startup programs. Previously Barry has served senior technology roles for Barclays Capital; The Royal Bank of Scotland; VMware; Westpac, TD Securities and Nordea Bank. Barry lives in Seaford, England and his favourite pastimes are golf and motor racing.

Jim Cownie, Principal Engineer, Intel. Jim is an Intel principal engineer and is currently architect for the OpenMP* runtime. He has worked on parallel computing since 1979 when he started at Inmos working on Occam and the Transputer. He served on the HPF and MPI committees, designing the MPI profiling interface as chair of the MPI-1 profiling sub-committee. Since joining Intel eight years ago, amongst other things, he has worked on the Pin profiling infrastructure, Intel® Transactional Synchronization Extensions (Intel® TSX) and OpenMP. He lives in Bristol, and would rather be skiing.

Herman Fick, CEO, fraXses. Herman has a strong background in IT Management (ex Group CIO), practical business solutions, general IT, and Big Analytics. As a founding member of Quantum System Integrators in 1993 and then of the Intenda Group of Companies in 2001, Herman is not afraid to put his energy behind his beliefs in the IT space. He combines his entrepreneurial approach to life with his love for solutions designing. The new fraXses technology and methodology for Big Analytics is a prime example. Herman has built the global 'go-to-market' strategy and the product roadmap while acquiring and integrating a new Machine Learning and Neural Technology company for predictive analytics. In the process, Herman has taken fraXses to the next level--from technology stack to application stack.

Matthew Grosvenor, Doctoral Researcher, Cambridge University. Matthew focuses on real-time and low-latency datacenter systems at the University of Cambridge Computer Laboratory. He is an experienced software and hardware developer and has worked for a collection of start-ups including Exablaze,Zomojo and Avalias as well as several research laboratories including NICTA (Sydney), Microsoft Research (Silicon Valley) and Microsoft Research (Cambridge). Since starting at Cambridge, he has twice won the ACM SIGCOMM Best Student Researcher award, as well as the Brendan Murphy Memorial Young Researcher Award. Matthew has a BEng (Hons 1) in Mechanical Engineering and BSc in Computer Science from UNSW (Australia), where he attended on a UNSW Co-operative Scholarship.

Neil Horlock, Director, Credit Suisse. Neil joined Credit Suisse (then known as CSFB) in 1998, from Financial Times Information where he became a core part of Credit Suisse's Market Connectivity development team and was central to the design and implementation of their in-house European Direct Market Data solutions. Neil rapidly became an SME in European exchanges, protocols and networks and helped grow the products beyond Europe to Asia and from Equities into other asset classes. In 2005, Neil designed and led the implementation of a (significantly) sub millisecond order routing solution for use within the proprietary trading business, a product that still forms the basis of the bank's high performance ultra-low latency trading platform. In recent years, Neil has represented the bank in a number of technology based consortia including projects such as BOAT and Turquoise and is currently part of the Nanostreams research project led by Queen's University, Belfast. Neil is a Director in the Trade Execution Technologies group at Credit Suisse and his responsibilities include oversight of the European Market Connectivity operation, Global Architecture for Market Connectivity, pre-trade risk and Regulatory Reporting functions, he is alleged SME for topics such as C++, market connectivity and the broad spectrum of low latency solutions, is an all-purpose trouble shooter and general-purpose busy body.

Martin Kleppmann, Apache Samza committer, former LinkedIn Engineer. Martin is a software engineer and entrepreneur, specialising in the data infrastructure of internet companies. His last startup, Rapportive, was acquired by LinkedIn in 2012. He is a committer on Apache Samza and author of the upcoming O'Reilly book "Designing Data-Intensive Applications" (http://dataintensive.net).

Matthew Knight, Marketing Director Financial Services, Solarflare. Before joining Solarflare in January 2014, Matthew was the Company President of Accensus, a company building an ultra-low-latency hybrid software/FPGA trading platform. Prior to that he worked at DRW Trading in Chicago in the role of Head of Labs focused on ultra-low-latency technology, and before that he worked at STAC in its early days, following almost a decade at Reuters.

Peter Lankford, Founder & Director, Securities Technology Analysis Center. Peter leads STAC®, which provides hands-on technology research and testing tools to the finance industry and facilitates the STAC Benchmark Council™, a group of leading financial institutions and vendors that engages in technical dialog and specifies standard ways to assess technologies used in finance. Prior to STAC, Peter was SVP of the $240M market data technology business at Reuters. Prior to Reuters, Peter held management positions at Citibank, First Chicago Corp., and operating-system maker IGC. Peter has an MBA, Masters in International Relations, and Bachelors in Chemistry from the University of Chicago.

Leo Martins, Managing Director, Pontus Networks. Leo has over 18 years of experience in architecture, development and support of message-oriented middleware systems, and has worked at number of Tier 1 investment banks (e.g. Barclays Capital, HSBC, Credit Suisse), as well as for middleware vendors such as TIBCO Software and Talarian. Leo has recently architected and deployed Informatica / 29west's UMS, UMP and UMQ products at HSBC, and prior to that, he architected and implemented the adoption of Solace Systems as a key middleware supplier to Barclays Capital. As a result of experience as a vendor and customer in the middleware industry, Leo has very close connections within TIBCO, Informatica, and Solace Systems. Leo received a B.S. in Computer Engineering, and a B.S. in Computer Science from Colorado Technical University, and his Executive M.B.A. degree from London Business School.

Federico Silla, Associate Professor, Technical University of Valencia. Federico received MS and PhD degrees in Computer Engineering from the Technical University of Valencia, Spain, in 1995 and 1999, respectively. He is currently an associate professor at the Department of Computer Engineering (DISCA) at that university, and external contributor of the Advanced Computer Architecture research group at the Department of Computer Engineering at the University of Heidelberg. He is also member of the Advanced Technology Group of the HyperTransport Consortium. Furthermore, he worked for two years at Intel Corporation, developing on-chip networks. His research addresses high performance on-chip and off-chip interconnection networks as well as distributed systems and remote GPU virtualization. He has published numerous papers in peer-reviewed conferences and journals, as well as several book chapters. He has been member of the Program Committee in several of the most prestigious conferences in his area, including PACT, ICS, SC, etc.

Rashmi Tank, MD, eFICC, Barclays. Rashmi runs the Statistical Modelling and Development team in eFICC Trading at Barclays capital. Prior to that he was in the Quantitative Analytics group at Barclays working on derivatives pricing libraries. He has a background in computational solid state physics.

Glenn Wright, Systems Architect, DataDirect Networks (DDN). Glenn has 20+ years of technical experience in the High Performance Computing industry, working on both server, networking and storage elements of extreme performance solutions. Glenn has been with DDN for the last 3 years, during which time he has helped clients across HPC industries solve the new "big data" high performance/low latency problems. Glenn has recently focused on technical infrastructure solutions in high capacity analytics for environments with the need for extreme levels of I/O between the storage and server components of the solution. Prior to DDN, Glenn was a senior architect at QLogic (Infiniband group) and also held many varied/international roles at Sun Microsystems.