Twentieth Greenside US Research Tour (May 2016)

Reflections on what was a great opportunity to learn more about the latest data domain developments and innovations coming out of Silicon Valley.

Normalising the Hype surrounding data

During the weeklong event, we received presentations from thirty-nine start-up companies, three Venture Capital firms (Intel Capital, Andreessen Horowitz and Sequoia Capital) as well as Microsoft and Splunk.
Trends and developments associated with the management, manipulation and exploitation of vast amounts of corporate, social and sensor generated data were hot topics. What was previously considered hype is becoming normal as organisations realise the critical importance of data, big and small. The compound effect of Moore's Law fast approaching critical mass and continuing improvements in data processing techniques look set to enable new paradigms of machine learning and artificial intelligence.

Start Up Big Data Firms are successfully eliminating barriers to adoption

New companies have been driving rapid growth of systems that support non-relational and semi-/un-structured forms of data, as well as increasing their analytic capabilities at massive scale and speed. Established big data and analytics solutions are evolving and maturing to bring them into line with enterprise IT standards and a number of start-ups have come about specifically to address gaps in solutions that make them difficult to integrate with existing business systems and operating models. This trend for retrofitting seems likely to persist as the rate of development of new raw capabilities increases. Big data start-ups are contributing, building and integrating components such as security, authentication, fine-grained role based authorisation and business continuity capabilities that customers expect from traditional enterprise relational database management systems. These, previously somewhat overlooked, capabilities are now becoming key enablers and differentiators within the ecosystem of emerging big data technologies, eliminating barriers to enterprise adoption.

If you want to get in touch please email info@ntegra.com

New Solutions allow for business users to replace data scientists.

As well as working to provide enterprise class systems, companies are building solutions that enable both business users and data scientists to fully realise the value of their data. There is growing demand from business users to have the same self-service access to insights that they get from traditional data warehouse environments. Companies such as Datameer are blurring the lines between traditional business intelligence and big data, enabling business users to discover insights in any data via wizard-based iterative point & click analytics and drag & drop visualisations, regardless of the data type, size, or source.

Companies are rapidly negating the need for swarms of experts in white lab-coats to be continually nursing corporate big data solutions. Self-service, self-discovery and automated commentary (describing why insights and analytic results have come about) are what's expected by today's business users.

Innovative companies humanise data access and simultaneously reduce cost.

Business users also want to reduce the time and complexity of preparing data for analysis, when dealing with a variety of data types and formats. Companies, such as MarkLogic and GigaSpaces have put a lot of focus on end user data preparation. Their customers can also dynamically asynchronously scale up or down the amount of storage and compute resources in the databases relative to the larger amounts of information stored in data lakes. Storage of data is comparatively cheap compared to the cost of the compute resource needed to process it so it makes sense to use the elastic provision of resources in the cloud, to ensure that compute is only paid for when it is actually being used. The effort to humanise IT (enabling people to intuitively interact with systems as opposed to systems asserting behaviours that enable interaction) is increasing. A number of companies demonstrated how data security can be moved into the background using biometrics and new algorithms to enable faster, non-intrusive authentication and authorisation. Companies like GoodData are providing both the tools and the expertise needed for organisations to collect, analyse and exploit data, allowing users to quickly and easily see the impact of changes made on performance. This leads to a more human approach, to performance improvement, based on feedback loops and proven success.

Schema-less database concepts and adoption of NoSQL technologies were a strong theme.

Adoption of NoSQL technologies and a preference for storing data in unstructured schema-less form (where data is applied to a plan or schema as its being pulled out of storage, rather than when it is written) were common themes. NoSQL and schema-on-read databases are becoming an established part of the enterprise landscape as the benefits of schema-less database concepts become more recognised and understood.

Enterprise death of the data warehouse is being replaced by cloud based data warehouse services.

Growth in the massively parallel processing (MPP) data warehouse segment has been slowing recently and the "death of the data warehouse" has been predicted. However, companies such as Cazena are driving a resurgence in the popularity and use of this technology in the cloud. Their solution provides self-service orchestration of cloud infrastructure in Amazon AWS RedShift and Microsoft Azure SQL Data Warehouse. Cazena uses these environments to provide, what it calls, Data Mart as a Service (DMaaS) alongside Data Lake as a Service (DLaaS) configurations on Hadoop and other schema-less databases. This enables on demand provision of data processing and analytics platforms, with other technologies including Google BigQuery likely to be included soon, giving customers seamless access to best of breed workload engines and heterogeneous infrastructures - something most enterprises would simply not contemplate on premise.

The emergence of the Internet of Things (IOT) is further driving cloud growth.

Data volumes from devices connected to the internet of things (IoT) is a further driver for petabyte scale growth in the cloud. Established companies such as Google, Amazon Web Services and Microsoft are developing IoT services to enable data to move seamlessly to their cloud based analytics engines. Services that ease the pain of wrangling and conveniently storing this data, are enabling companies like Arundo to develop predictive solutions that raise asset utilisation and performance in industrial companies. They do this by combining sensor and transactional data with deep domain knowledge and experience to reduce maintenance costs and avoid unexpected outages using machine learning techniques.

Ultimately the need for human literacy may overtake IT literacy.

Using machine learning algorithms to enable systems to learn how humans work (not the other way around) helps to spot patterns and connections between activities and performance, leading the way to more innate decision support and, ultimately, autonomous decision making capabilities. Driverless cars, drones and robotics are pushing back the boundaries of what's possible in this domain. Developing technologies so that interaction is driven by human preferences and needs, rather than technology capabilities, will come to the forefront during the next year. IT literacy may not be a differentiator for individuals soon, but human literacy and understanding will become an essential component of future systems.

Discovering the latest Security developments.

The 2016 tour also provided a great opportunity to learn more about the latest Security developments and innovations coming out of Silicon Valley.

New firms are highlighting concerns that organisations are failing to spot.

New InfoSec and cyber security companies highlighting a broader spectrum of concerns that should be considered holistically. Multiple layers of security aimed at managing specific risks are needed, and these should overlap to provide strength, depth and dependability. Ways of working are continually changing, leading to more opportunities for data to be targeted.

Organisations cannot assume they have enough security in place to prevent them from being compromised. They must continue to defend themselves and assume they're about to be breached, and have appropriate controls in place to manage the aftermath of a security incident. Most organisations are struggling to understand the risks they face and to take appropriate steps to be able to respond if and when they are compromised.

Security teams are struggling to keep up with the new ways that individuals, good and bad, can access, use and interact with organisational data and information. New devices, new apps and new services all pose additional cumulative risks, and many do not have security controls built-in. Development of new services and apps by small companies on tight budgets often mean security is poorly implemented, if thought about at all. This complex and diverse threat landscape is growing and IT security teams simply do not have the time, resources or awareness to keep up.

A combination of approaches is needed to overcome the contradictory needs of protecting data and allowing access from unknown environments.

Some organisations resort to using red teams to test their IT systems and provide the insights needed to update and improve their security controls (technology, people and process). Organisations can also create fake, network connected, machines (such as the TrapX honeypots) to divert hackers away. These approaches help to provide some confidence and can be effective. However, as well as protecting data, organisations need to ensure that the right people can access the right information, when they need it, often from external environments where there are many unknowns. These contradictory requirements have led to a number of big data and analytics companies, for example Prelert, to re-purpose their capabilities and concentrate on individual usage patterns and behaviours. The ability to understand and baseline normal behaviour enables detection of abnormal activities and anomalies.

Hacker journeys are identifiable, in the same way that customer and employee journeys are, within time series data collected and logged across all the potential touch-points and routes through the corporate IT landscape. SS8 showed us how this can help to highlight risks from suspects of interest or devices of interest and illuminate areas where different security layers must interact and flex to form an appropriate defensive barrier which is only permeable to appropriate (normal) users and usage.

Innovative start-ups have tools to help increase security as software is being developed.

To harden and help improve the quality and lineage of new software products, start-ups such as BlackDuck and Tinfoil have developed automated scanning mechanisms that can be integrated into development pipeline processes. BlackDuck is able to detect open source components and libraries, that are commonly included in builds, and hence provide developers or end-users with an assessment of their licence exposure and risk. Tinfoil scans newly compiled executables to identify all known security vulnerabilities and then raises defect reports in the developers' bug tracker (complete with instructions on how to debug the code and eliminate the vulnerability). Asserting Tinfoil into the Continuous Integration development cycle is like having an infallible security expert with coding skills in the development team.

Employee behaviour is a key area to focus on.

Outside of key mission critical areas such as defence, in reality IT is currently largely uncontrolled due to the proliferation of uncontrolled devices and regular use of uncontrolled networks. Employees have little or no training and their knowledge of IT security is, at best, poor. They move between organisations, cross-pollinating careless behaviours that lead to contagion in process vulnerabilities that, in turn, provide huge opportunities for hackers to exploit. For these reasons, security needs to be considered and planned at all levels: from employee identity; to building security into bespoke applications; to managing devices; to tracking intrusions.

The top, and most obvious, vulnerability layer is no longer an individual's ID and password, but his or her identity itself. From just the name of an employee, a hacker can gain an entry point (for example, making use of social media to access personal details combined with a knowledge of an organisations remote access solutions can often provide enough detail to gain initial entry into corporate systems). Once inside the lobby, brute force password attacks can be completed in less than a few second due to the immense computing power that is now available, on demand, in the cloud.

One innovative approach that reduces the chances of an employee making basic mistakes was demonstrated by Menlo Security. They protect against cyber-attacks from the web and e-mail by providing an Isolation Platform which insulates content and eliminates malware in the cloud. Users web sessions, and all active content whether good or bad, is fully executed and contained within the Isolation Platform. Only safe malware-free rendering information is delivered to users' browsers. No active content leaves the platform, so malware has no path to reach an endpoint, and legitimate content does not need to be blocked in the interest of security.

The race between innovative new IT solutions and the cyber threat is on!

In 2016, we saw that already overstretched security teams were also having to think about, previously unforeseen threats and new technology responses from outside their traditional domain and skillsets. SkySafe provided an interesting description of their anti-drone systems which can be used to protect airspace and detect, and bring down, drones that may be eavesdropping corporate secrets by listening in on conversations, or photographing industrial or military installations. Pindrop described their anti-fraud and authentication solutions for enterprise call centres. Within 10 seconds their technology can generate a risk score by simply listening to the background noise on a telephone connection and assessing whether the call is really coming from where it purports to be from. DeltaID made a good case for the use of iris recognition as a better bio-metric than face and fingerprint. Face recognition is a challenging technology, with changes in lighting, hair style/colour and other looks resulting in a variable user experience. Fingerprints are a better and more popular bio-metric but they can be problematic as they may be effected by weather, age, work and many other factors. Using the iris as the bio-metric results in a much better accuracy.

The above examples and many other interesting insights, from new companies, helped to illuminate the broad extent and continuing growth of the cyber-threat surface. This can be seen as a proxy for, and measured proportionally with, the rate at which new innovations and ever more complex IT solutions are being delivered.