



vLAB CIO Reverse Pitch Challenge Summary January 26, 2016

Niel Nikolaisen – CTO of OC Tanner - Legacy App Modernization

Breaking legacy apps into Lego Blocks in order to improve business speed and agility.

Many of us (most of us?) have a decent-sized portfolio of legacy applications that need to be modernized. As we develop mobile apps or need to increase business and IT agility, these legacy applications are a barrier to speed and innovation. They are often complex, monolithic structures that perform mission critical tasks. Even if they were originally well-architected (not very likely) they have been built on and built around and bolted together in a way that makes them difficult to change, risky to even touch and in our way when we need to move fast. We could simply replace these legacy applications but that would take a significant amount of time and resource. So, we tend to work around them rather than deal with them – and not dealing with them means we will continue to be slow and a barrier to our agility and innovation.

So, what makes them hard to deal with? Here is my list:

1. The applications are complex because they are interconnected. Think of them as a large hairball or Gordian Knot. Touch one strand and it might break things you had no idea were connected to that strand. Thus, it takes lots of time to make and test changes.
2. They are often built with diverse tools, frameworks, technologies and approaches. Some of those are in different states of obsolescence.

What would help? Rather than replace the hairball, I want to break it into loosely-coupled pieces that connect in a logical way. I need to do this both vertically and horizontally. By vertically I mean that I need to separate the UI/UX from the logic layer and the logic layer from the data layer. By horizontally, I mean that I need to separate the logical “services” from each other. This sets the stage for improved agility. If I need to revise user authentication, I change only the user authentication service as it is loosely-coupled with everything else (like user profile). This also sets the stage for improved reuse. If I have multiple applications that need an authentication service, I can use just the one because I have sort of abstracted it out of the hairball and made it available for use to other applications. In effect, I want to break my applications into the equivalent of Lego blocks. With a reasonably small number of Lego blocks, I can build a nearly infinite number of designs.

I could make this happen if there were a way or a tool (or set of tools) that I could use to:

1. Discover that application code / elements connects with what.
2. Discover which transaction flows move through the UI/UX and data layers (so that I can separate them).
3. Discover which code / elements are rarely, if ever, used.
4. Determine which are the natural application elements (like authentication or profile or ...).

The challenge with this is that these elements and flows are inside of the hairball application. So, a traditional approach like an ESB might not help as typically an ESB connects the various hairballs. What I need is to get inside the hairball. But, I am not smart enough (or even close to being smart enough) to figure out how to get inside the hairball, understand how to best break it apart and then couple it loosely to the sources and consumers of its data and transaction outputs. Thank you.



Raffaele Piemonte – VP of Applications and Infrastructure, Center Light Health System - Sunset of Legacy Applications/Products

Keeping the legacy data but not the legacy cost.

Many companies are moving from old or to better products/solutions. This leaves data in the old systems and issues with realizing true ROI due to lingering licensing/maintenance fees. A great solution would figure out a way to efficiently migrate the data with associated metadata to a data store. This data store will allow access and extensive reporting features or be able to leverage common reporting engines without the product/application being live. Ideally the data would be migrated to an open source solution so that it would keep down licensing costs around the database structure. The reporting solution should include some metrics on how often data is being accessed and viewed so that the application/product can be properly retired.

Christopher Wilson – Principle Cloud Architect, Verizon Enterprise Solutions - Event Classification across the IT Infrastructure

Provide a service, collection of services, 3rd party software or plugin which can classify events occurring across the IT infrastructure into incident management levels.

Explanation: Though there are systems which exist for event correlation (Splunk, Zenoss, IPSoft) categorizing and classifying those events into appropriate incident management queues (ServiceNow, Remedy, IPSoft) is typically performed as a one-time event by an administrator and filtered and routed thereafter. Similar events, events of similar classes, are often dropped by filters which prevent incident managers from being overwhelmed, but lead to missed opportunities for speedy or proactive resolution.

1. As the customer I want the ability to import historical incident management data into the system. These data may contain incident management queues, ticket details, times, and resolution details.
2. As the customer I want the ability to import historical application and event logs into the system. These data may contain system log events, errors, application errors, customer defined events, etc.
3. The system shall inspect the incident management and event data and perform the following:
 - a. Cluster events which are similar. These clusters may be clustered by time, customer, application, natural language captured within the system logs or incident management system, or any of the above.
 - b. Provide classification of events based on supervised and [hopefully] unsupervised learning techniques. These may include novel language models, distance metrics gleaned from clustering data, or other advanced Machine Learning or statistical methods.
 - c. This step shall contain a holdout set to determine model accuracy.
4. The system shall provide a usable web interface for the customer to use to understand events, event classification, and the accuracy of the model(s) provided.
5. The system shall update its model based on these historical data and accept new events and route to appropriate incident management queues.
6. If unseen events arrive which are of a user defined threshold (business critical events) classify these events appropriately or recommend new queues be created.
7. Allow the customer to tag/flag events as misclassified and take these errors into account in any model updates in steps 1-3.

Definition of Done: Done shall be a demonstrable artifact from the user story defined or any agreed upon part of this user story/epic. These artifacts are listed below:

1. The system shall demonstrate importing historical log and incident management data into the system
2. A user interface (preferably web based, but command line is ok in early stages) shall allow customers to browse these events visually.
3. Visually represent the data clusters.
4. Visually represent the data classifications seen and recommended classifications (Sankey diagrams nice here ... I think).
5. The system will run (manually, scheduled, or automated) a series of machine learning models with the goal of correctly classifying events into the appropriate incident management queues. Model accuracy shall be estimated and produced visually.
6. The system shall accept unseen data from other/random system or event logs and attempt classification. Model accuracy shall be provided.
7. New incident management queues or re-classification of events into existing queues shall be recommended.

Those first steps just help guide administrators to be better at their current job and I think that first step is great. Show me my events, show me how they classify and may be re-classified manually and that covers step 1-4 above. The next set of cases as you would start chucking unknown data at the system and re-building your models, you'd have to scale the system, etc. The next step, after the math, business case, exit strategy, etc. pan out you'd scale the system by throwing random systems at it, random applications, see how the models rebuild and test accuracy by getting feedback from the administrators or another pull of the data and see if the event was manually reclassified.

Dr. Kaijun Zhan – Director of Global Infrastructure & Engineering Computing, Atmel Corp - Back-up, Tape is Fraught with Problems

Back-up - Cloud is not always the easy answer.

In many companies, tape based backup is still the norm, yet we all know the cost, complexity and reliability – imagine you have many global offices and you need libraries and tapes and logistics etc. So cloud based is the direction we all like to go.

However there are a few questions today that prevent people from moving to cloud backup as mainstream technology:

1. The cloud providers (AWS, Azure...) they do not have backup/archive tools – they only provide the “media”.
2. The data replication guys, they ensure the data is transferred effectively, de-dup, compression, or whatever, this is a critical core technology element as you need to transfer large amount of data over to the cloud. However they typically don't have the application – enterprise grade like Symantec .
3. The backup guys, Symantec, Tivoli etc., they have the nice application, but don't have the core technology to replicate data.

So you see there are things here and there in place but yet no single person can make a statement: “ I will transform your tape backup environment, to the cloud, globally, with cost efficiency, but transparent to you so you would not feel any different than what you do today”. There are consumer type of tools today or desktops. But not enterprise level.



Robert Stevenson – Chief Strategist, Hitachi - Open Stack Integration and Orchestration with Chef and Puppet

Need for an Open Stack light for IOT use cases.

The challenge is trying to collaborate puppet and chef with service providers via product suppliers. I am starting to see a ServiceDesk model that follows this problem but was looking for an open source alternative.

Vince Tripodi – VP of Engineering, Associated Press - Content Processing

A new look at Metadata for Media Companies' use cases.

The challenge in AP's content production is to "develop intelligent systems" to increase efficiency, automate processing where possible and filter content for enhanced product offerings. The goal in this area is to harness technology to lower production costs and at the same time add new value to our products through smart filtering and custom packaging. Early efforts in this regard include the development of our industry-leading metadata system for organizing content, our automation of stories from data in business news and sports and the launch of our new unified customer portal that presents content across media type in "story-centric" packages.

I'd pitch on the value of automated application of content metadata and its value to AP to both reduce current expense around production and lead us toward 2020 goals. AP creates close to 75K pieces of unique content a day: stories, videos and photos. Platform that authors, acquires, enriches and distributes that content globally for B2B and over mobile. It's the "Enriches" piece that I'll be talking about today, and how AP's metadata drives product development and customer value. We've developed an industry-leading metadata system for organizing content. That platform is the source of my pain today ... on-prem, labor intensive, 10 years old and tuned for text and not photo/video.

What is metadata:

1. Taxonomy - ontology. Fixed terms that describe breaking news.
2. Rules that apply taxonomy to content.
3. Platform that applies taxonomy to content using rules - using proprietary on-prem licensed technology.

Types of metadata we generate: 4200 subjects, 2100 geographic locations, 1300 organizations, 93,000 people, 43,000 publicly-traded Companies, top level subject areas, arts and entertainment, business, demographic groups, environment and nature, events, general news, government and politics, health, lifestyle, living things, media science, social affairs, sports, and technology.

How we do it:

1. Rules-based system – each term in the taxonomy has an associated rule.
2. People and company tagging is based on mention, with significant disambiguation rules to ensure accuracy. 85% recall is target for quality.
3. Subject, geo and organization tagging is based on more complex rules, and strives for "aboutness."

What we use it for:

1. Product development.
2. Improved search and discovery.
3. Automated aggregation, syndication and distribution of related content.
4. Richer and more relevant content products and services.
5. Reduced time to market for new products and services.
6. Reduces editorial workload, creates efficiencies.
7. Content interoperability.



Some use cases:

1. Allows customers to follow a company, person, or topic over time.
2. Enables us to deliver specific products, e.g. Green technology, Elections packages, 2016 Olympics hometown athletes, etc.
3. Aggregate AP & partner content.

Limitations:

1. Text based only - (stories, photo captions and video scripts).
2. Photo and video analysis tech.
3. Supports English language content only.
4. Limited sentiment analysis.
5. Relevance and confidence scoring is challenging.
6. Proprietary on premise technology.
7. Breaking news introduces new tags/terms which require manual taxonomy updates & rules to be written.
8. Cumbersome maintenance of rules that apply taxonomy to content.

Opportunities:

1. Cloud based tech platform.
2. Crowd sourcing.
3. AI & machine learning content analysis.
4. Natural language processing.
5. Big data solution applicable to media companies and individuals.
6. More relevant, local, targeted advertising.

What we want: a rule-free, automated, cloud based platform that derives contextual metadata for text, photo and video content to support a breaking news environment.