Maximizing Velocity with SOA @ Yammer

Mike Ihbe
Senior Infrastructure Engineer
twitter.com/mikeihbe
github.com/mikejihbe

Saturday, May 19, 2012

Hi everyone, I'm Mike Ihbe, a senior infrastructure engineer at Yammer. Today we're going to be talking about Maximizing Velocity in a software engineering organization and how Service oriented architectures can help you manage that. Now, the title is pretty buzz-word filled so let's break it down. What do I mean by velocity?

Maximizing Velocity with SOA @ Yammer

Mike Ihbe
Senior Infrastructure Engineer
twitter.com/mikeihbe
github.com/mikejihbe

Saturday, May 19, 2012

Hi everyone, I'm Mike Ihbe, a senior infrastructure engineer at Yammer. Today we're going to be talking about Maximizing Velocity in a software engineering organization and how Service oriented architectures can help you manage that. Now, the title is pretty buzz-word filled so let's break it down. What do I mean by velocity?

What is Velocity?

$$\vec{V} = \frac{d}{t}$$

Saturday, May 19, 2012

Well, it's distance over time. From a software engineering perspective, velocity is the value created by your organization. This is a familiar concept, so it's easy to reason about. It's essentially about moving quickly, but more than that, velocity is a vector. Consistent direction is essential to effectively achieving your goals and maximizing the production of value.

At Yammer we think of distance as business value. It's the value we can create for our customers. Every time we release a new feature, fix a bug, or improve performance it adds to this value. Things like refactoring and tackling technical debt fit into this metaphor too. They count as friction. The forces your engineering organization can bring to bear must overcome this friction or you won't be producing any value. So, when I'm talking maximizing velocity, I really mean creating as much value as quickly as possible.

So, what is SOA?

What is SOA?

Technology

Software design paradigm enabling organizational flexibility through composable, loosely coupled, fault tolerant business services with clearly defined, published interfaces.

+

People

A human-organizational structure enabling small, directed, decentralized, autonomous, extremely effective engineering teams.

It's a software design paradigm enabling organizational flexibility through composable, loosely coupled, fault tolerant business services with clearly defined, published interfaces. That's textbook. We'll talk a little about these things, but I'd like to add a human perspective.

SOA is also about people and how we work. It's an organizational structure enabling small, directed, decentralized, autonomous, and extremely effective engineering teams.

Technology + People

Saturday, May 19, 2012

Most of us in here are software engineers. We all know technology is hard. Human organization is harder. Doing both well is nigh impossible. This talk is about some of the lessons we've learned at Yammer about how to organize your people and your technology to build a scalable organization that's laser focused on building value quickly with SOA.

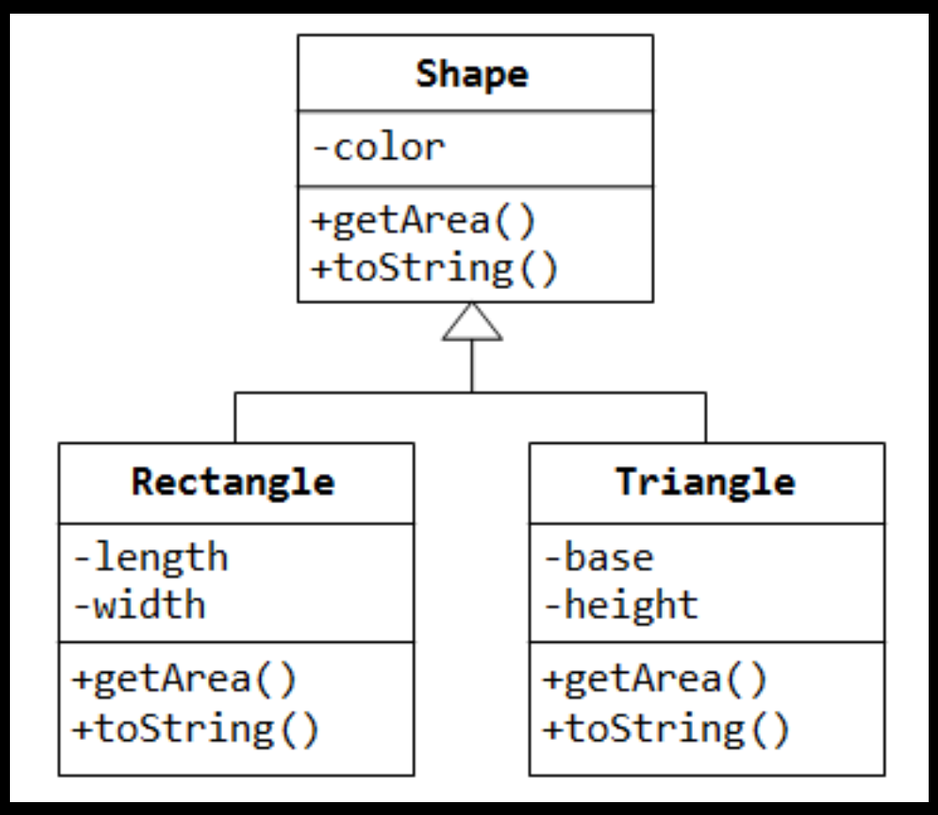
That textbook definition of SOA wasn't terribly explanatory, so let's wrap our heads around it a little more. We can begin by contrasting it to the alternatives.

VS the monolith

- Large, complex, unfocused
- Difficult to reason about
- Shared state/data

Large single code base applications inevitably run into the same problems. They're unfocused and complicated because they try to solve every problem. It becomes impossible to keep the entire program in your head. Developers step on each other's toes. SOA provides an answer to these concerns by stealing some principles from object oriented programming.

OOP Similarities



- Encapsulation
- Focused functionality
- Clear interfaces
- Reusability

Encapsulated concerns and focused functionality make the pieces of a complex system easier to reason about. Clear interfaces hide implementation details and make services composable to accomplish even more complex tasks.

The last way to think about SOA is the human organization behind it.

Organizationally

"Organizations...are
constrained to produce
designs which are copies of
the communication structures
of these organizations."

-Conway's law

This is Conway's law. "Organizations...are constrained to produce designs which are copies of the comunication strucutres of these organizations."

Basically this is saying that if you hire a firm with 3 engineering teams, you'll end up with a product that has 3 subsystems and the quality of the interfaces between those subsystems will reflect the communication quality between those teams. Conway was a smart man. There are studies out of Microsoft Research and Harvard that help corroborate this.

SOA, when wielded properly, can help build an organization that takes advantage of Conway's law to improve quality.

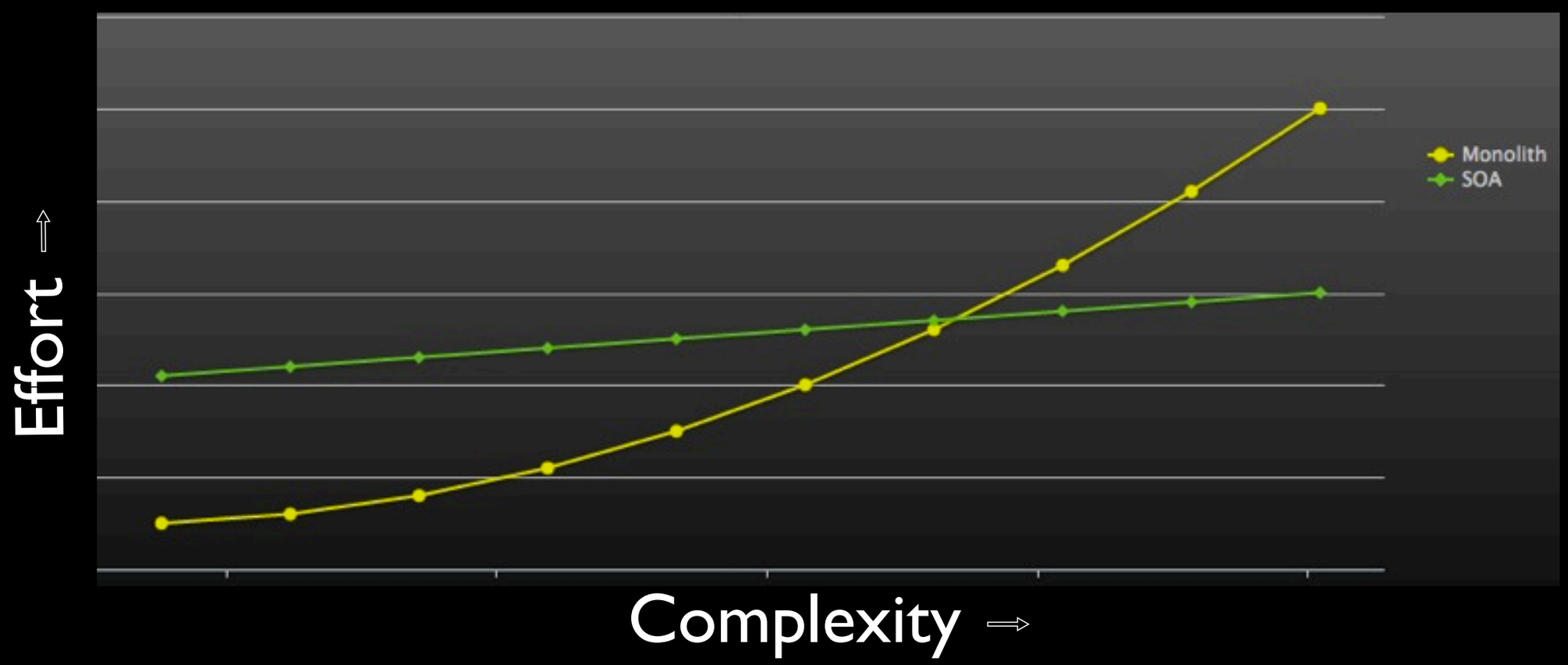
Panacea?

All that makes it sounds like SOA is all glitter, rainbow and unicorns. It's no panacea.

As with most engineering, there are very real costs behind all those benefits.
With a single, undivided code base you're able to dance around class boundaries, access data layers directly, and avoid all the overhead that managing many disparate production services entails.

There is value in that, but at some point the overhead of managing many services becomes less than the overhead of dealing with all the complex, wild-west engineering that was happening in your monolithic application.

Costs



Monolith: $\frac{x(x-1)}{2} + D$

SOA: $Cx + 3D$

Saturday, May 19, 2012

So, what do these costs look like? I've attempted to model them here. Complexity is my hand-wavy aggregate of the number of people working on a project and LOC. Effort, is the cost of building and maintaining a project.

We can see that the monolithic application's effort grows quadratically. The classic wisdom here is from Mythical Man Month - the effort needed to communicate within a team grows at a rate of n(n-1)/2, we also include the difficulty of the project as D.

To contrast that, building systems has some significant overhead. Systems programs are supposedly 3x more expensive than creating the product. Because we're building a small, focused service, the number of people and LOC are bounded, and that's the key to keeping effort growing linearly. It will always start higher than the monolithic approach, and it will always end lower.

In reality, there's a missing factor in this second equation to account for the complexity of the entire system of systems, but that factor would be based on the service dependency graph of a specific deployment and should be small for any given new project. The point is: as your System's complexity increases, SOA becomes much more viable.

When is SOA worth it?

The rest of this talk is about maximizing the velocity of a large software project - and at what points SOA becomes useful.
For many projects, becoming service oriented can be a natural progression, and we'll discuss Yammer's experience in this area. We've found that some of the ideas behind SOA, especially decentralization, make for extremely productive engineering teams.

It's probably intuitive that maximizing velocity looks very different depending on the stage of your project and organization. So this is my perspective on how to maximize velocity at the different phases of a project development lifecycle superimposed on the the human development lifecycle for your entertainment.

Babies

Saturday, May 19, 2012

We all start as babies. This is the new project stage. The team is small, communication is easy, you're probably all in the same room and you can just talk to each other face to face. Everything is warm and soft and cuddly.

People treat you nicely just because you're new to the world, even though you’re probably puking all over the place.

Baby Tech

- Web framework




- Database



- @Yammer
  - Rails
  - Postgres
- Normal Architecture
  - 1 Web / 1 Db

Technologically speaking, things are also pretty ideal. Usually, you've picked your favorite web framework with a simple database backend. From an application standpoint, these frameworks can get you really far for a surprisingly long time. Velocity wise, they're absolutely worthwhile. Feature development is fast, development environments are pretty standardized. You actually save on communication costs by leaning on the convention over configuration paradigms that a lot of these frameworks offer. Yammer has leveraged the heck out of rails - I think we have one of the largest rails apps in existence. We're mostly prototyping at this point, so the architecture story is nothing to write home about.

Maximize Velocity

1. Get to Market
2. Find a fit

Saturday, May 19, 2012

Maximizing velocity at this stage is pretty straightforward. It's usually a matter of getting something out into the world as quickly as possible and determining market fit. Some technical debt accumulates here and that's totally normal. You're going to try some things that will fail and collect dust in your code base. You're going to have other things that are hideous from a technology design perspective but end up as successful products. Survival is the only goal here. If someone looks at your code next year and throws up a little bit in their mouth, that's a totally first-world problem.

Once you've launched, you can move on to...the terrible twos

Terrible Twos

Saturday, May 19, 2012

Now, I’m not a father, but I hear this is actually a good thing. Our little kid is making an impression on the world, learning to express itself. At this point you've probably got customers, which is awesome. They might complain a lot, which sucks. Maybe things are getting slow.

Two Tech

- Caching
- Feature services
  - Solr / Lucene / Redis / whatever
- @Yammer
- Memcached
  - Solr / tsearch / Redis
- Normal Architecture
  - n webservers
  - Master-slave dbs

Saturday, May 19, 2012

Technology wise, Memcache or some other caching layer usually gets thrown in about here. You probably have a database slave. That's good - losing data is embarrassing.

We started to venture into services a little bit here too with a SOLR application. We didn't have the expertise to manage it, so we fell back to postgres fulltext search, using this gem acts_as_tsearch.

Velocity wise, this wasn't terrible - our project still wasn't that big the team was still reasonably sized. Communication wasn't that stressful - we were definitely helped because we were building an internal collaboration tool. As we started getting traction, tsearch started causing problems for our database replication layer, so we did eventually move to a simple Lucene application. We also created another service to handle our "online now" presence feature that was backed by redis here.

Maximize Velocity

1. Best tool for job
2. Focused team & tools

Maximizing velocity during your twos is pretty pragmatic. Stay focused, use the best tools for the job. We've started our approach the crossing point into SOA being more efficient, but we're not there yet. Since we happen to have some services cropping up, let's talk about bootstrapping SOA.

Bootstrapping SOA

@Yammer -
Core Services team to own service creation & maintenance

Saturday, May 19, 2012

So far, we've been adding services as our featureset demands. This is what I meant by it being a somewhat natural process. The services we've built have been dictated by the best tools for the job. Organizationally, we haven't added too much complexity. This may not look like it, but it's a fork in the road. You've started making services, you've probably diversified the technology stack -- Is your organization going to back these services? Are they going to build monitoring infrastructure for them? Are you going to hire dedicated people to maintain them? Are they going to build other services to help segment your quickly growing main application? This was a turning point for Yammer. We decided to make a Core Services team that would own these things moving forward. The important thing to remember here is to be deliberate about your decision making. Don't over engineer too early -- be incremental and flexible in your changes.

The next stage of development is learning to speak...

"Reliability"

"Uptime"

"Disaster Recovery"

"MTTR"

"MTBF"

Learning to Speak

Saturday, May 19, 2012

We're starting to grow up at this stage. We're learning words like "Reliability", "Uptime", "disaster recovery". Things are actually starting to fail occasionally, so we're thinking about minimizing MTTR and maximizing MTBF. It's not that you didn't know these words before, it's just that you haven't had a lot of cycles to spend on them.

This step is about establishing and enabling operational processes. Your system is probably interesting enough that an operations team has come on board. So they're working on things like automating disaster recovery paths. They're probably also spending time putting the beginnings of a complex monitoring system into place. Most companies do this, but it's not always with the intent of pursuing service oriented design. Regardless of the reasoning, monitoring is a critical cornerstone in a service oriented world. It's obviously important to know that your services are running, but as teams and services establish SLA's there must be a system in place to ensure compliance and to immediately detect regressions or your system will inevitably degrade over time. Planning for these needs will improve your velocity as the organization and number of services grows.

Speaking Tech

- Backups
- (semi) Automated Recovery
- Monitoring
- @Yammer
  - Smorgasbord :-(
  - Normal Architecture
    - Smart LBs
    - Ganglia / Nagios / Cacti / whatever
    - In-house recovery

Saturday, May 19, 2012

The actual technological choices diverge significantly between organizations at this point, so I'll talk more about strategies.

In any complicated system, there are going to be failures. You'll never be able to account for every failure mode. So, rather than focus on mean time between failures, Yammer tries to optimize the mean time to recovery for our services. Generally speaking, optimizing MTTR will allow your organization to make changes more quickly. Being prepared for these failures by mitigating their impact and automating recovery is the best way to ensure a system keeps running smoothly.

It's becoming popular for developers to be on-call for the services that they build (at Amazon and Google among others). At Yammer, we script and optimize the recovery path as much as possible to enable ops to quickly recover service outages. Developers are still on the hook for bugs, but are typically the 2nd tier responders.

Monitoring is a complicated topic, and building a complete monitoring solution for a SOA ecosystem is a tall order, especially with off the shelf open source tools available today. There are lots of them available: ganglia, nagios, cacti, etc -- but none do everything that we'd like them to and they're all painful to look at. We use a smorgasbord of them at Yammer, and we're just starting to look into customizing them enough to make them intuitively useful.

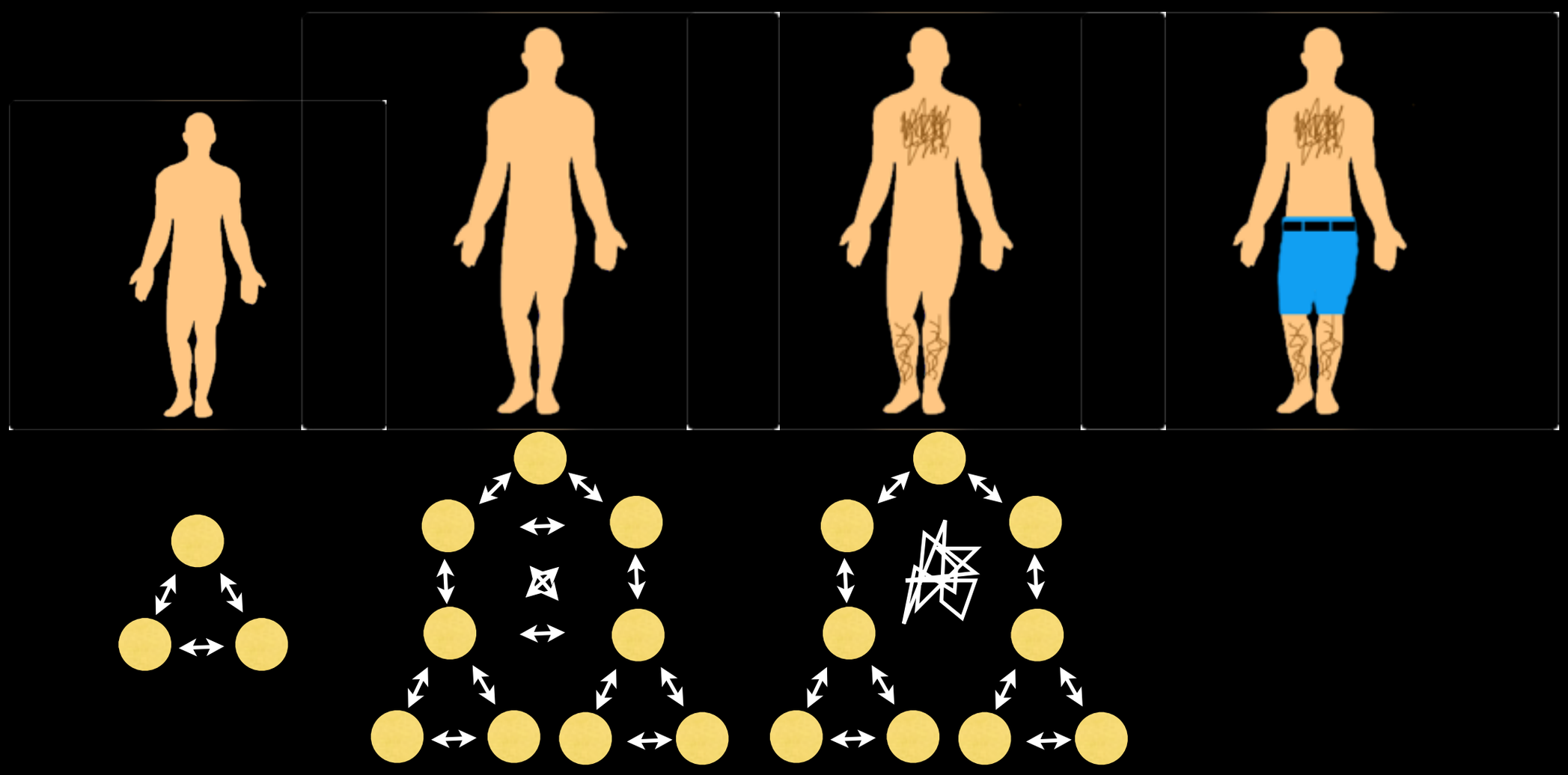
Maximize Velocity

1. Standard provisioning
2. Monitor

Saturday, May 19, 2012

The only operational advice I have to give is to standardize your server setup. It saves you a world of hurt later as you build solutions for automated provisioning, monitoring, and sources of truth.

As I said before, Monitoring is a cornerstone of SOA. You need to be able to enforce SLAs, detect regressions, understand what your systems are doing, and know where problems are coming from.



Adolescence

Saturday, May 19, 2012

At this stage you might start noticing some changes in your body. Hormones are raging through your body and things are starting to get interesting. You're growing like crazy, things that ran smoothly before, like communication channels, are starting to get hairy, parts of your body/organization that you didn't spend much time thinking about before, like security and compliance, are becoming much more important to your day to day activities. Your outlook on the world is also probably changing. You realize the world doesn't revolve around you anymore, if you want to get laid and/or succeed with customers you need to start dealing with grown up issues. You also have to start planning for the future.

How an organization deals with these growing pains has an immense impact on how feasible SOA is going to be for them in the long term. This is a very good problem to have. This is the crossover point. This growth phase is an indicator that the business needs to be considering SOA – focusing too much on it beforehand will probably just slow you down with needless bureaucracy and make it less likely that you'll even reach adolescence.

The canonical SOA example tends to be Amazon.

@Yammer -
Cross functional teams
flat, decentralized

Saturday, May 19, 2012

They got to this stage and continued merrily building their massive C++/perl application. During adolescence we're usually still strapped for personal wealth -- which in the computing business really means time -- so Amazon just bought some extravagantly huge databases to help them grow up. Eventually this became problematic, so they undertook the exceedingly painful process of switching to a SOA from their monolithic, single-database-backed application.

There's a lot of ways to learn from their mistakes and capitalize on their successes. From a human organizational standpoint they made small "2 pizza teams" to help keep communication overhead low. Yammer takes this a step farther. We spend a lot of time thinking about Conway's law and how to harness it.

A lot of organizations divide their departments vertically or horizontally, but we've found that this inevitably leads to silos within the organization of people working on the same thing day after day. This kind of arbitrary pigeon hole forces people to get attached to their "turf" and inhibits communication and decision making.

When we're creating a new service or feature we put together a "cross-functional team" with representatives for every aspect of the project. This creates an autonomous, well-informed, decentralized design process that yields well-designed, isolated systems. The key here is that these teams are ephemeral. They come together, solve a problem, then each team member moves on to something else. This prevents Conway’s law from shaping our solutions, we can instead focus on the right answer from an engineering perspective. This sort of structure naturally leads to a relatively flat organizational structure and lends itself well to a transparent and trusting organization.

DISTRUST

Saturday, May 19, 2012

Speaking of trust, trust is a big deal in a service oriented world. As a developer, when you're dealing with a remote service, a certain distrust is healthy because it forces you to deal with error conditions more strictly than you otherwise might, which makes your system more robust. The flip side of that is, because you're all in the same organization, you can typically rely on the people behind the services you're using to deliver quality product. After all, if they don't, you know where they sit. That level of trust also leads to happy, responsible engineering teams.

Thrift

ProtoBuf

XML

RPC

WSDL

SOA Technology

REST

SOAP

UDDI

JSON

Saturday, May 19, 2012

This is also a good time to be thinking about the technology aspects of SOA. What are your services going to look like in production? There's a lot of wisdom out there that says SOA is great with heterogeneous systems so services should be protocol independent and everyone can pick whatever protocols make them happy. That's a load of crap. If you can avoid that complexity, you should. Services can get the same velocity wins from convention over configuration that we saw with our early framework choices. There's no need to create multiple interface types for different services. That doesn't add any value. Standardization here is great. Having the same response formats, data protocols, monitoring interfaces, deployment stories, and dependency management systems will make your world infinitely easier than trying to manage tons of unique and special snowflakes that all speak different languages. Homogeneity is your friend. All of those tools will also help you maximize your long-term velocity by making it easier to get your new services to production quickly.

Adolescent Tech

- Varies widely - Tooling, Provisioning
- @Yammer
  - Dropwizard - Jetty, Jersey, Jackson, Metrics, Guava, Log4j, Hibernate Validator
    - Quickly deploy monitored, alerting, metrics reporting, logging fat Jars
- Crapapult - Deploy artifacts from CI
- Partie - Framework for distributed BDB-HA KV store

Saturday, May 19, 2012

Adolescent technology varies widely, but it's mostly focused on tooling and provisioning.

At Yammer, we have an opinionated service framework called dropwizard. It combines many of the libraries and tools that we've found useful while building java sevices. It allows a new developer to deploy a production ready, simple service that's monitored, alerting, reporting metrics, and logging in under 30 minutes.

We have a deployment tools for these services too. Any developer can deploy passing builds from our continuous integration tool.

We've also abstracted a distributed storage framework called Partie that's used by a number of different services.

A start to real services:
Artie - "RealTime" Cometd delivery to listening interactive clients
ROUS - Remote object upload service
Tokie - handles oath authentication for non rails services
Search pipeline:
Querie - lucene keyword search in yammer threads

Maximize Velocity

1. Standardize tools
2. Minimize time to production

Adolescence is really the time to invest in the future. Build tools that make it easy to build new services and you'll find that it becomes natural to create them.

And now we're on to the final stage, adulthood.

Adulthood

Saturday, May 19, 2012

Congratulations! You've made it! But adulthood isn't all fun and games. This is the longest, hardest part of the SOA lifecycle. It turns out that responsibility is a bitch. For the first time, you'll encounter problems that haven't been seen before that need custom solutions. You'll need to stay informed, pay the bills, and raise your family into well-monitored, independent functional units that maintain their service level agreements.

Adult Tech

- @Yammer - Specialized Domain Services
  - Feedie - 10+ Billion messages, 45 node cluster
  - Flatterie -> Dexie -> Completie / Querie

Saturday, May 19, 2012

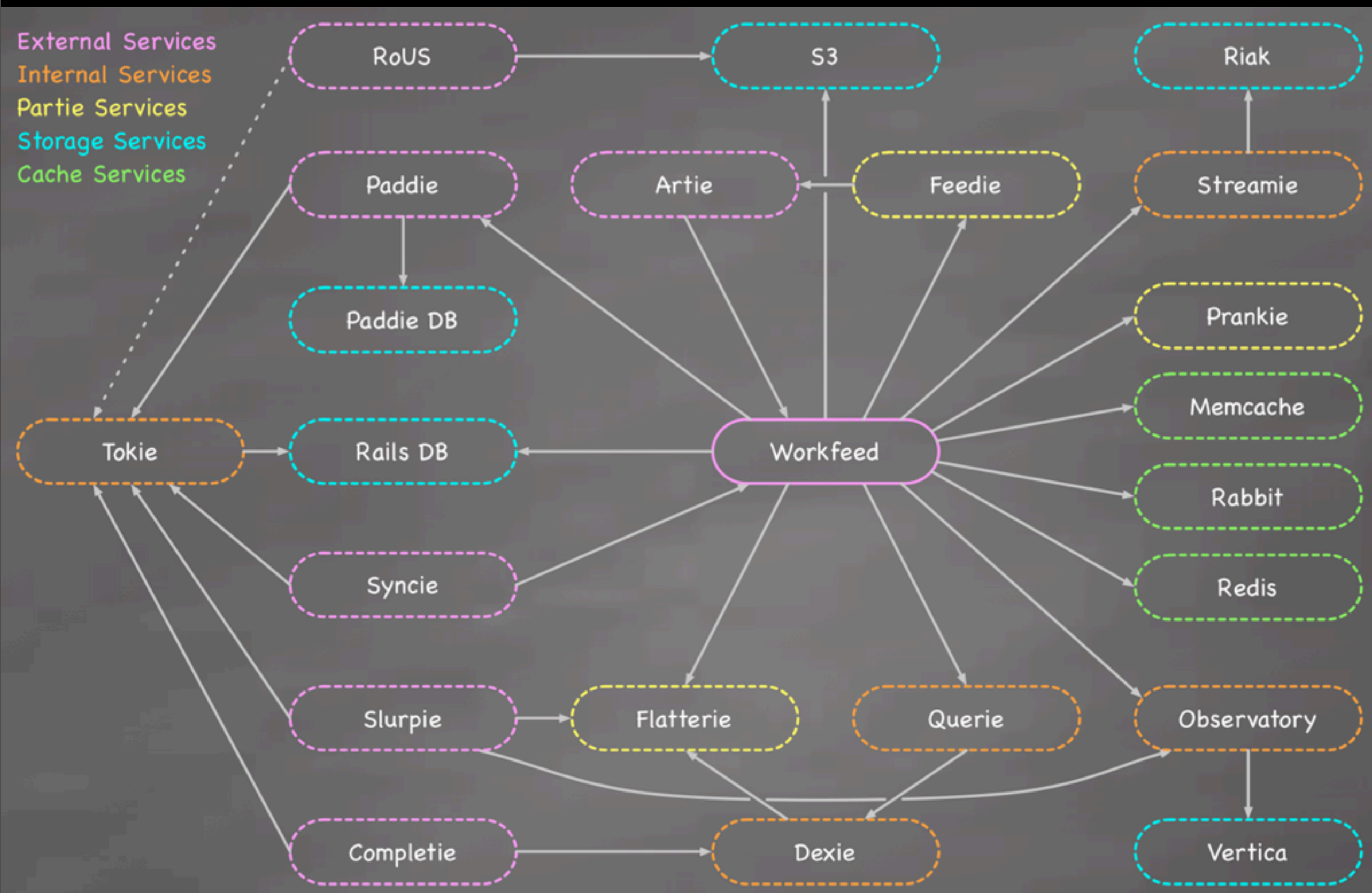
This is where the technology story tends to get very interesting. I won't even attempt to make sweeping generalizations about tech here because every company is different and will require different solutions for their domain.

At Yammer, we still have a massive Rails app, but we've started componentizing, and continue to add new services constantly.

We have a couple of cools ones. Feedie is a distributed feed storage system. It’s a 45 node cluster holding more than 10 billion messages.

We also have an interesting search stack. Flatterie is a service that handles denormalization and allows cursor based querying. This feeds into both our data export tool and our index building service Dexie. Dexie creates search and autocomplete indexes that are used by the search services Querie and Completie.

I probably don’t have time to discuss it, but I thought you might be interested -- Here’s a dependency diagram for services at Yammer.



Tokie - OauthToken based authentication
Rous - file uploads
Paddie - realtime document editing
Syncie - directory sync
Slurpie - data export
Completie - autocomplete
Dexie - index building services
Flatterie - Cursor based data denormalization
Querie - Search
Observatory - event aggregation
Prankie - Ranking service
Artie - Realtime delivery
Feedie - Message feed delivery / storage

SOA Setup Tips

- Internalize SOA
- Direct, Decentralize, Divide & Conquer
- Monitor Everything
- Limit dependencies
- Handle Failure Modes
- Keep it simple

Saturday, May 19, 2012

Internalize it. Make it part of the culture.
Have a unified global vision and execute it in a trusting, distributed fashion.
Know what's going on in your systems and in your ecosystem.
Limit dependencies wherever possible
Always try to return something useful to the user
Finally, and most importantly, make your systems as simple as possible, but not simpler.

<<PAUSE>>

yammer

is hiring

Mike Ihbe
Senior Infrastructure Engineer
twitter.com/mikeihbe
github.com/mikejihbe

Lastly, my shameless plug. Yammer is hiring and it's an awesome place to work.

That's me, find me online or come talk to me after, I'd be delighted to speak with you.

Thank you!

<<PAUSE>>