

Scriptable Infrastructures for Scientific Computing

2010 Amazon Web Services Genomics Event

Deepak Singh Hates Me.

- Why do I have to speak after James Hamilton?
- *Whatever I did man, I'm really very sorry.*

Who am I?

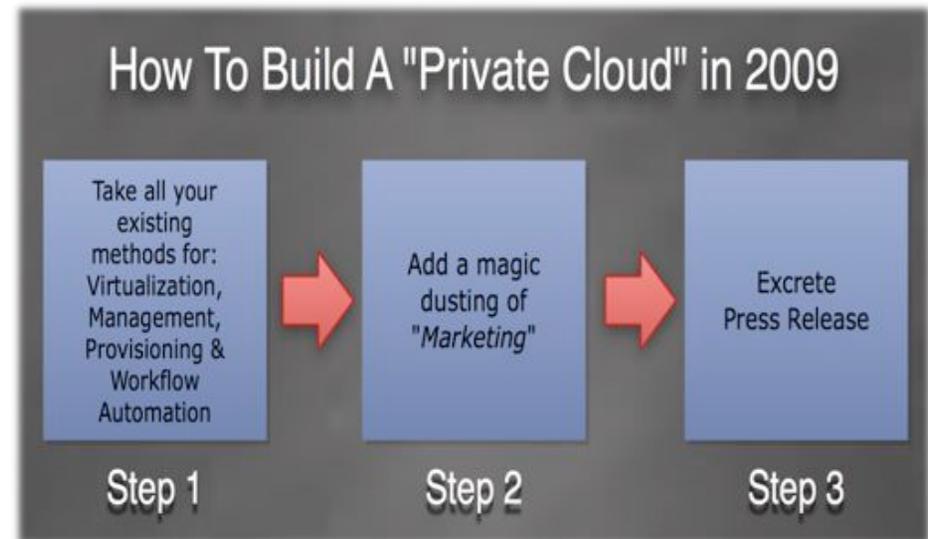
- I'm from the BioTeam
 - Independent consulting shop
 - Staffed by scientists forced to learn IT to get our own research done
- Found a fun business niche
 - Bridging the “gap” between science, IT & high performance computing
- This matters today because ...
 - We've been doing production informatics work on Amazon AWS since 2007
 - Can speak from multiple AWS perspectives (Customer, Developer, Integrator)



Scene from ancient history in a cloud-enabled world...

I am not planning to ...

- Explain what “the cloud” is
- Explain how useful “the cloud” can be in our field
- **Talk about how utterly silly “private clouds” are****
- Talk about why Amazon owns this space right now
- .. Each of these topics is either old news or could be a self-contained talk itself



*** I have seen this '09 graphic come to life in the real world. Psychic.*

Yet more things I plan to not talk about ...

- Mapping high performance computing (HPC) workflows onto cloud platforms
- Strategies, Reference Architectures & Best Practices
- Technical challenges & workarounds
- Horrible mistakes I've made and valuable lessons learned
- ... many of the upcoming speakers will be covering these topics (and more!) in detail

However if that stuff interests you ...

- Take advantage of the upcoming coffeebreak
- Perk yourself up for an interesting day
 - The agenda looks fantastic

{ or visit }

- <http://bioteam.net/aws>
 - AWS Training & Info
- <http://blog.bioteam.net>
 - Talks, Howtos, rants, etc.



Shady marketing trick: stick the shameless plug within the talk rather than the end ...

Topics For Today

- Laziness
- Beauty
- Careers
- Money



*The real reason I come to Seattle, shhhhhh
(mocha made by Kelli @ Inner Chapters Café & Bookstore)*

Laziness

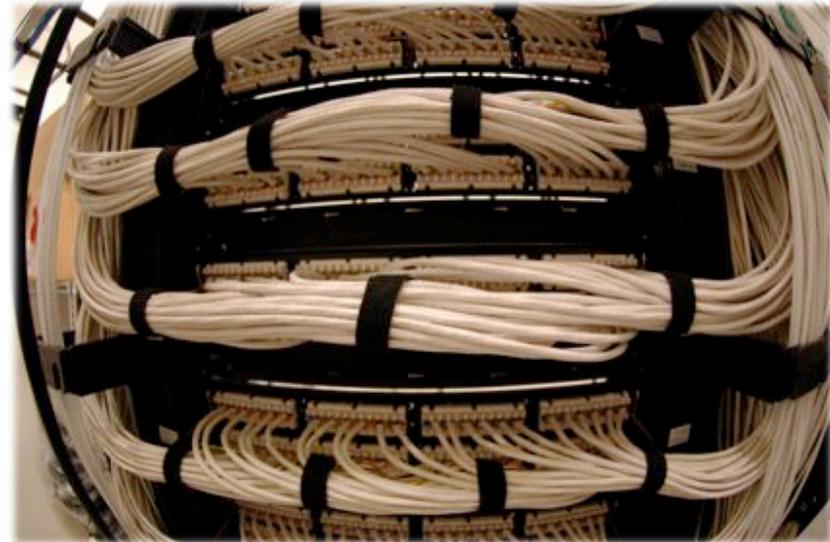
- Larry Wall's 1st Great Virtue:
 - *"... the quality that makes you go to great effort **to reduce overall energy expenditure. It makes you write labor-saving programs that other people will find useful, and document what you wrote so you don't have to answer so many questions about it"***
- Scriptable IT Infrastructures are the latest boon for the perennially lazy (like myself)



Note subtle Amazon product plug above ...

Beauty

- Call me a nerd but this stuff is amazing:
 - “Scripting my Datacenter(s)”
 - Orchestrating complex systems & workflows with a few lines of code
 - Infrastructure managed like source code



Can you believe that the Broad Institute @ MIT let me into their telco closets & machine rooms?

Careers

- Radical change ahead
 - Cloud Computing will rewrite many of our roles & job descriptions



Nerds under glass

Money

- If a cloud talk occurs without mention of \$
 - ... did it actually happen?

Greetings from Amazon Web Services,

This e-mail confirms that your latest billing statement is available on the AWS web site. Your account will be charged the following:

Total: \$101.68

Please see the Account Activity area of the AWS web site for detailed account information:

<http://aws-portal.amazon.com/gp/aws/developer/account/index.html?action=activity-summary>

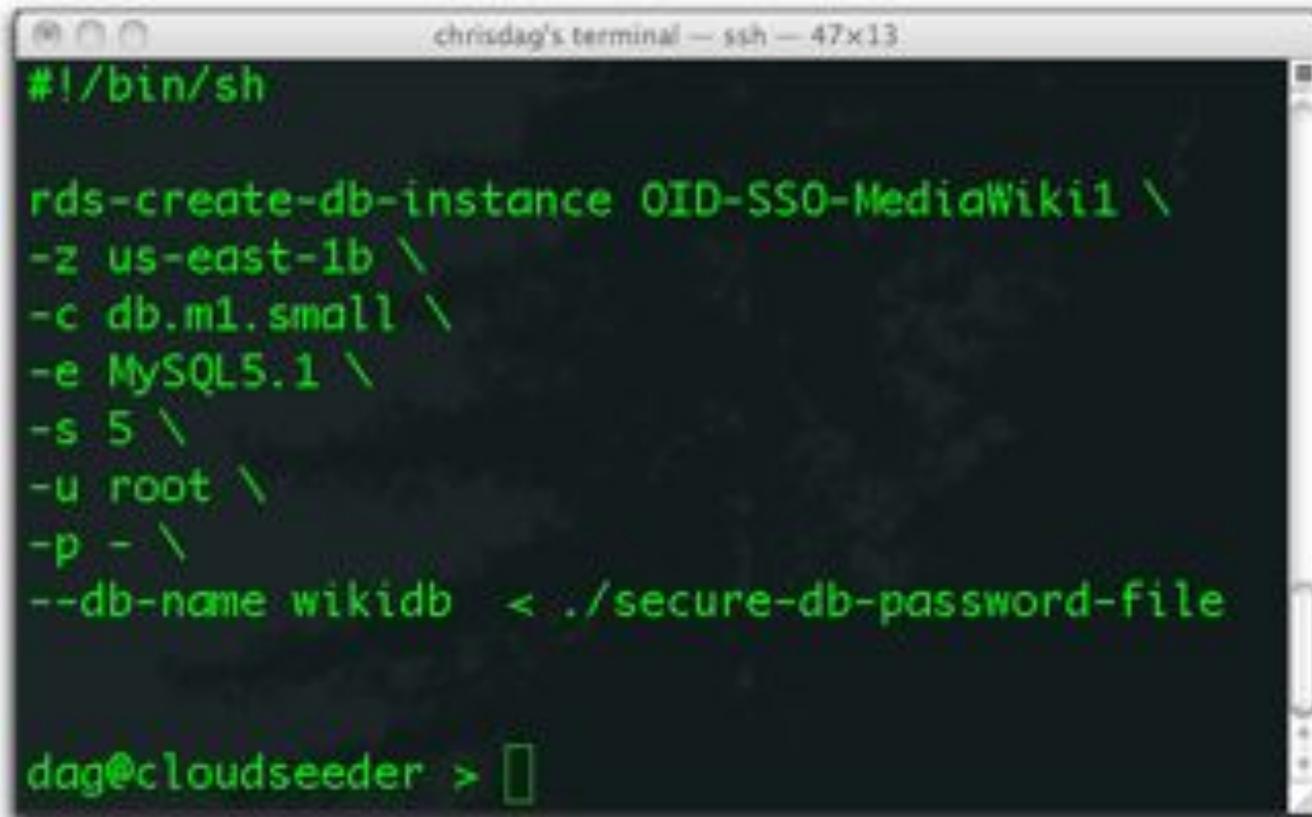
BIOTEAM

Enabling Science

Laziness
Laziness

chris@bioteam.net - <http://www.bioteam.net>

“Scriptable Infrastructure” is a BIG DEAL



```
chrisdag's terminal — ssh — 47x13
#!/bin/sh

rds-create-db-instance OID-SSO-MediaWiki1 \
-z us-east-1b \
-c db.m1.small \
-e MySQL5.1 \
-s 5 \
-u root \
-p - \
--db-name wikidb < ./secure-db-password-file

dag@cloudseeder >
```

This single command will start a 5GB managed MySQL database in the Amazon cloud for \$0.11/hour. The database is **automatically** patched, managed and backed up. Planned enhancements include auto-scaling & snapshots.

It's ALL scriptable ...

- Servers
- Storage
- Operating System(s)
- Network
- Provisioning
- Management
- Monitoring & Scaling
- Accounting

Not hype. Real.

- Every facet of our IT infrastructure can now be automated and remotely controlled via simple scripts and API calls
- Benefits go way above and beyond simple IT Operations work, server “lights out management” features and what local VM systems provide
- *Nirvana for lazy nerds like myself*
 - *Concentrate on getting my work done, not babysitting my datacenter racks*



```
top - 10:10:40 up 22:21, 1 user, load average: 0.10, 0.16, 0.18
Tasks: 102 total, 1 running, 101 sleeping, 0 stopped, 0 zombie
Cpu(s): 0.0% us, 4.0% sy, 0.0% ni, 96.0% id, 0.0% wa, 0.0% hi, 0.0% si
Mem: 2523776k total, 2507368k used, 16408k free, 41660k buffers
Swap: 589816k total, 208k used, 589608k free, 1984196k cached

  PID USER      PR  NI  VIRT  RES  SHR  S  %CPU  %MEM     TIME+ COMMAND
 20252 dag        16   0 6172 1140 836  R  4.0   0.0   0:00.08 top
   1 root        16   0 4772  640 536  S  0.0   0.0   0:02.75 init
   2 root       34  19   0     0  0  S  0.0   0.0   0:00.04 ksoftirqd/0
   3 root        5 -10   0     0  0  S  0.0   0.0   0:01.11 events/0
   4 root        5 -10   0     0  0  S  0.0   0.0   0:00.05 khelper
   5 root        7 -10   0     0  0  S  0.0   0.0   0:00.00 kthread
   6 root       13 -10   0     0  0  S  0.0   0.0   0:00.00 kacpid
  86 root        5 -10   0     0  0  S  0.0   0.0   0:32.12 kblockd/0
  87 root       15   0   0     0  0  S  0.0   0.0   0:00.00 khubd
 110 root       20   0   0     0  0  S  0.0   0.0   0:00.00 pdflush
 111 root       15   0   0     0  0  S  0.0   0.0   0:13.78 pdflush
 112 root       16   0   0     0  0  S  0.0   0.0   0:28.31 kswapd0
 113 root        7 -10   0     0  0  S  0.0   0.0   0:00.00 aic/0
 259 root       24   0   0     0  0  S  0.0   0.0   0:00.00 kseriod
 492 root       19   0   0     0  0  S  0.0   0.0   0:00.00 scsi_eh_0
 514 root       15   0   0     0  0  S  0.0   0.0   1:20.84 kjournald
1228 root        6 -10   0     0  0  S  0.0   0.0   0:00.00 kauditd
```

So easy an Ipad can control it.

Scriptable Infrastructure

- For the first time some of our IT infrastructure might be 100% virtual and entirely controllable via scripts and APIs
- It's not rocket science
- Anyone can drive this stuff
 - Especially motivated researchers



BIOTEAM

Enabling Science

Beauty

Scripted Infrastructure – Creating AMI Bundles

```
ec2-bundle-vol -d /mnt \  
-k /root/.ec2/pk-6LGHW [REDACTED] WHU7.pem \  
-c /root/.ec2/cert-6LG [REDACTED] HSWHU7.pem \  
-u 6099-7144-1117 \  
-r i386 -p 32bit_ManualChef
```

BO

Scripted Infrastructure – Upload bundle to S3 bucket

```
ec2-upload-bundle -b cloudtraining/32bmanchef \  
-m /mnt/32bit_ManualChef.manifest.xml \  
-a AKIAJFXS50PTB52QFJFA \  
-s BLr2w [REDACTED]0xp+GaA5/5dv
```

BORI

Scripted Infrastructure – Register & receive new AMI ID

```
ec2-register -d "32bit CentOS, BioTeam Chef Managed Server (Manual Registration)" \  
-n "32b_chef_manreg" \  
-K /root/.ec2/pk-6L[REDACTED]HU7.pem \  
-C /root/.ec2/cert-f[REDACTED]SWHU7.pem \  
-a "i386" \  
cloudtraining/32bmanchef/32bit_ManualChef.manifest.xml
```

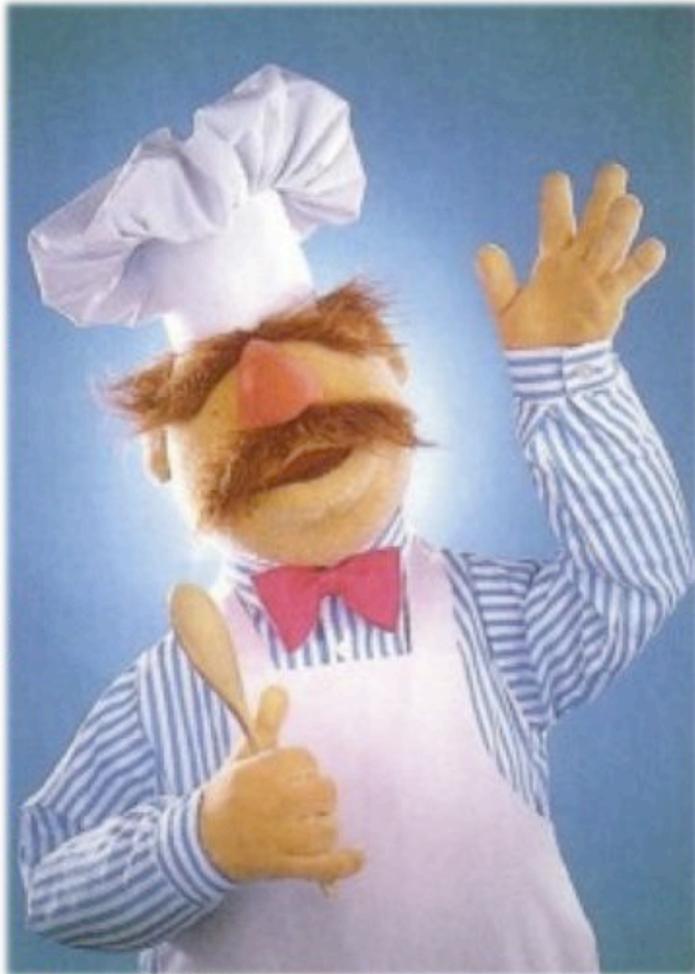
BORING.

Beauty

- “Scriptable Infrastructure” is just the baseline
 - The cool stuff happens when we build on top of these capabilities
- AWS enables us to **orchestrate** vast arrays of complex systems, pipelines, workflows & applications
 - *Without leaving the hammock*
- *Orchestrated systems working in concert are a beautiful thing.*



Systems Orchestration



- **Bork!**
- OpsCode Chef
 - <http://www.opscode.com>

Chef is

At a high level ...

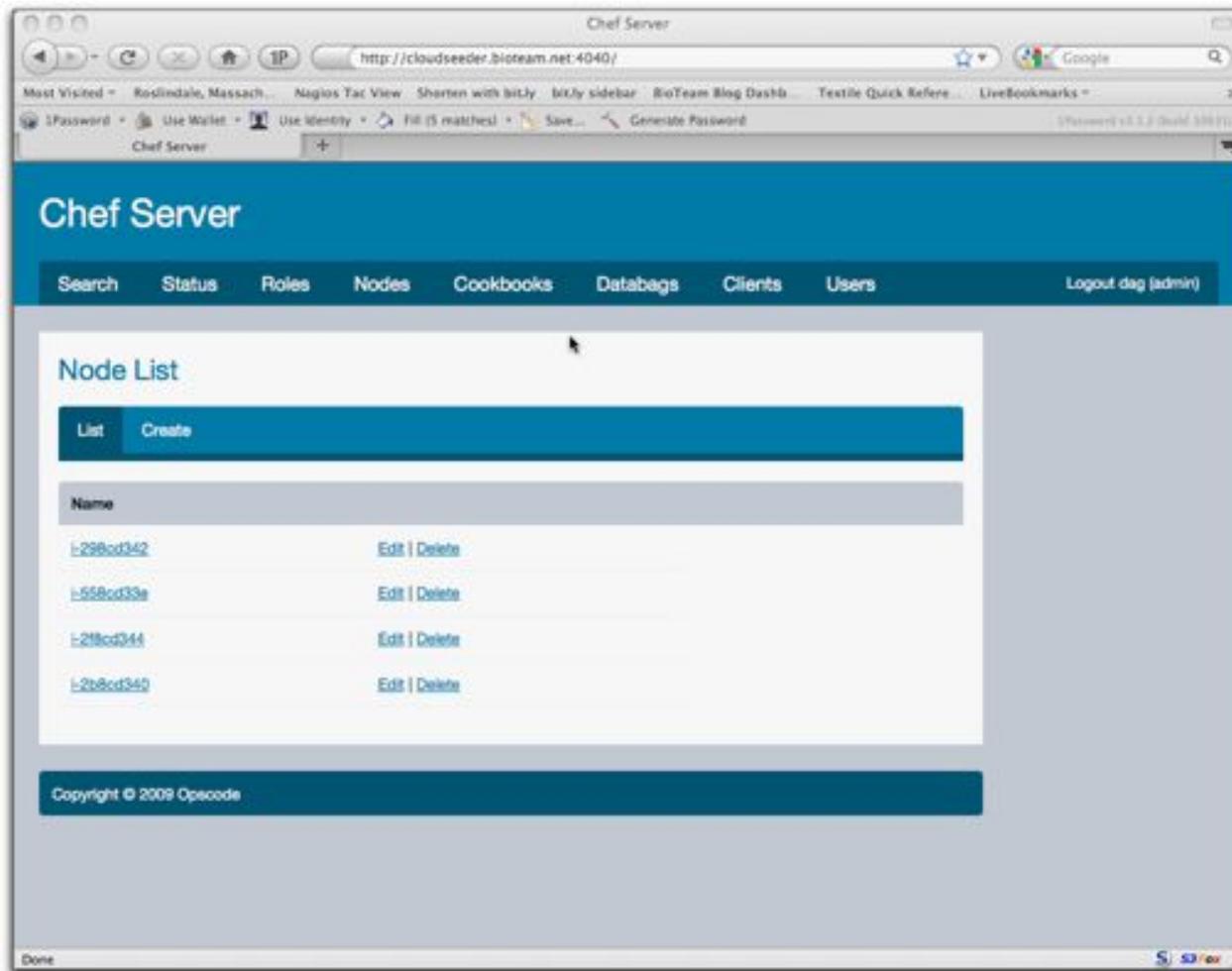
- Library for configuration management
- A standalone configuration management system
- A full featured systems integration platform
- **... natively aware of cloud platforms and cloud instance metadata**

Cooking with Chef

Why?

- New product built from years of CM experience
- Experience & techniques we can leverage
 - 'Puppet' developer spent years with 'cfengine'
 - Chef developers spent years as 'puppet' consultants
- And at 2010 BioITWorld Cloud Workshop
 - ... first three speakers mentioned that they used Chef - totally unplanned & unexpected

Chef Server Web UI



As a result of Chef ...

BioTeam now manages only 4 Linux AMI's in the cloud

- 32bit Linux CentOS & Debian
- 64bit Linux CentOS & Debian
- ... using knife, chef-solo or Chef Server we can orchestrate AWS services into whatever we need in a matter of minutes.

Chef understands cloud metadata

Attributes

Attribute	Value
▶ block_device	
▼ cloud	
▼ private_ips	
0	10.242.110.3
provider	ec2
▼ public_ips	
0	67.202.56.58

Snippit of our Grid Engine recipe

Cookbook: gridengine

Attribute Files

Recipe Files

default.rb

```
# Adam Ereut
# Version 1.0
#
# Goal:
#   Install and auto-configure a functional
#   Grid Engine 6.2u5
#

hosts = Array.new
localhost = nil
qmaster = nil

# search for all nodes in this ec2 reservation
nodes = search(:node, "reservation_id:${node[:ec2][:reservation_id]}")

nodes.each do |n|
  # node own's record, store in localhost
  if n.name == node[:name]
    localhost = n
  end

  # use launch index to determine qmaster
  if n[:ec2][:ami_launch_index] == "0"
    Chef::Log.info("Setting qmaster to #{n[:ec2][:public_hostname]} for this reservation")
    qmaster = n
  else
  end
end
```

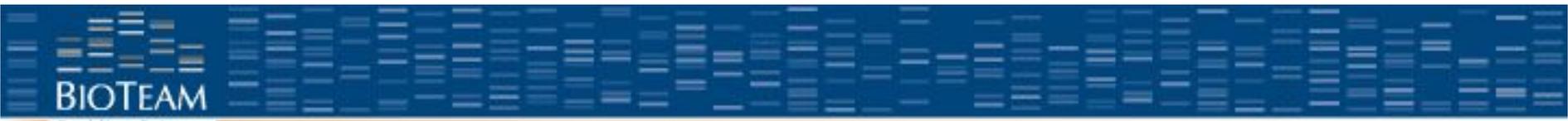
Chef lets you ...

Treat your infrastructure as code

- Manage configuration as idempotent **resources**
- Put resources together as **recipes**
- Group recipes into **roles**
- Track it all like **source code**
- **Configure** your **systems**

Systems Orchestration: Wrapup

- AWS gives us **on-demand resources & elasticity**
- “Scripted Infrastructures” give us **agility**
- Chef orchestrates & delivers **capability**



BIOTEAM

Enabling Science

Careers

chris@bioteam.net - <http://www.bioteam.net>

Aka “Career Politics & Scriptable IT”

- *What happens to IT roles when anyone with a web browser can instantly launch (and manage) a complex cluster, software pipeline or massive database?*
- Radical restructuring of the lines between
 - Research staff & Investigators
 - IT Operations Staff
 - IT Support Staff

“Career Politics & Scriptable IT”

- Something you should address sooner rather than later – seems clear that roles & responsibilities might change significantly
 - Erosion of the traditional lines between IT & Research
- Friction can also lead to issues when attempting true-cost accounting
 - “... did he really forget to include the cost of the datacenter diesel delivery contracts & ongoing service on the HVAC system?”

“Career Politics & Scriptable IT”

- My \$.02
 - This is a good thing
 - End result is that IT professionals spend less time “keeping the lights running” and more time enabling/assisting with the actual goals of the organization
- Maybe this will happen:
 - Research end users become the system admins
 - IT professionals become the system orchestrators & architects
 - Excellent career news for hybrid Science/IT types

BIOTEAM

Enabling Science

Money
WOLFGANG

Money

For anyone seriously looking at IaaS Cloud Platforms

- Can't escape it
- Critical to have a solid understanding of the financial issues

Money

- IaaS is all about leveraging economies of scale
 - Providers can charge far less than your own true internal costs while still making a profit
- **If you don't honestly assess and track the true cost of providing IT services internally, you won't have the data you need to make informed decisions**
 - ... also more exposed to manipulation by people with personal agendas to push

Money: Cost-of-service Accounting

Amylin Pharmaceuticals Example:

- Hired a consultant specializing in IT cost guidance
- Huge effort to construct a spreadsheet that tracked the real cost of delivering each IT service
 - Result: When compared against actual budget, the spreadsheet was accurate to within \$2K out of a 20M budget
- Incredible benefits from this data
 - Realized it cost \$2M/year to run HR internally
 - Many IT staff simply “keeping lights running” and not driving business or scientific success
 - Information shared widely, senior managers really did not like being associated with the most expensive services
 - ... lead to organizational changes & operational methods that deliver huge recurring savings

end;

Questions?

Comments/feedback welcome
<chris@Bioteam.net>

Plug: AWS Fundamentals for Science & Engineering
<http://bioteam.net/services/cloud-training.html>