Thanks for having me …

For me … some interesting cloud integration suites to investigate further …
http://www.snaplogic.com/what-we-do/integration-platform/snapcenter.php
http://integration.pervasive.com/UsageScenarios/ApplicationIntegration/CloudIntegration.aspx
http://www.boomi.com/resources/whitepapers


Richard Seroter is the Lead Architect for the R&D division of Amgen, a Microsoft MVP, blogger, author, trainer for Pluralsight and contributing editor to InfoQ.

Who I am …

[Where are we]

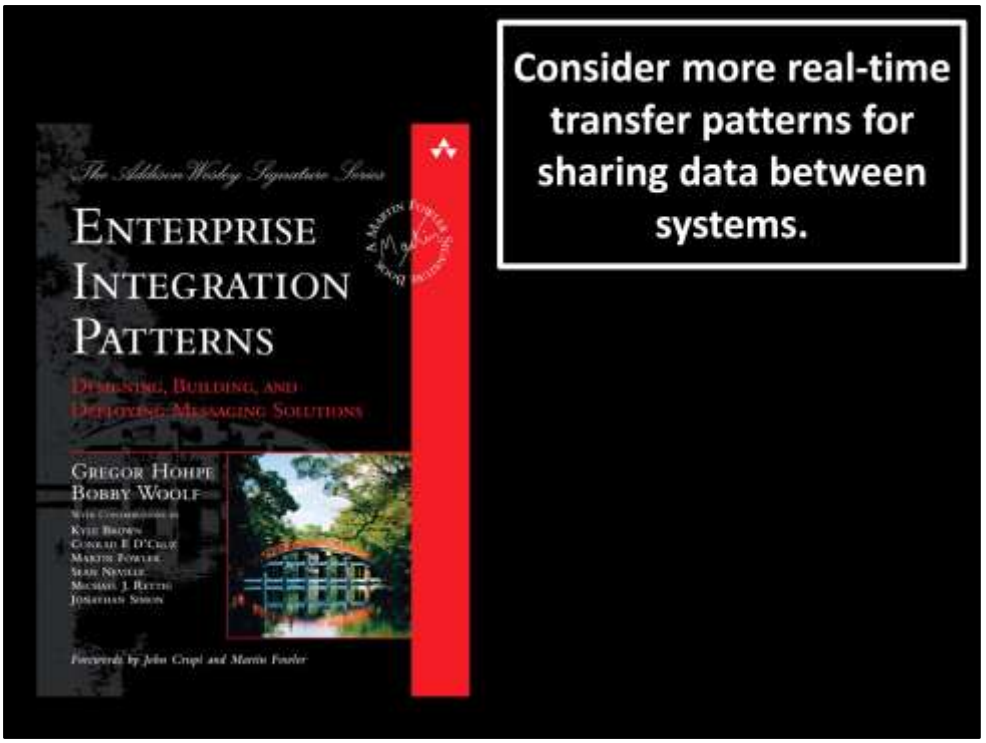
- Increased popularity towards using packaged software as a service or building apps in cloud platforms
- More than 95 percent of organizations expect to maintain or increase their investments in software as a service (SaaS) and more than one-third have migration projects under way from on-premises to SaaS, according to a survey by Gartner, Inc.
    - Currently, communications (52 percent), utilities (51 percent), and banking and securities (49 percent) industries rank highest with respect to SaaS deployed across the horizontal and vertical-specific categories sampled. In 2012, those industries ranking highest with respect to their plans to use SaaS include federal government (33 percent), banking and securities (22 percent) and wholesale trade (20 percent). Beyond 2012, top industries considering SaaS are manufacturing and natural resources (37 percent), wholesale trade and retail (each 29 percent).

[What's the problem?]

- SaaS not really built with data exchange in mind
    - Information Week survey (http://www.informationweek.com/news/cloud-computing/software/231002362?queryText=SaaS+Integration) of SaaS satisfaction areas showed "ease of deployment" #1 and integration last
- Easy to see these as "off site" and either do limited integration, or, irregular batch loads
- Three scenarios: on-premise to cloud, cloud to on-premise, and cloud to cloud
    - Each have unique challenges, but each are valuable and necessary scenarios

[where do we want to be?]

- There are tried and true integration styles for sharing data/processes between systems
- We're taking that reliable chassis and applying it to new technologies



[how do we get there?]
A seminal book on the topic is the EIP book by Hohpe and Woolf
Discusses a handful of communication styles and integration patterns. We'll focus on three that are called out there:

- Shared database
- Remote procedure invocation
- Messaging / async

For each pattern …

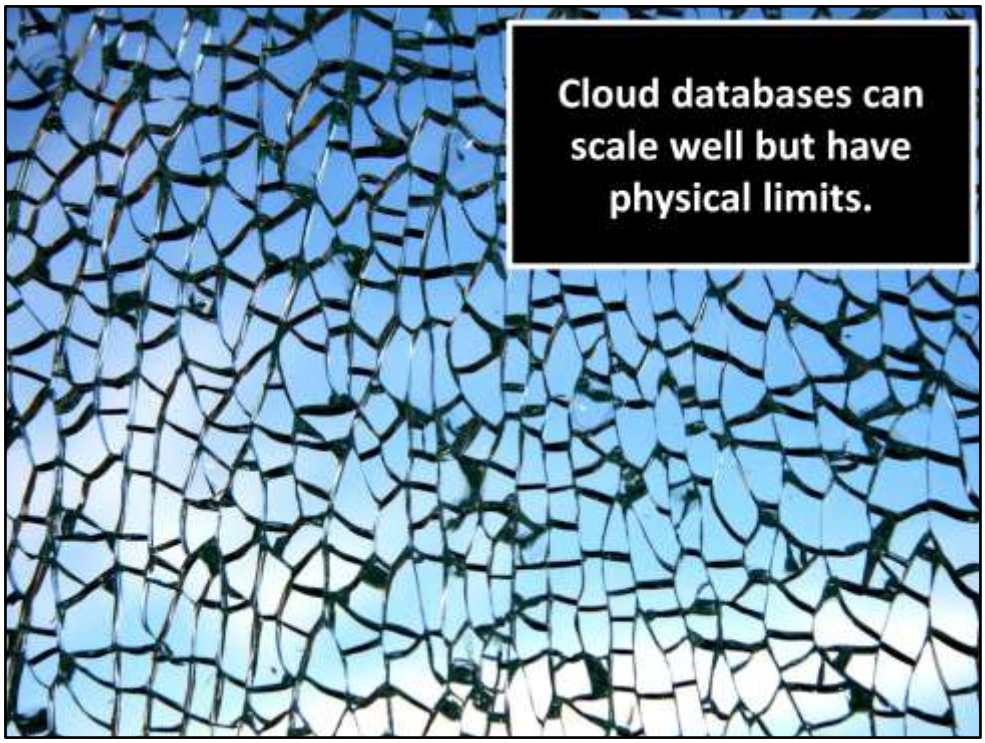
- Define it
- When it makes sense
- Constraints
- Cloud considerations
- Cloud technologies
- Demo

When it makes sense …

- Sharing via ETL or file isn't timely
- What if you need the latest data?
- Could be for reporting, reference data or transactions
- Use when you have common data (and/or structure) but different CONSUMING interfaces
    - Think of multi-tenant apps where you share a DB, but partition by user
    - Could be an ODS or data mart where you want common schema used by others
- Get a single view of the data (no need to force each end to define a data format)
- All dependent systems are consistent at the same time
- Single data access strategy (SQL syntax)

Challenges

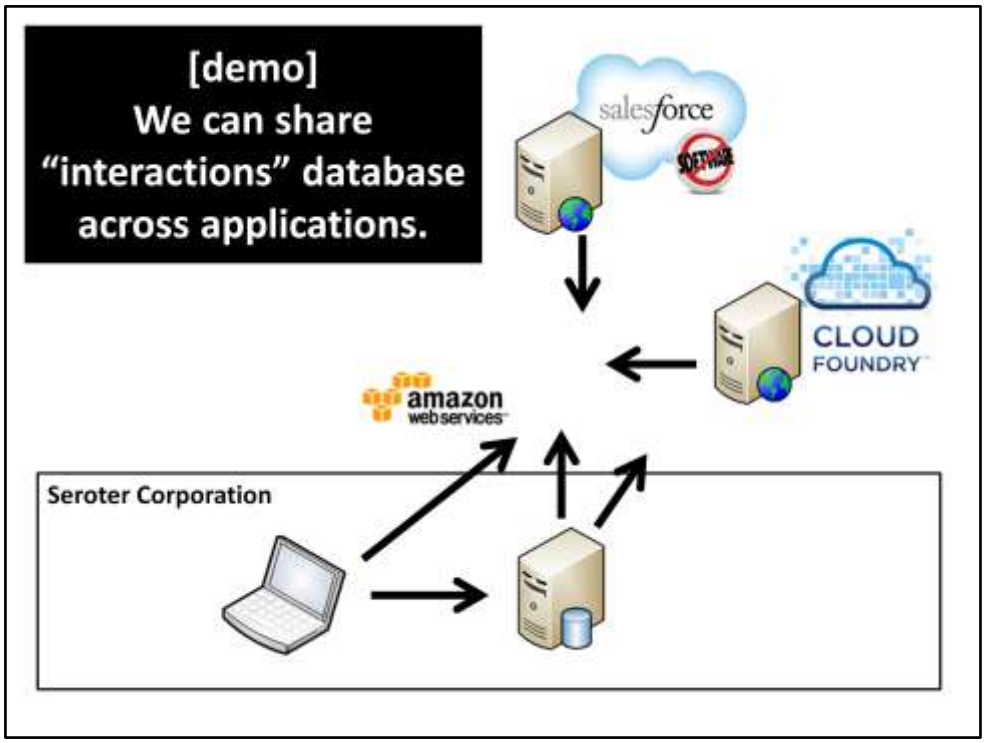
- ** Design **
- Tough to agree on formats (for ODS, mart, transactional records)
    - Unified schema that satisfies everyone? Commodity only, or, strong leadership to form a standard (or extensibility)
- ** Contention / Performance **
- Less likely to use when have multiple apps manipulating same transactional data
    - Could have data records defined where you could change parts of one record while someone changes parts of others
- May get fewer inconsistencies, but still have issues of simultaneous updates
    - Can try to handle with transactions, but transactions are often the enemy of scalability
    - http://www.ics.uci.edu/~cs223/papers/cidr07p15.pdf
- Try to avoid updates / conflicts between applications sharing the data
    - New rows only, or read only (reporting scenario, or reference data)
    - Avoid deadlocks
- Could get poor performance if apps are distributed and all accessing over WAN
- ** COTS support **
- Packaged apps rarely accept an external database as its source
- Opposite of “Shared nothing” where nodes are self sufficient http://en.wikipedia.org/wiki/Shared_nothing_architecture
    - May mean sharding http://www.codefutures.com/database-sharding/



Cloud Considerations

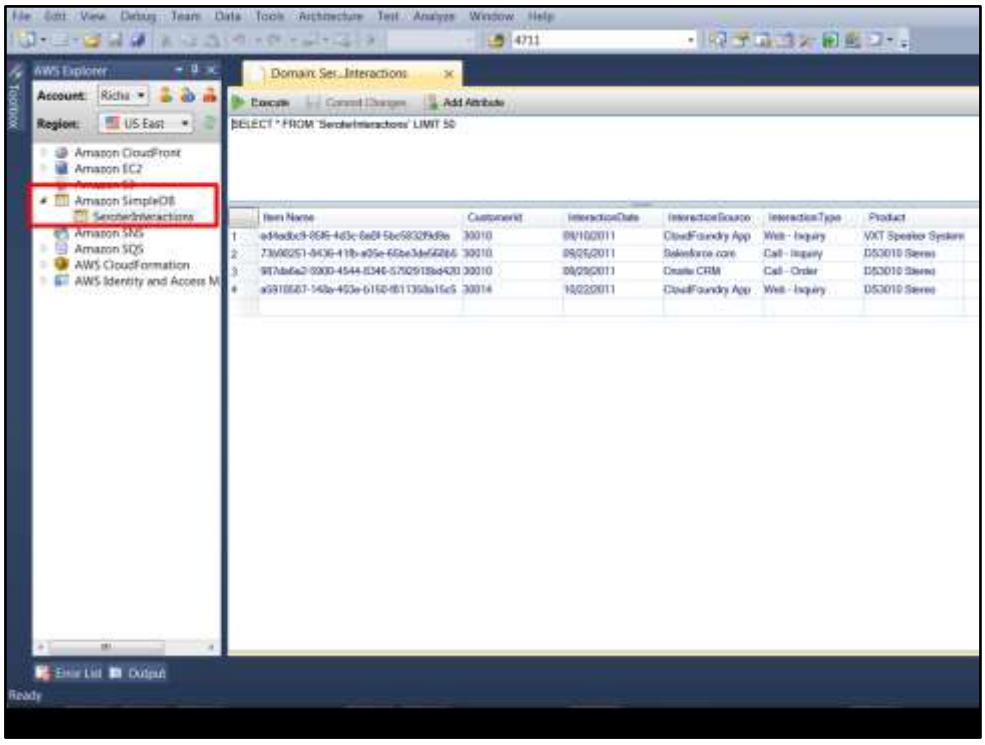
- ** Access protocols **
- Cloud provides either DB or web protocol access
- Have choices as to how to access
    - If remediating existing apps, may want the option of using “standard” ODBC APIs and not have to change much code
    - For newer (or rebuilt) apps, many services provide a RESTful API for data access
- ** Identity **
- Identity providers
- If shared, how do you apply granular access?
- ** Performance **
- May have good performance
    - Sharding built in to most of these (split based on domains/groups)
    - In some cases have eventual consistent reads, but SimpleDB supports option to have consistent reads http://aws.amazon.com/articles/3572?_encoding=UTF8&jiveRedirect=1
    - Conditional PUT and DELETE with expected values (to prevent overwriting fresher data)
        - Can still do eventually consistent read and trust the Conditionals to enforce consistency
        - Could use version or timestamp value as part of Conditional

- ** Different DB storage options **
- Could use cloud RDMS if you don't have massive scale needs
- Can use a schema-less product like SimpleDB
- ** Provider limits **
- Writes are throttled to SimpleDB, so can use sharded domains if expect more than XYZ puts/second (do batch)
- http://practicalcloudcomputing.com/post/712653349/simpledb-essentials-for-high-performance-users-part-1
- May have limited transactions
    - Across tables/entities for schema-less like Azure tables
- ** Options **
- DBs
    - Structured RDMS
        - Amazon RDS
        - SQL Azure
        - Database.com
        - Database hosted on IaaS platform like AWS
    - Schema-less
        - SimpleDB
        - Azure Tables
        - Google AppEngine Data Store http://code.google.com/appengine/docs/python/datastore/overview.html
            - Optimistic concurrency
            - Distributes data when necessary
            - Limits in number of calls per minute
    - Blobs
        - S3
        - Azure Blob Storage

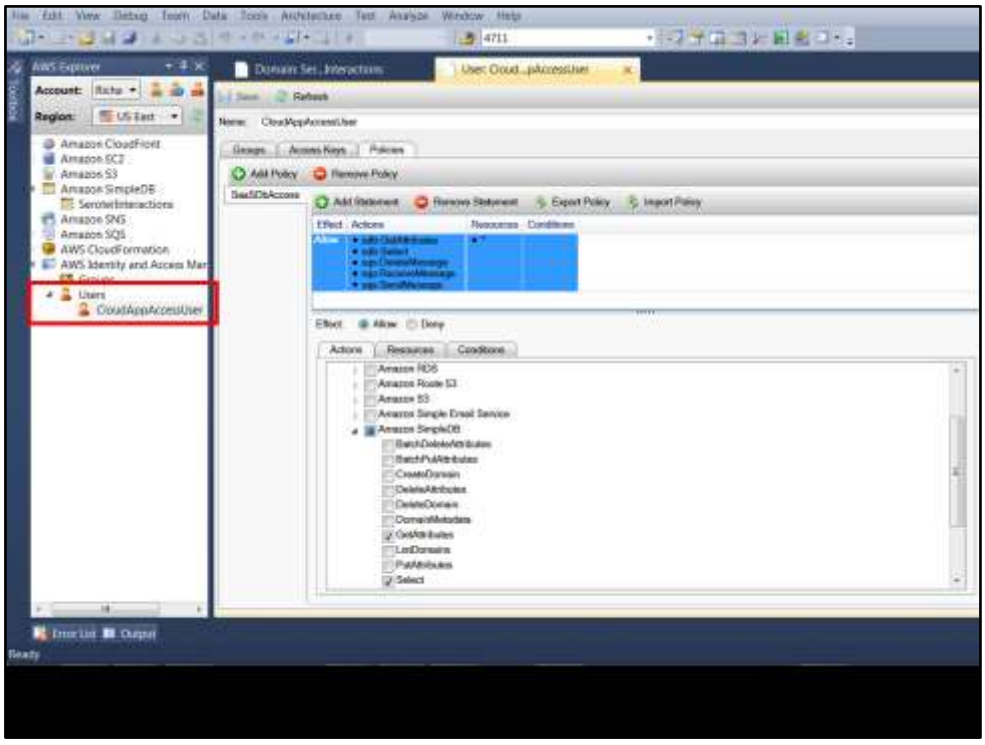


Demo

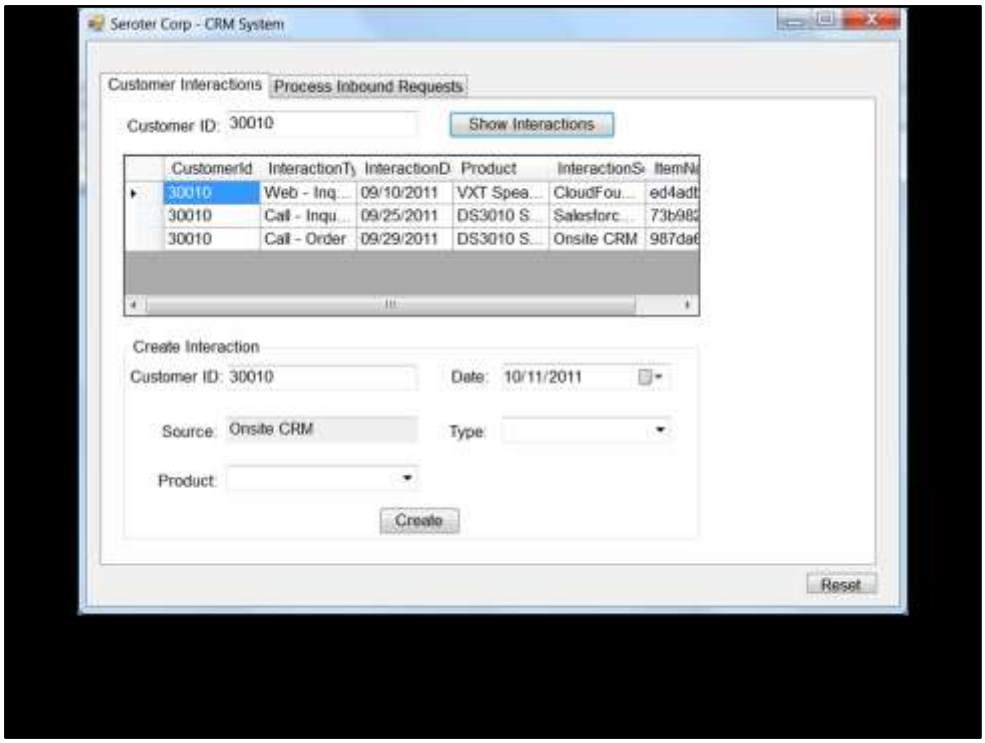
- Slide shows "before and after"
- Demo start …
  - Show db values from VS 2010; Show IAM and credentials specific to a user account
- .NET app
  - Retrieve items; Open app and don't put customer ID in; then put just the 30010 account in
- Cloud Foundry
  - Ruby app where I query the database via REST after building signed string
  - http://seroter-cloudintegration.cloudfoundry.com/lookup/30010
  - http://seroter-cloudintegration.cloudfoundry.com/lookup/30014
  - Show Ruby class; VMC to show running app instances
- SFDC
  - Custom code that looks at shared DB
  - https://c.na11.visual.force.com/apex/InteractionHistory?id=003A0000001YNPe (30010)
  - https://c.na11.visual.force.com/apex/InteractionHistory?id=003A0000001YycD (30014)
  - Show SFDC code
- .NET app
  - Add new item; Refresh SDFC and Cloud Foundry apps



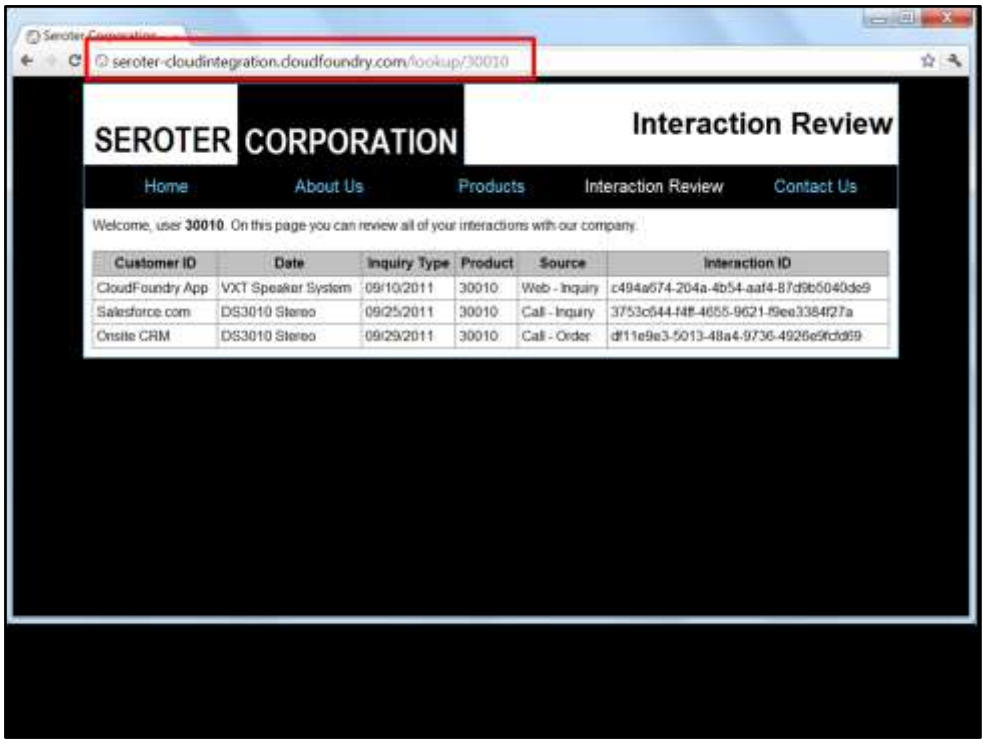
Can view my Amazon Web Services (AWS) SimpleDB content via Visual Studio 2010 plugin. Notice that I have four rows of data for two different customers.



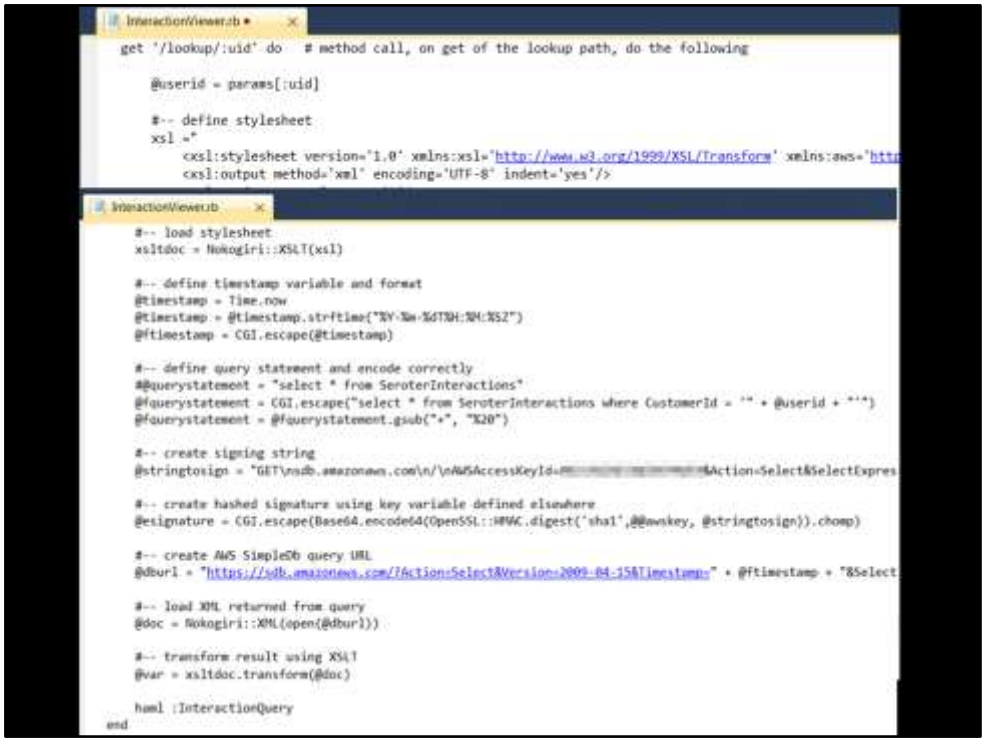
I can also do granular role-based permissions in AWS and restrict who can create/update/read data



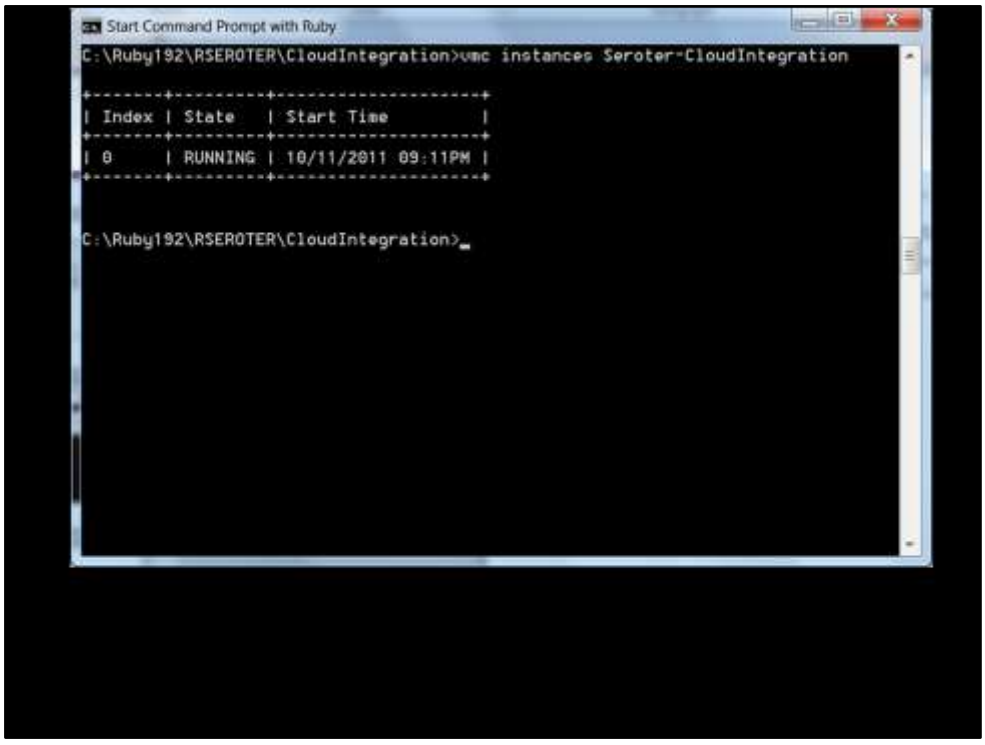
I built a client application that pulls all the interactions from AWS SimpleDB for a given user



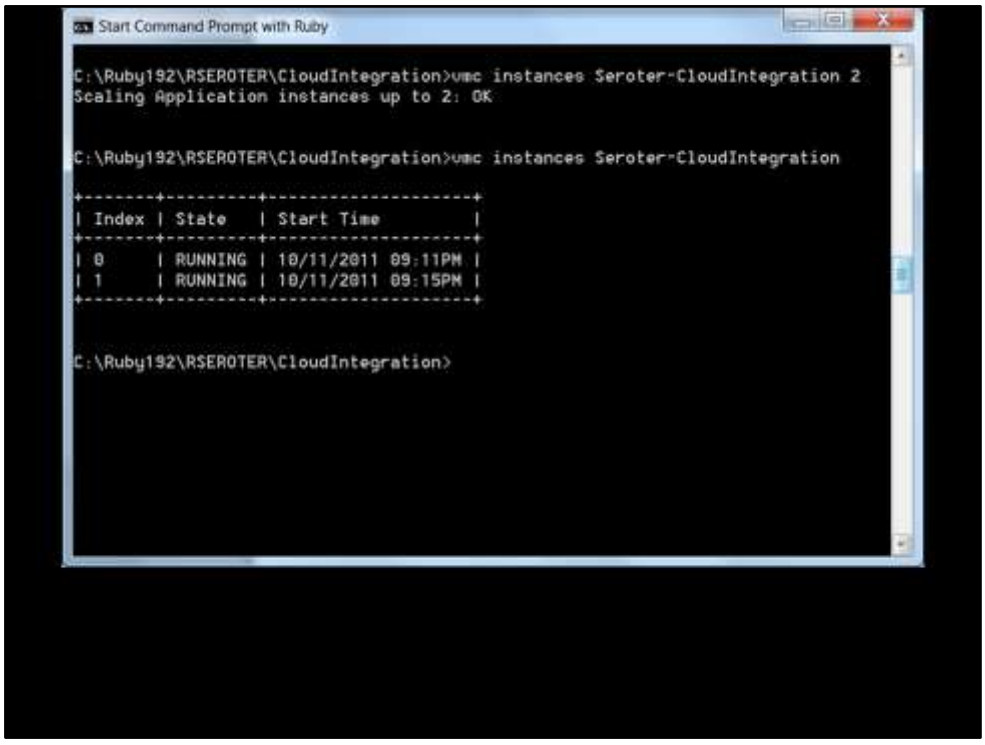
I've also built a Ruby web application hosted in VMWare's public/private cloud called Cloud Foundry



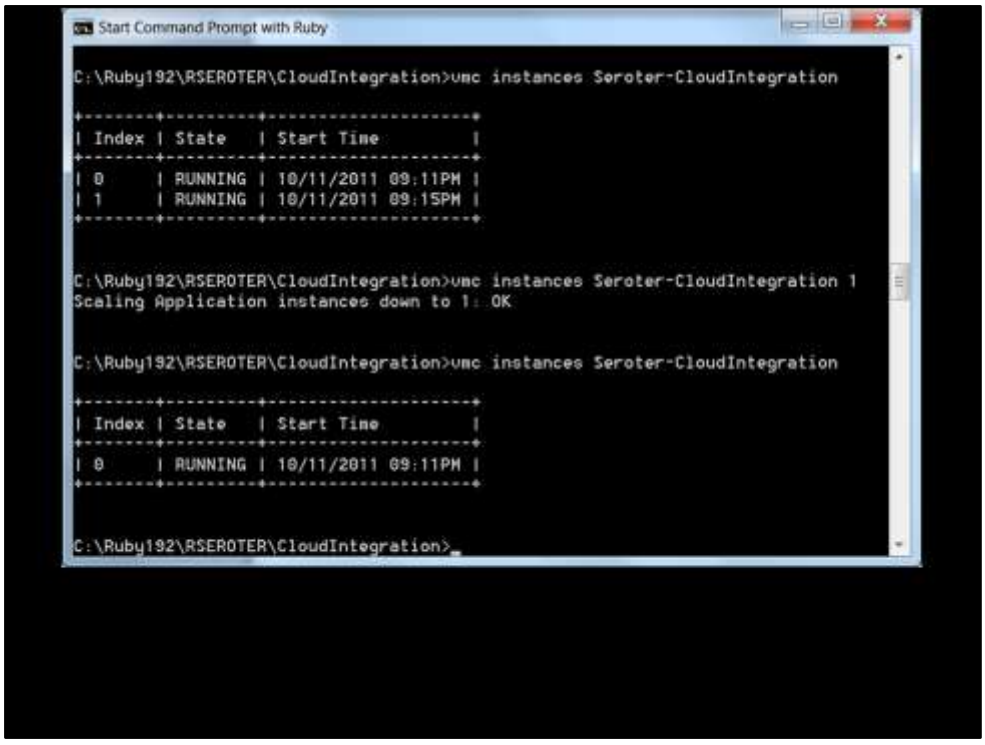
My Cloud Foundry Ruby code consumes the AWS SimpleDB data via web services



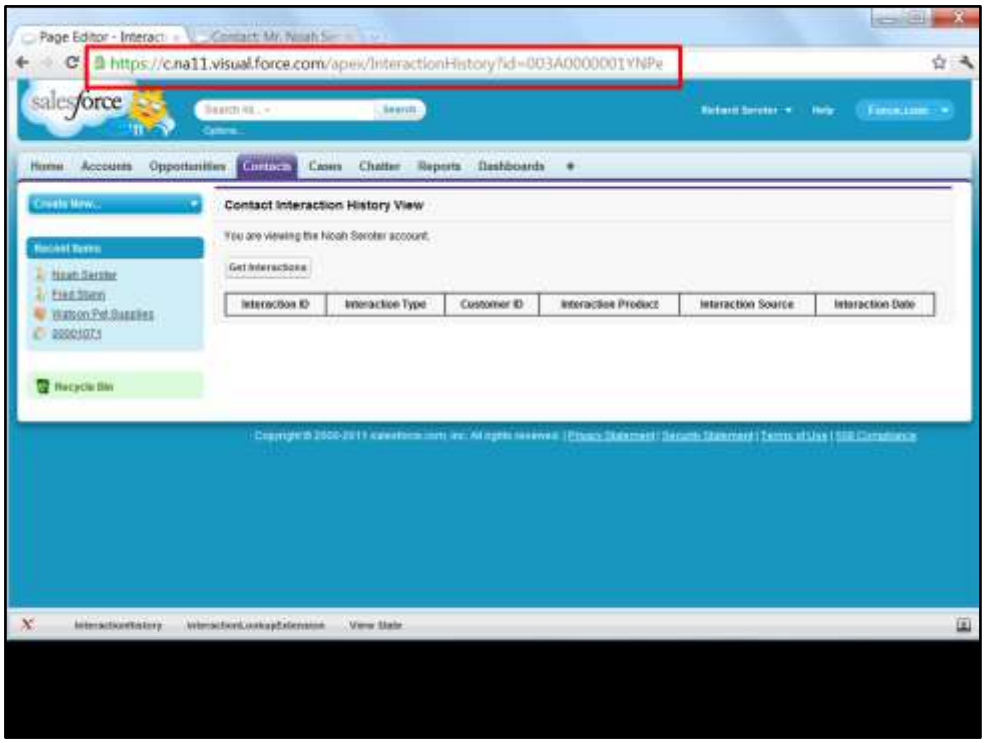
You can see how many instances of my web app are deployed in the VMWare cloud



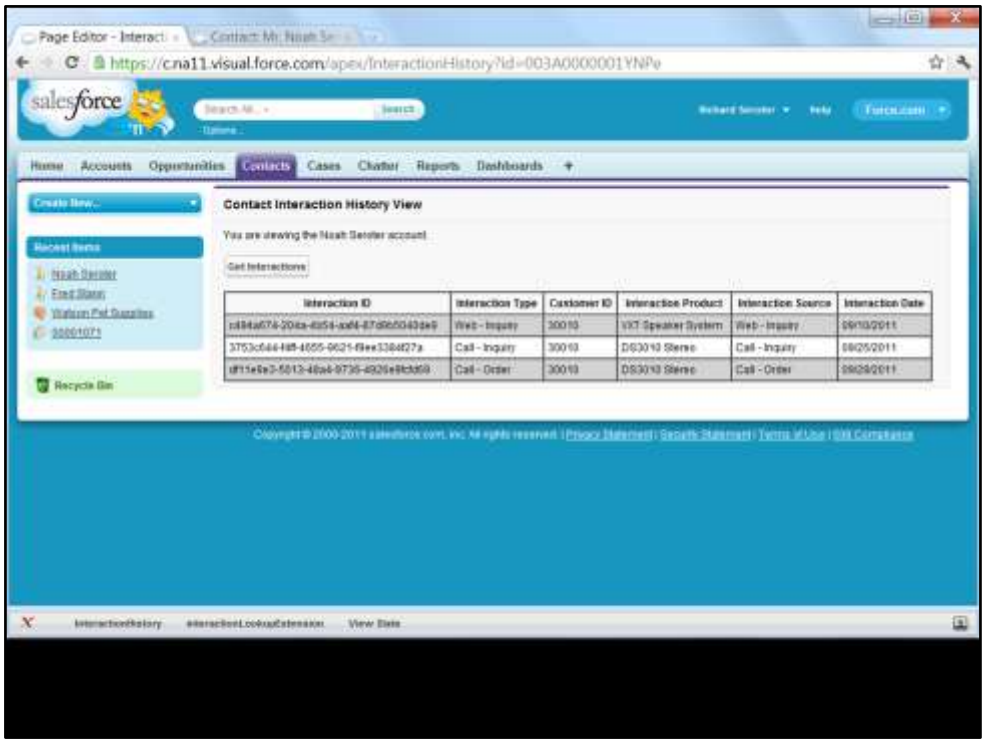
With one command, I *instantly* jump to two instances. Immediately load balanced.



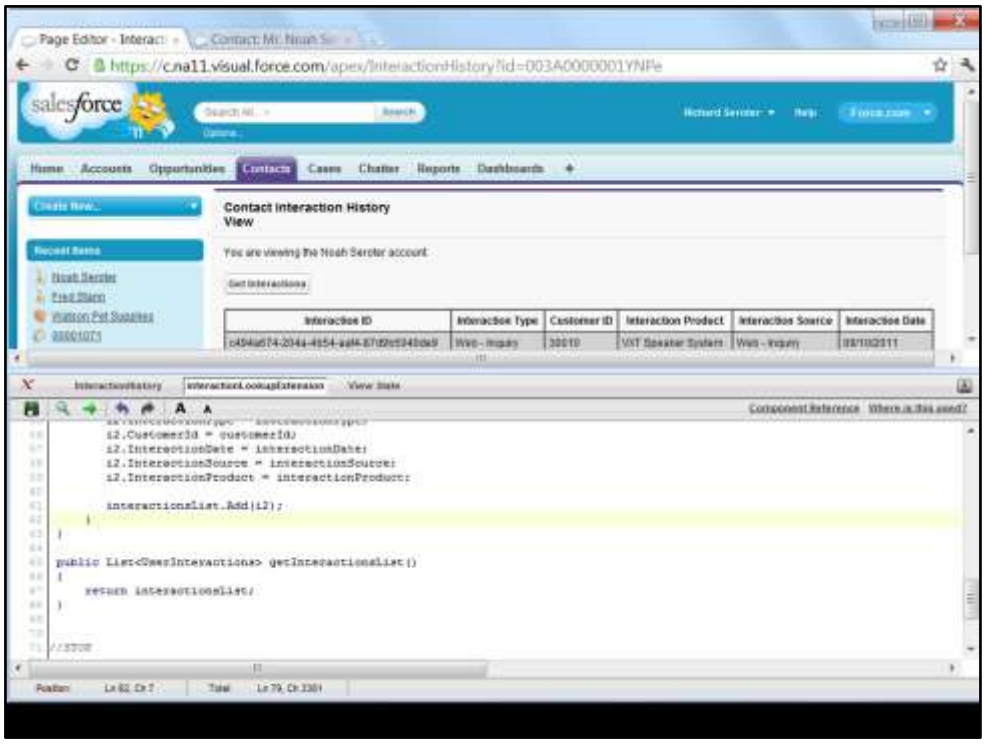
I can then *instantly* move back to a single instance.



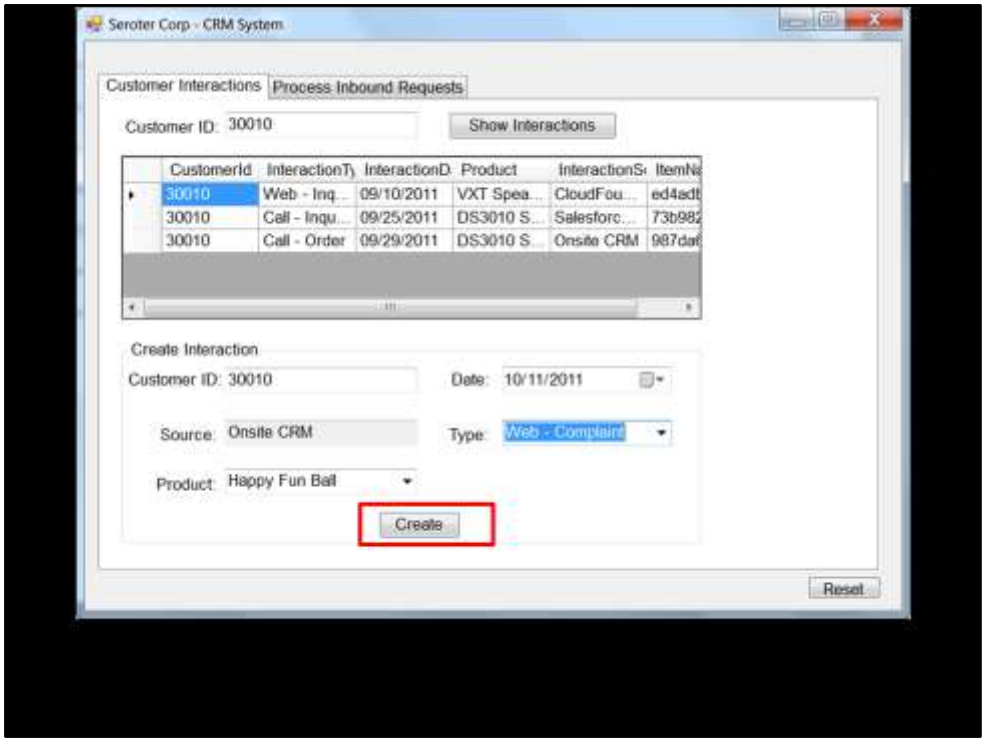
I'm now in my Salesforce.com account where I built a custom page to pull SimpleDB data



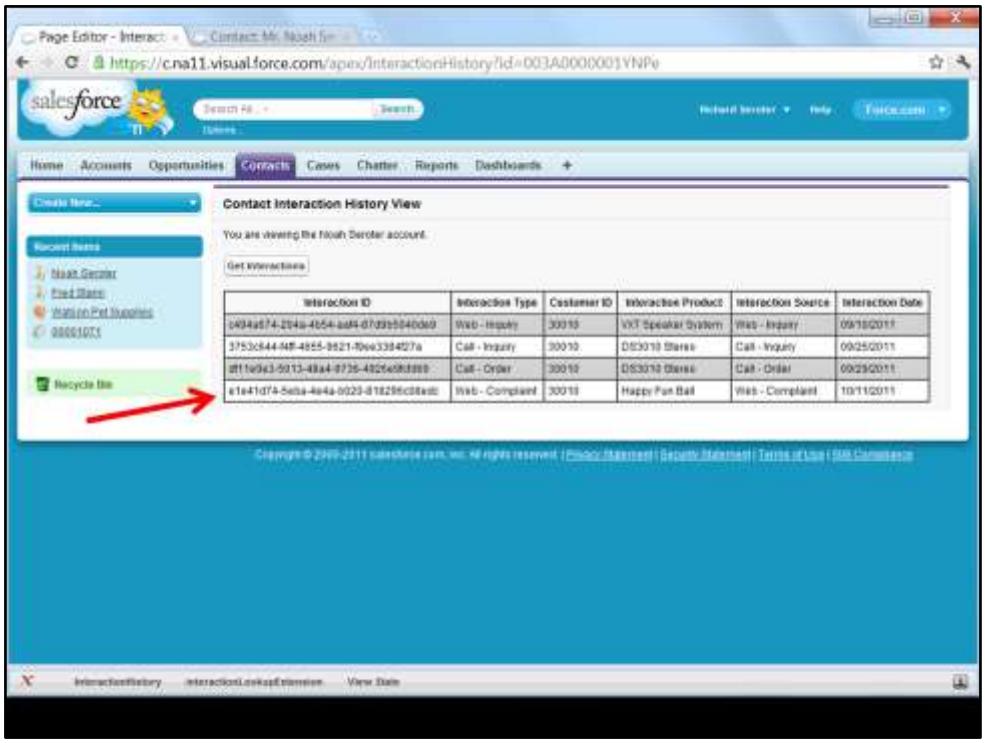
By clicking the button on my page, I retrieved all the interactions for this Salesforce.com customer



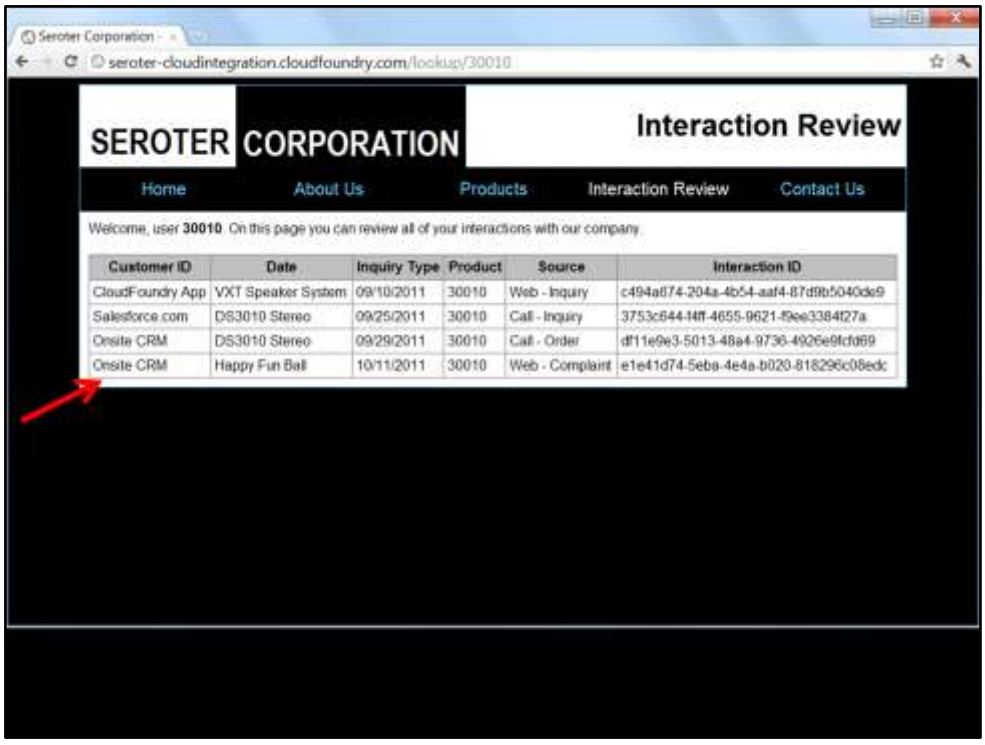
My code show that I consume the AWS SimpleDB service in a very similar way to Cloud Foundry app



Within my on-premises app, I added a new record to my SimpleDB database



That row is now immediately visible in Salesforce.com



… and within my Cloud Foundry app!

When it makes sense …

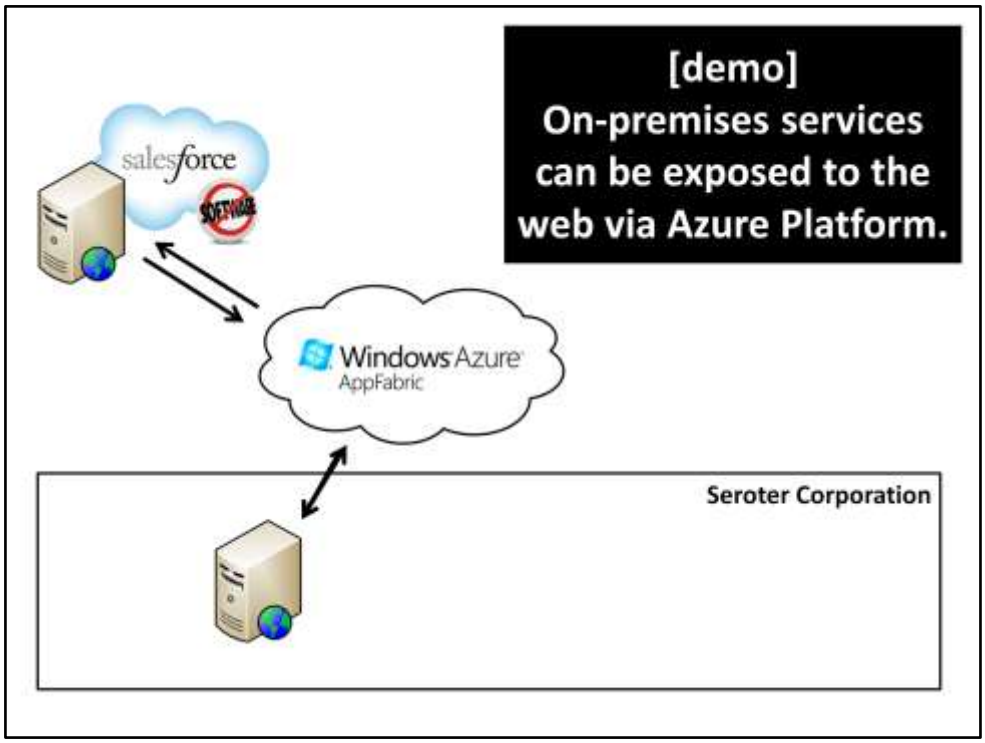
- Have multiple apps, built independently and data/processes need to be shared
- Just sharing data not always enough; processes, workflows too
- Data stays with it's source
- SOA – business services that initiate action based on invocation
    - Encapsulation that hides internals
        - Hide underlying changes
        - Can be responsive to necessary changes vs. getting everyone on board with (shared) database changes
    - Abstraction gives you coarse functions instead of granular function calls
    - Interoperability
        - Rely on HTTP and XML/JSON
    - Reusability
- Very familiar pattern for developers (request/response)
- Good for fine grained functions and mashup services
    - Want data before moving to next step
- Often don't need guaranteed delivery or a broker since you can just retry the request

Challenges

- ** Coupling **
- Not as bad as using a shared database
- Still have relatively tight coupling
    - Hard to change one piece
- ** Security / Capability gaps **
- Have to support a wider variety of capabilities by service providers
- Lack of uniform security strategies, inconsistent support for transactions
- Depends on protocols that can be consumed
- ** Locking **
- Blocking call for the sender
    - Receiver processes exceptions (pro and con)
- ** May be transient data **
- If a straight lookup without persistence … not available for reports, workflows etc
- ** COTS support **
- Many COTS systems don't natively expose their capabilities as services
- Some have gotten better, but you still see hyper-granular or abstract services exposed by leading vendors

Cloud Considerations

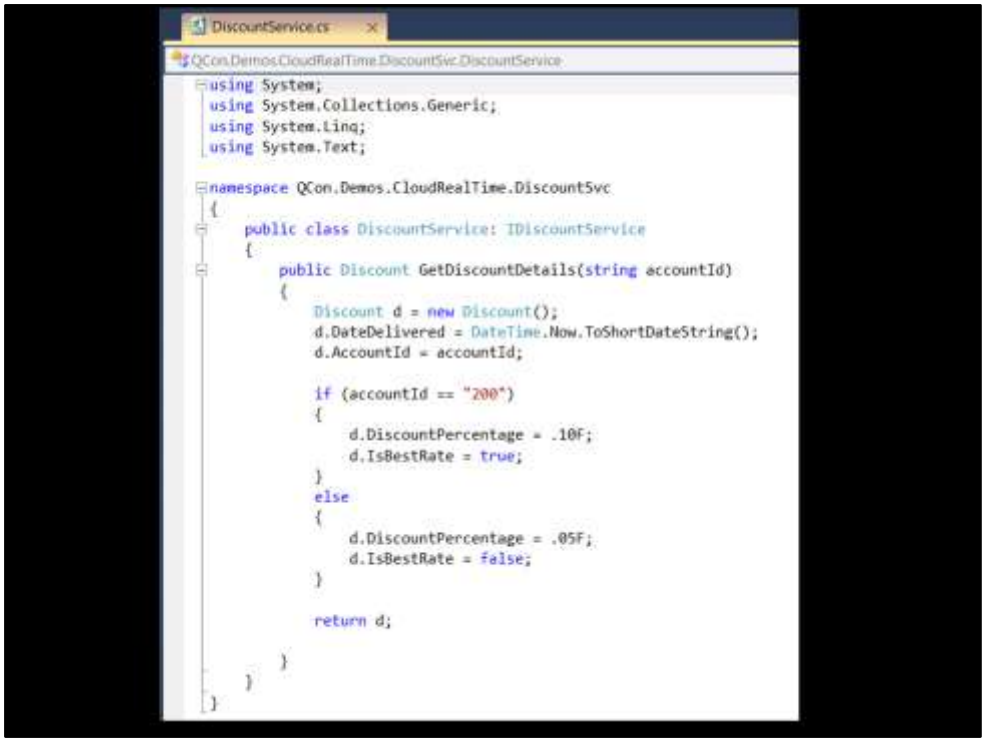
- ** Security **
  - How do you consume onsite services?
  - How federate when doing cloud-to-cloud
- ** Latency **
  - May require double hop if we first get a token and then make actual request
  - Given that this is a blocking call, may need to use AJAX design
- ** Access **
  - May require specific protocols (HTTP)
  - Not going to see broad support for DB-specific protocols
- ** Option **
  - Cloud to on-premises
    - Could use internet facing proxy service that forwards request to back-end system/DB
    - Could use VPN between caller and target system
    - Relies on cloud app that supports VPN
      - Amazon VPC, Google Data Connector, Azure Connect
    - Could use Windows Azure AppFabric for cloud based relay service (not durable)
  - Cloud to cloud
    - Leverage web services on both sides; Uwith cloud integration provider



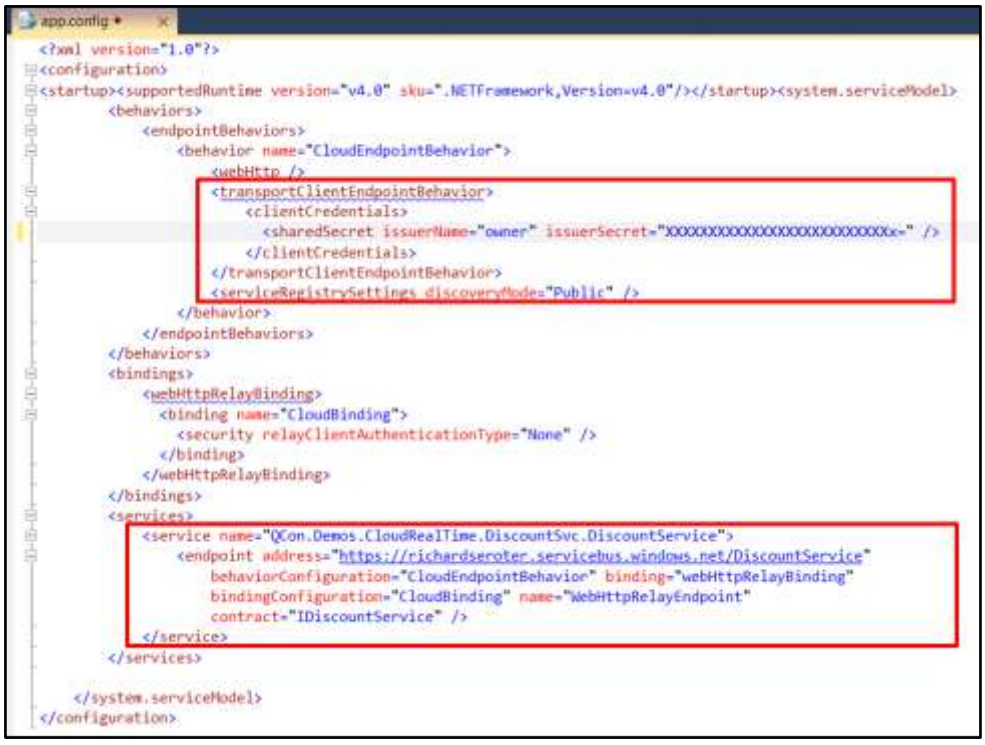
Demo

- Show on premise service
  - REST contract
  - Implementation
- Show what makes it “cloudy”
  - Cloud bindings
- Start service
- Show in registry
  - http://richardseroter.servicebus.windows.net
- https://c.na11.visual.force.com/apex/DiscountLookup?id=001A000000YBX3C
- Show SFDC call (with token for security)
- Call service
- Call service from account with different ID

I've built a custom WCF service that uses a RESTful (vs. SOAP) web service strategy

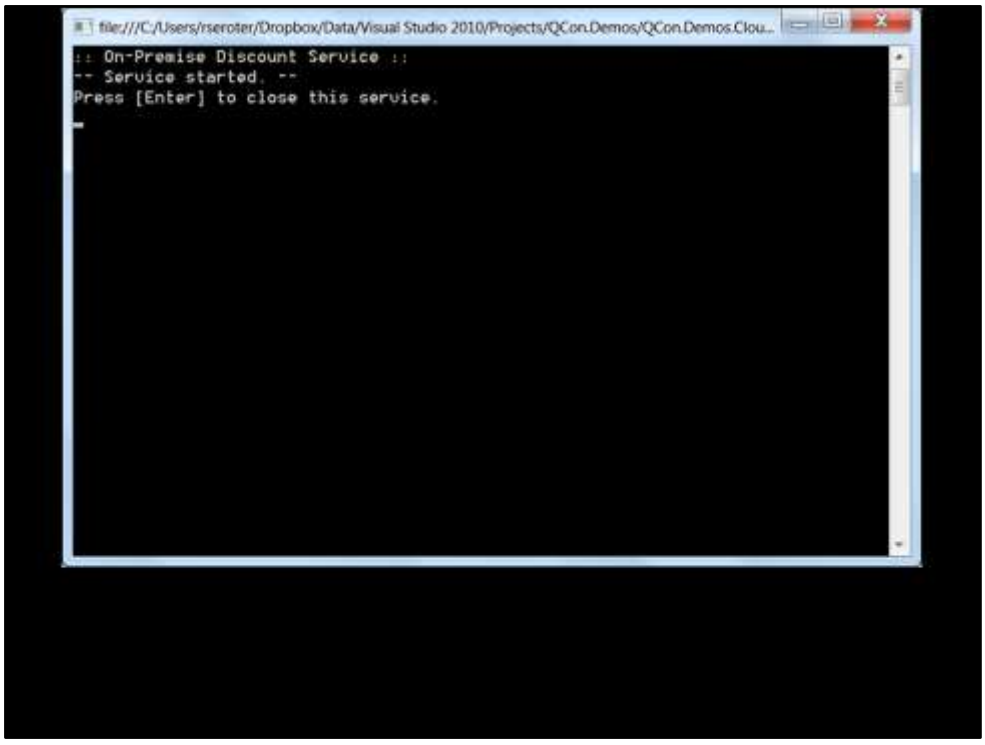


The implementation of this service simply says if the user ID is 200, the discount is 10%. Otherwise, 5%.

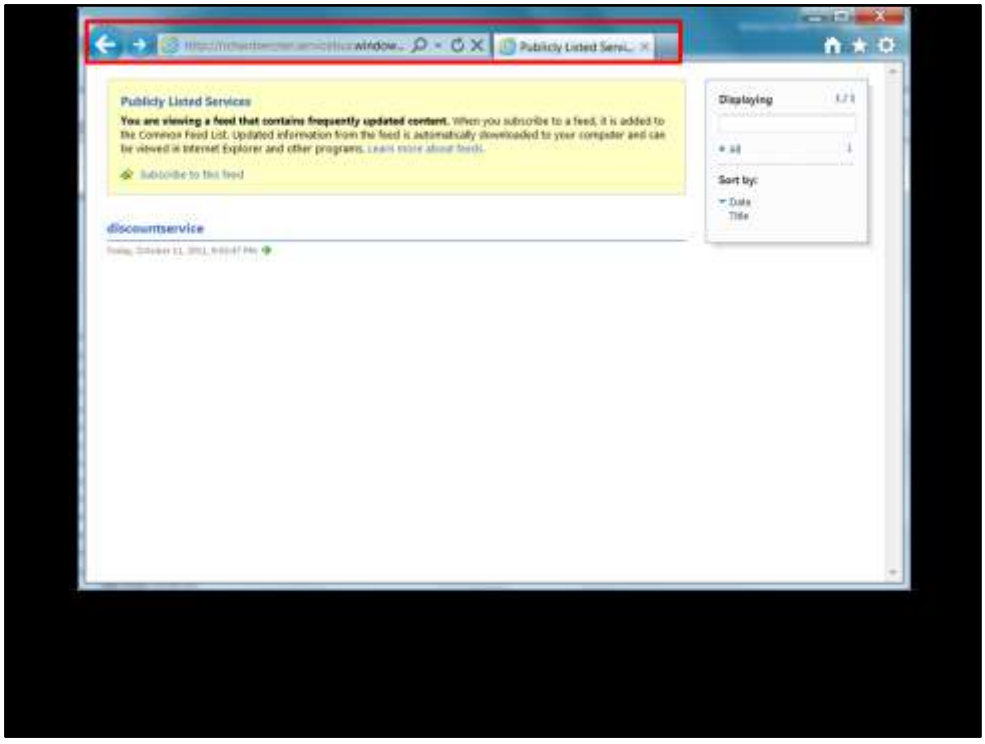


This service connects to Windows Azure AppFabric via this configuration. What this does is create a secure, two-way tunnel to the Microsoft cloud.

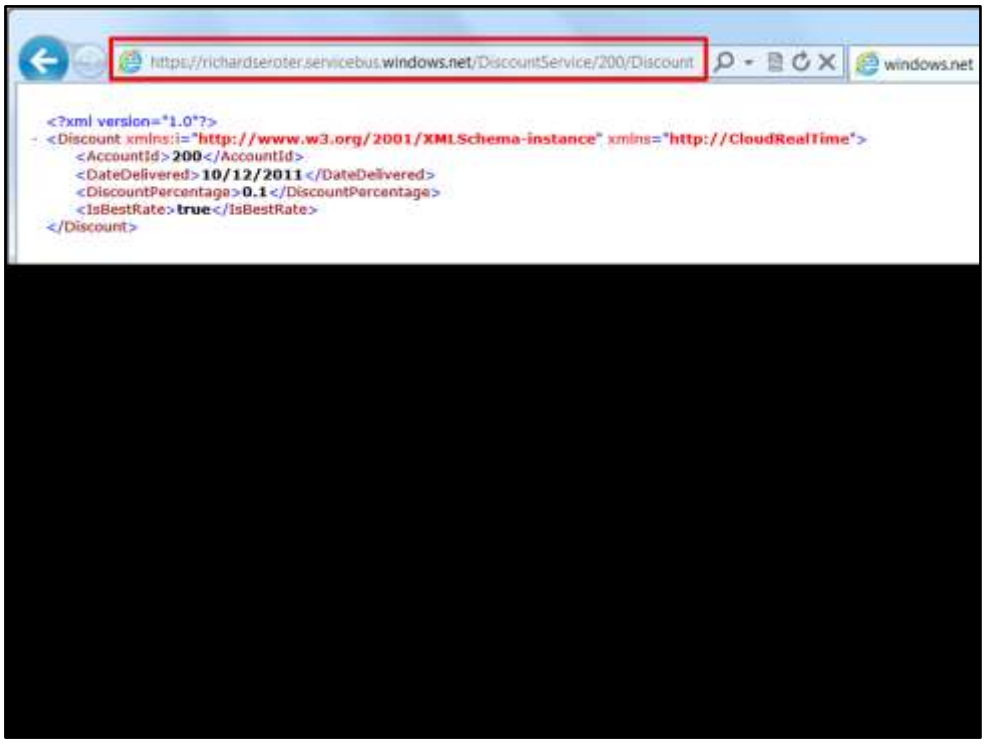
Messages sent to http://richardseroter.servicebus.windows.net/DiscountService will get “relayed” to my on-premises web service!



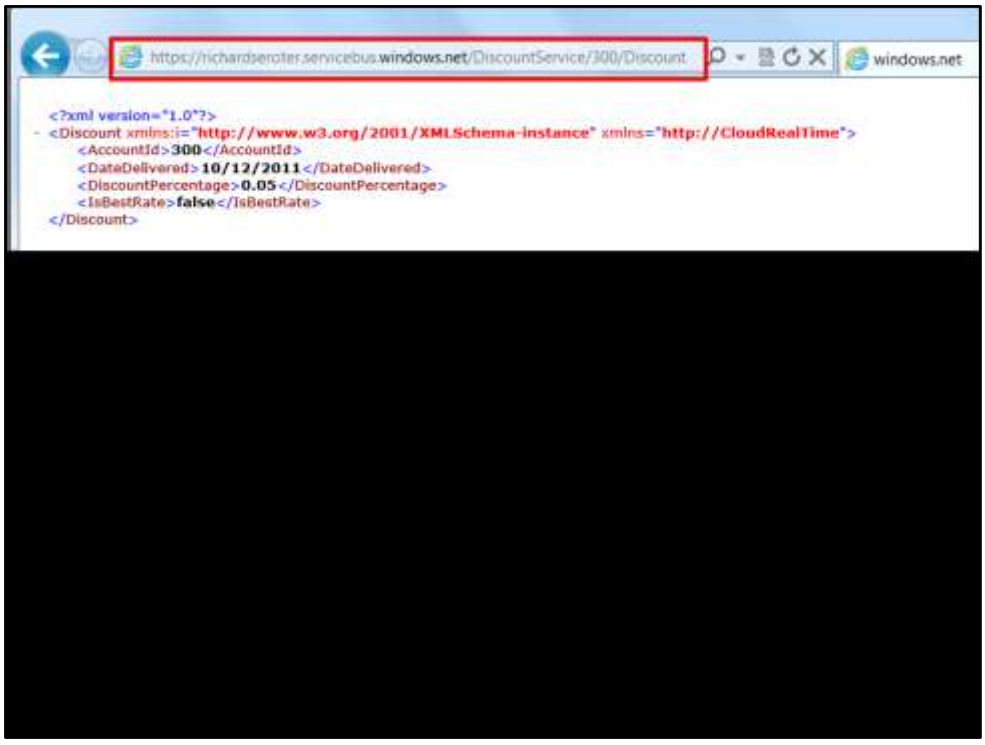
I start up my on-premises service, which initiates the binding to the cloud



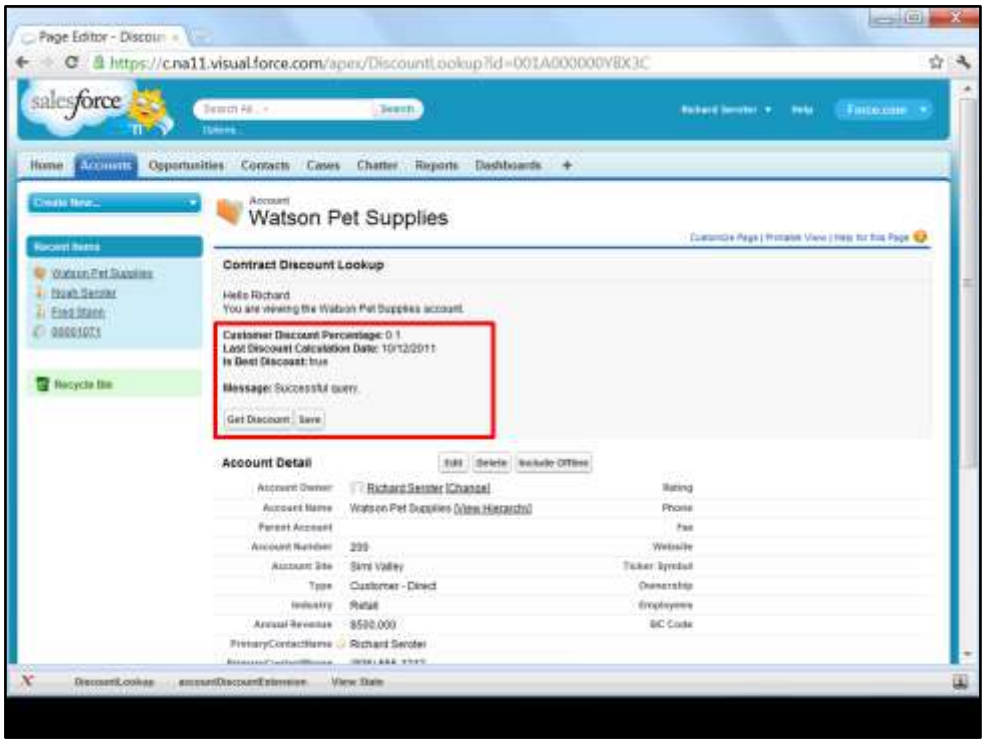
I can see my on-premises service exposed in my cloud registry



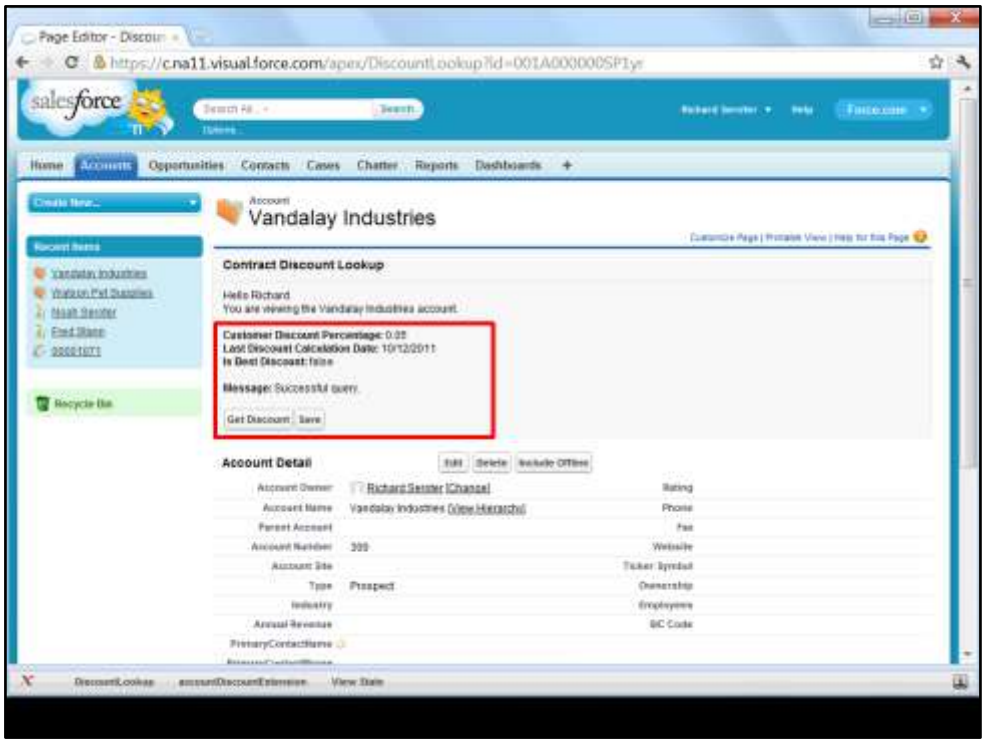
Because I have turned off caller security, I can test my service using any browser.



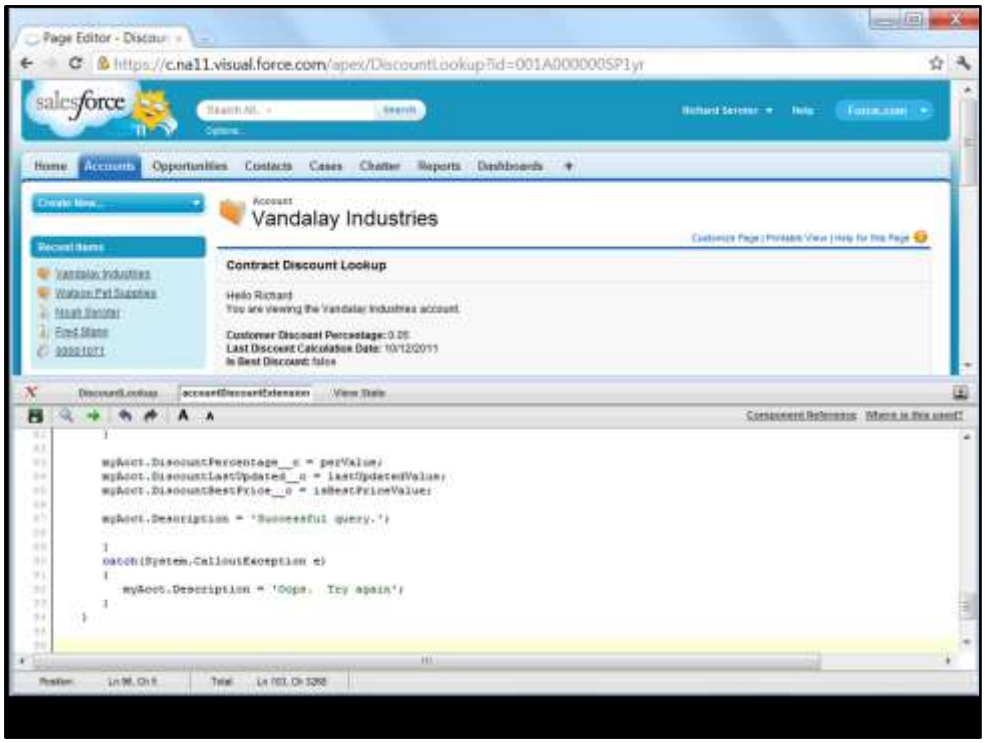
Note that I get a different value when I use a customer ID besides 200



Within Salesforce.com, I have a custom page which consumes the cloud relay service and executes my on-premise business logic in real time!



For a different account ID (besides 200), a different discount rate is returned



Behind the scenes, this code connects to the relay service and optionally passes in a security token.

When it makes sense …

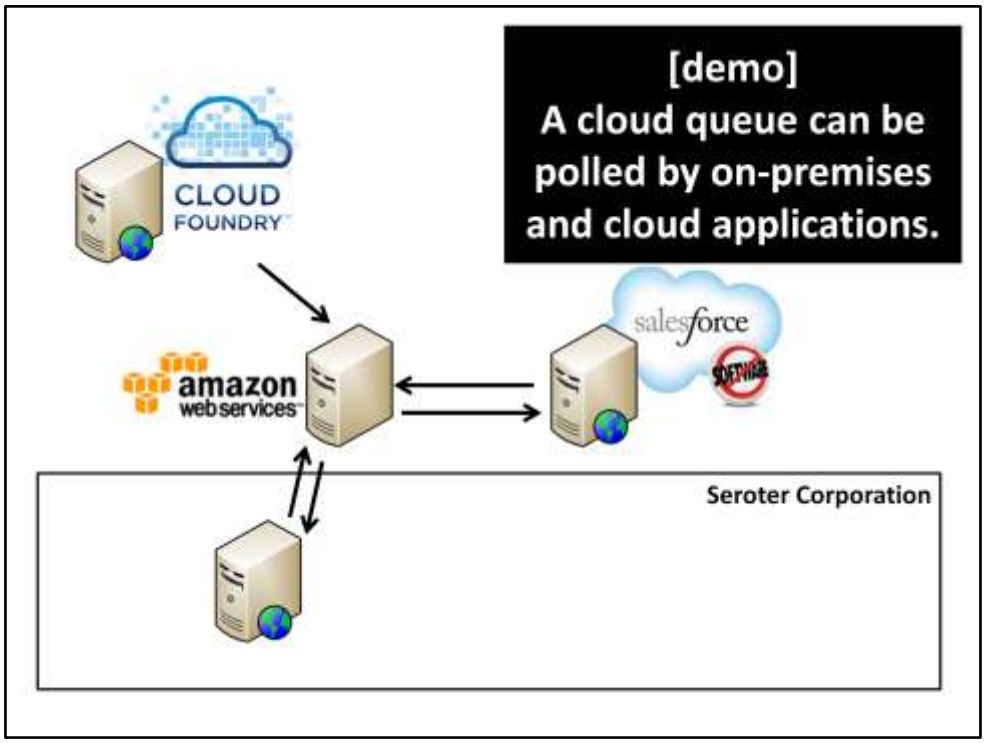
- Share data between systems in a responsive and scalable way
    - Caller doesn't have to wait
    - Think of sending an email; it doesn't block Outlook until your recipient gets it
- Want to do broadcast (pub/sub) or multicast (defined recipients)
- Caller doesn't need to care where it goes
- Supports disconnected applications (not online at the same time)
- Have multiple replicable units (bus scenario)

Challenges

- ** Not real time synchronization **
- Inherently includes eventual consistency
    - No simultaneous or instant updates
- ** May need durability / reliability **
- Without durability in the middle, stand the chance to lose data
- A router can be used if you want to direct the messages to places unknown to the caller
    - Router can also do activities like data transformation, protocol bridging, or workflow
- Idempotence needed in many cases
    - Data shared my tell receiving system to go get data (handle dupes ok)
- ** App support for receiving or sending async messages **
- Few applications natively share data asynchronously

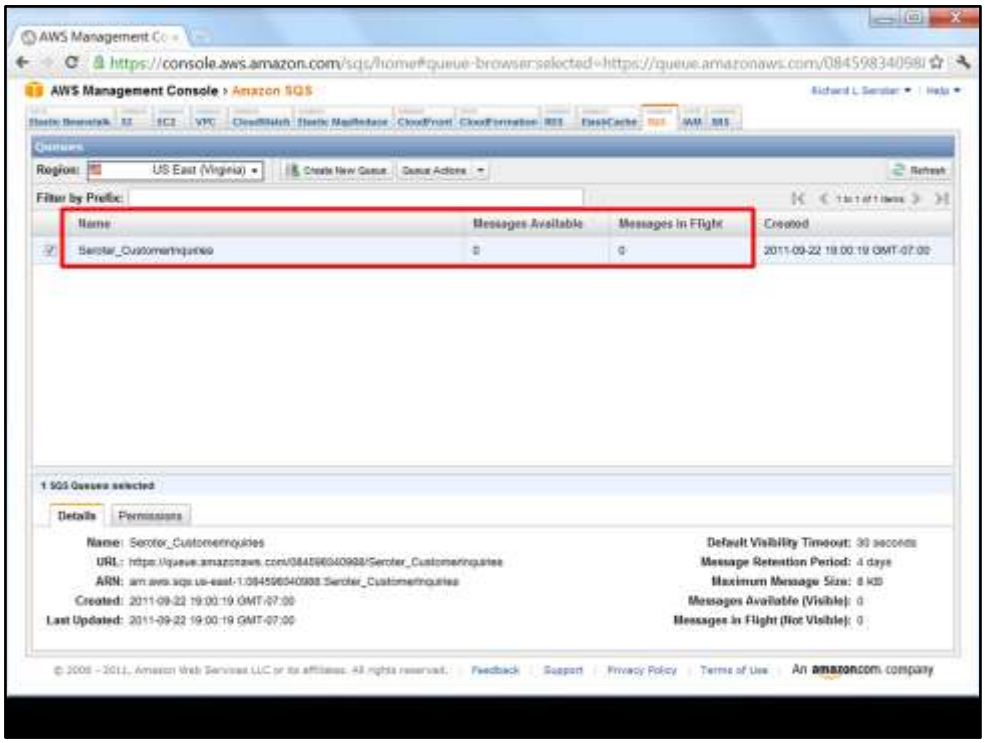
Cloud Considerations

- ** App support **
- Seems even fewer cloud apps share or receive async
- ** Security **
- May still need to poke a hole in the firewall
- ** Provider limits **
- Cloud vendor limits on polling (e.g. SFDC)
    - May want to put lots into queue and “peek” to retrieve/delete only certain ones
- Distributed nature of cloud arch could result in unexpected (lack) of data
    - Distributed queues may not return all items on each poll (machine sample via Amazon)
- **Options **
- Some SaaS platforms bake in async push (e.g. salesforce.com)
- Could leverage a cloud queue that everyone talks to
    - Or other cloud broker (DB, message router)
    - Could work for both cloud-to-cloud or between cloud and on-premises (both ways)

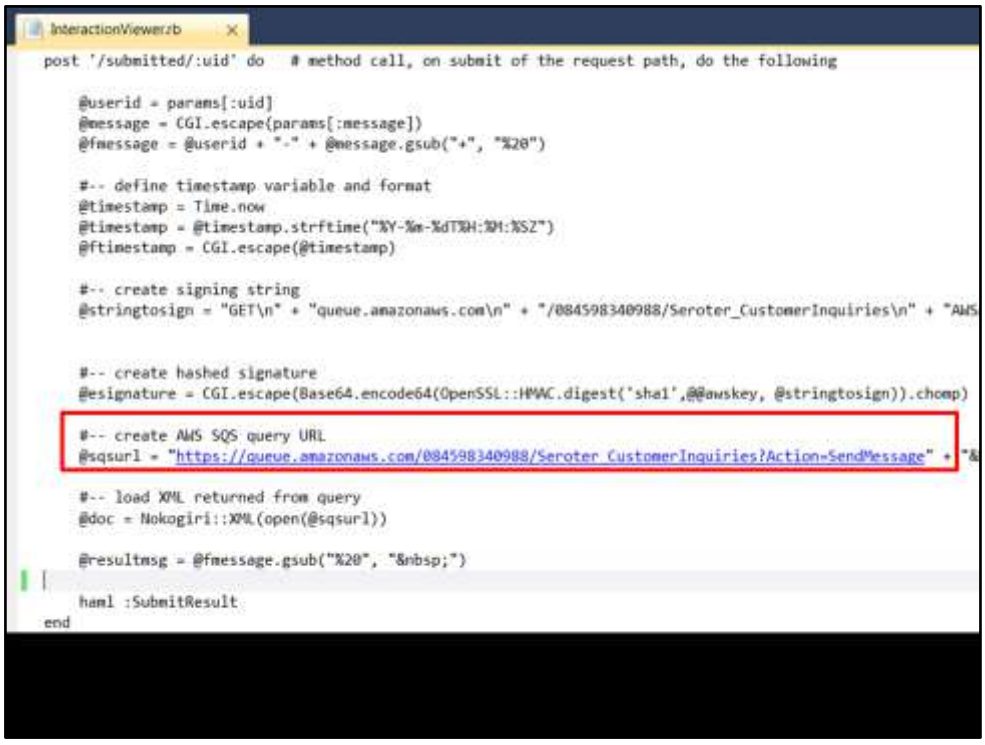


Demo

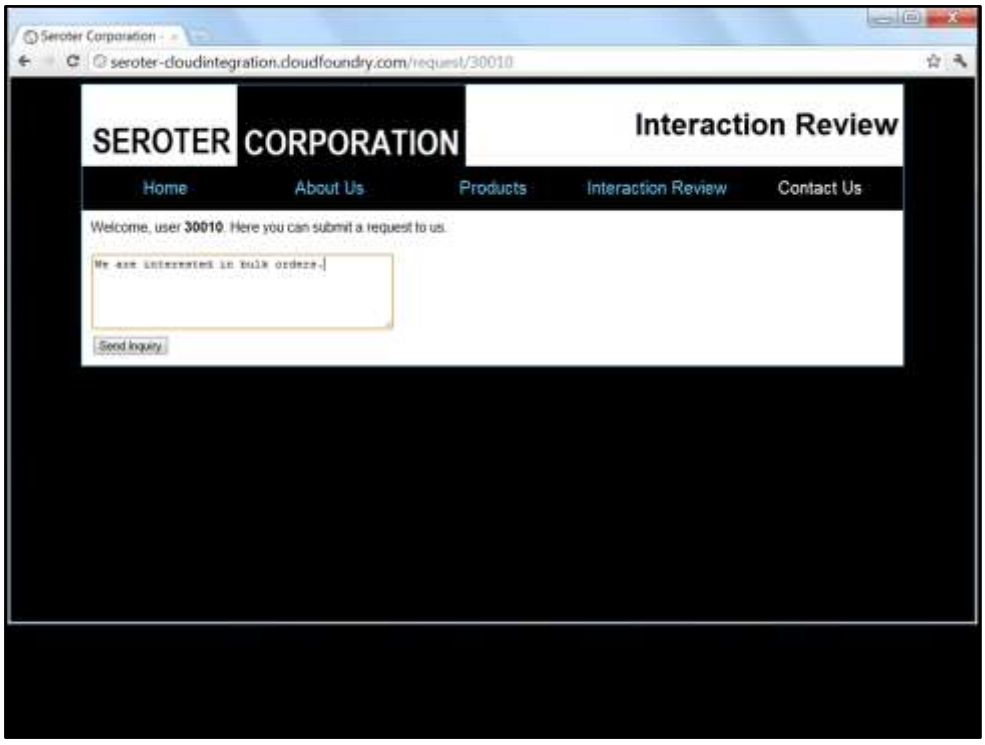
- Show queue on AWS console
- Show CF call that puts stuff into queue
- Show queue in AWS console
- Call Pull from on-premises client
    - Do retrieve plus delete
- Now empty if you look at AWS console
- Go to SFDC
    - Have code that can call “poller” every hour and update SFDC by creating a “case”
    - Show poller interface
    - https://c.na11.visual.force.com/apex/CaseView
    - Can also trigger manually and check queue and create “case”
    - See AWS queue is now empty



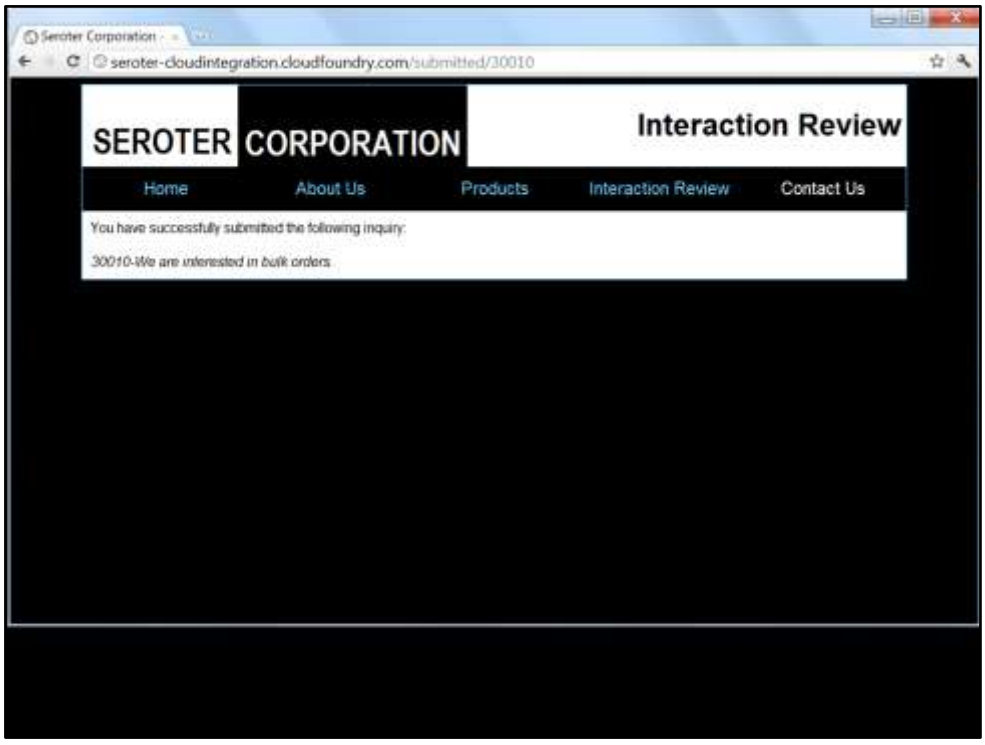
I have an AWS Simple Queue Service (SQS) queue created



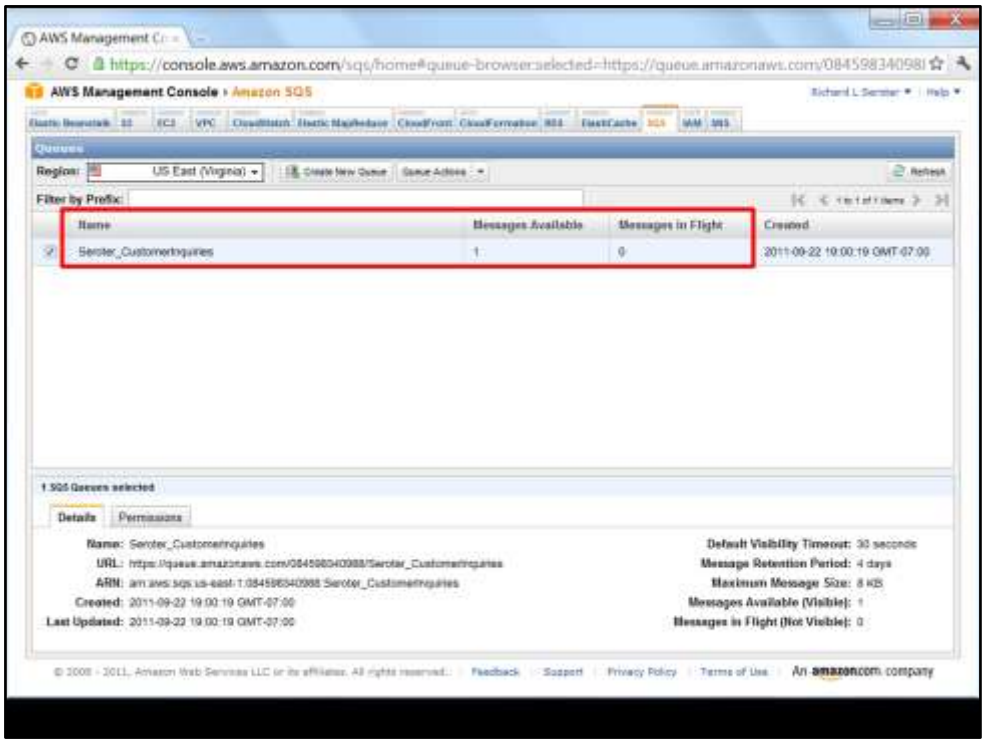
Within my Cloud Foundry app, I build up my request to SQS and add a message to the queue



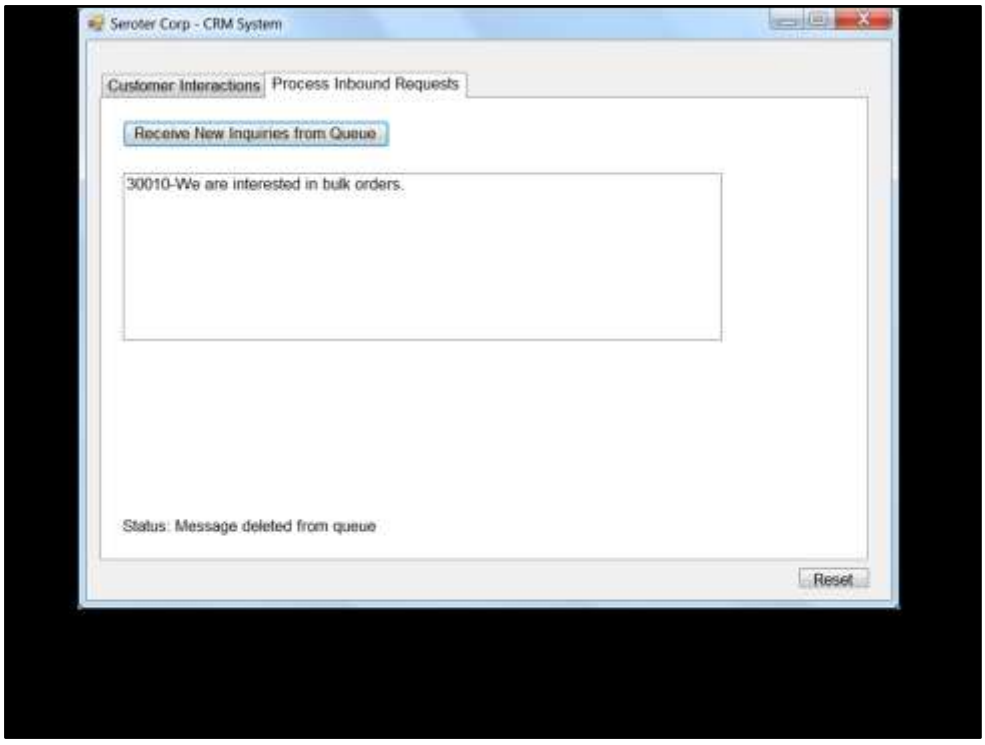
In this Cloud Foundry app, I send “feedback” messages from websites to a queue where either an on-premises application, or other cloud app, can read them.



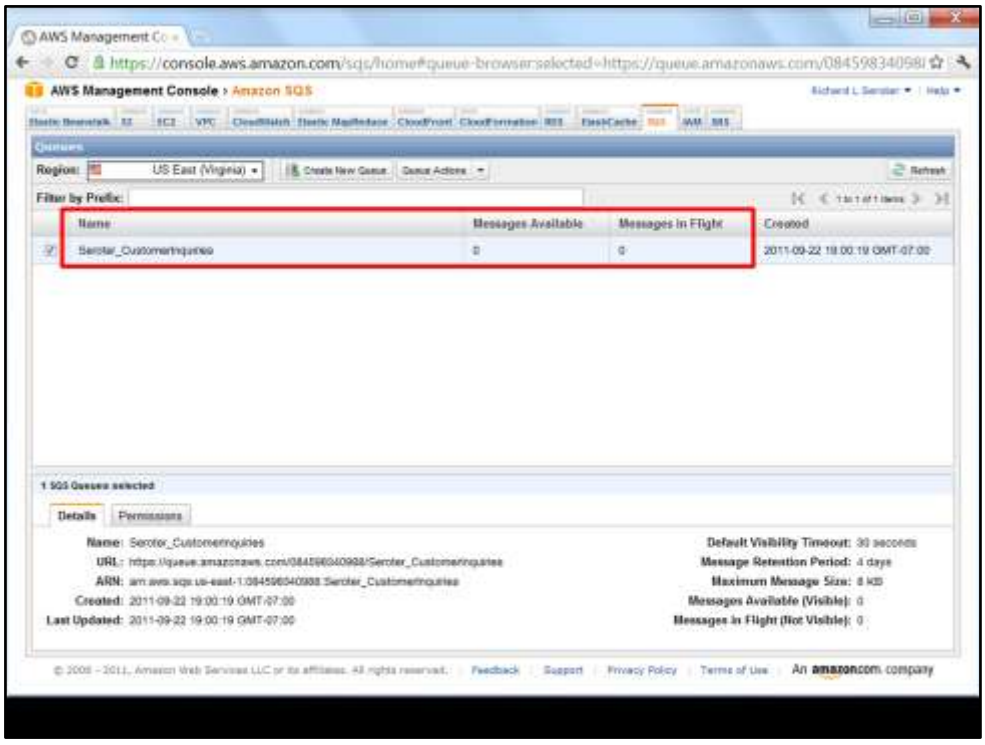
A message has been added to the queue.



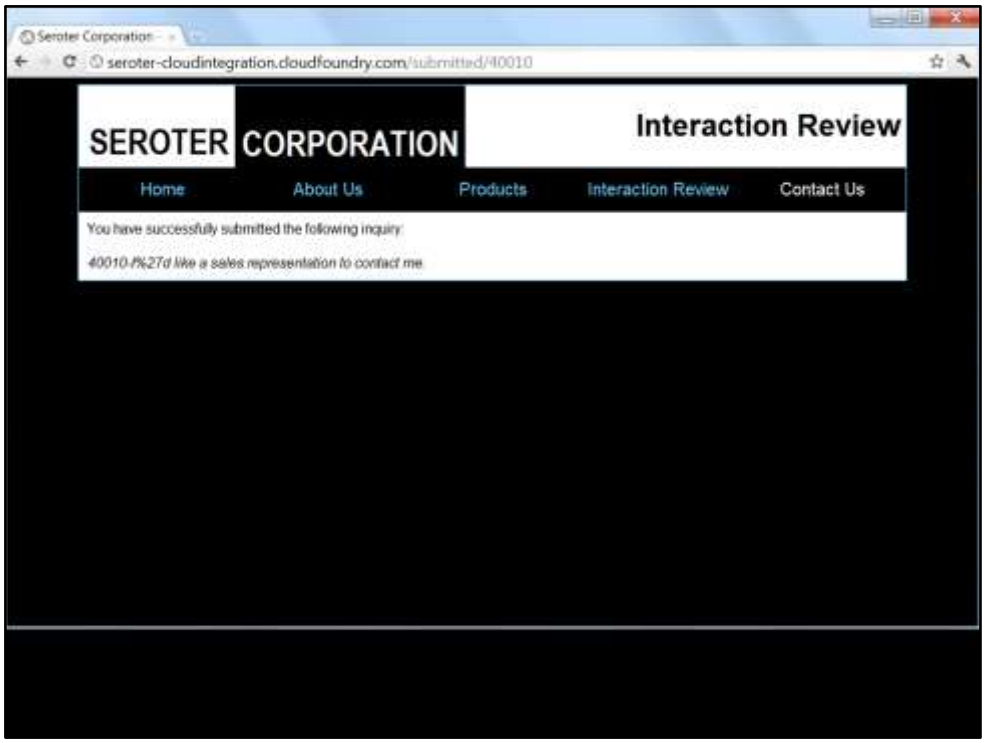
I can see that my queue has a message in it.



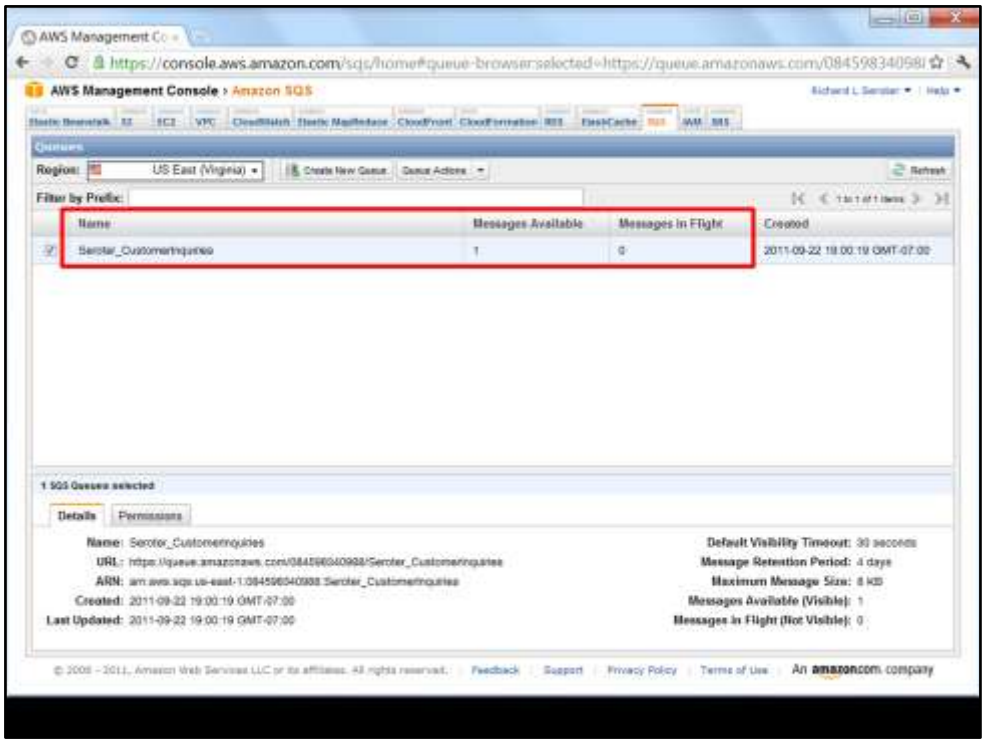
Within my on-premises app, I can retrieve and then delete that message (so that other queue readers don't get it).



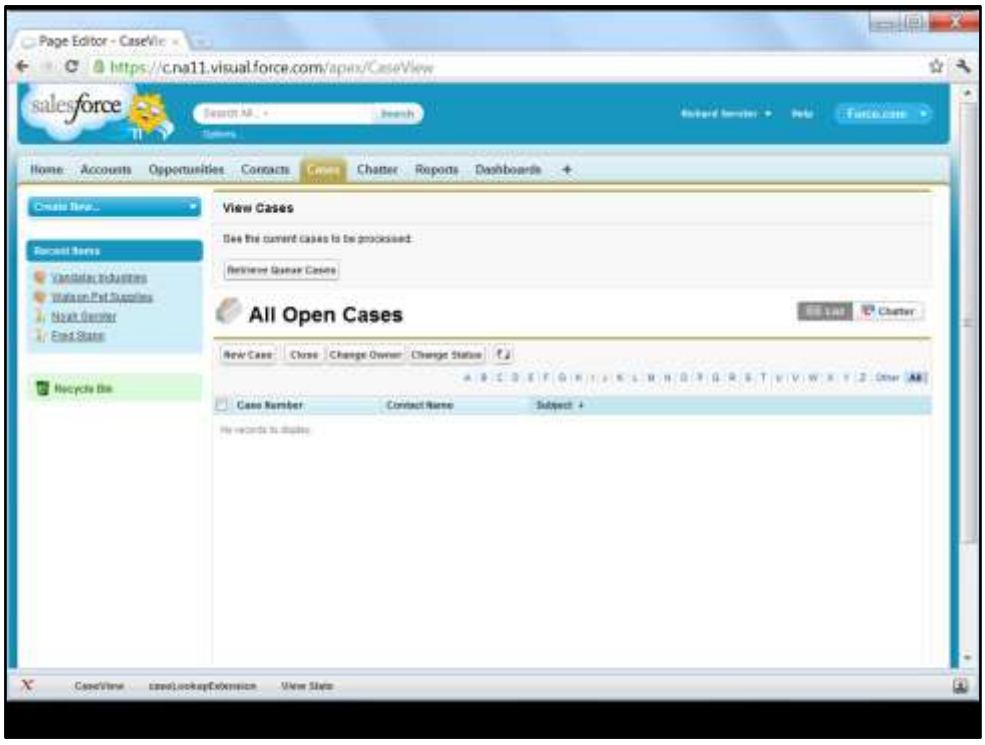
My queue is now empty.



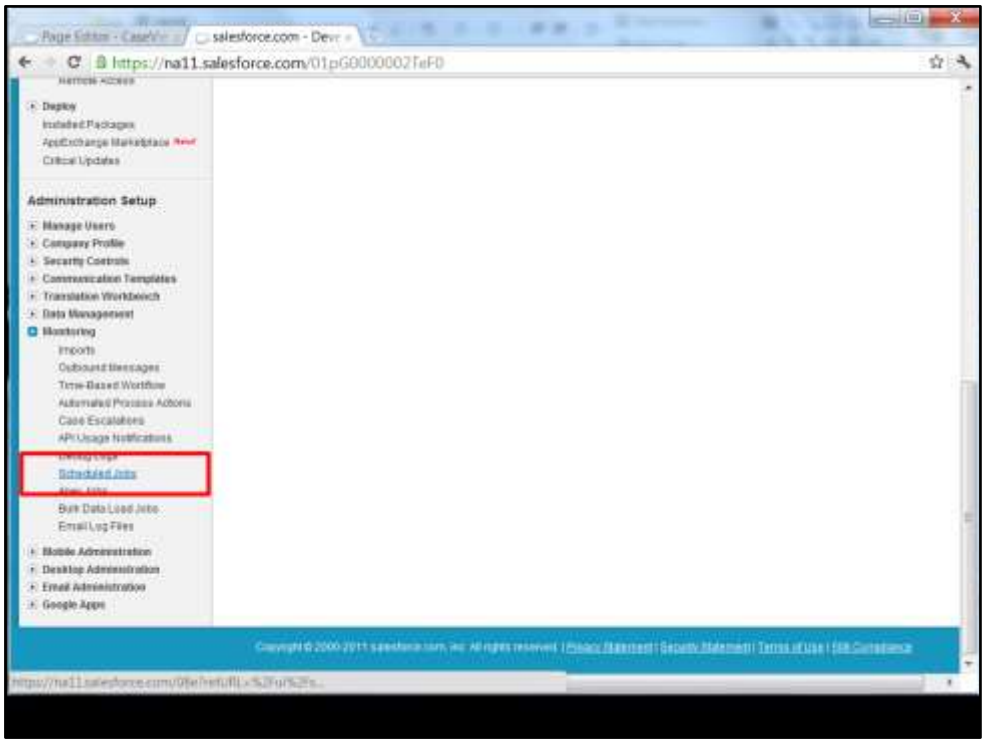
I sent another message for a different customer.



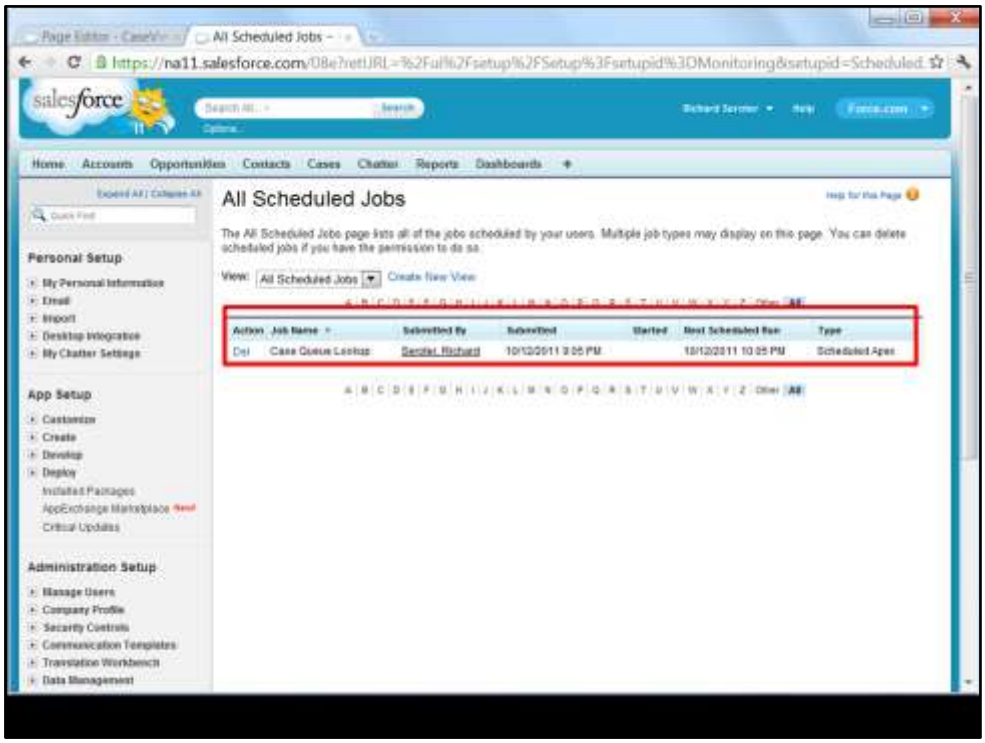
My queue now has one message



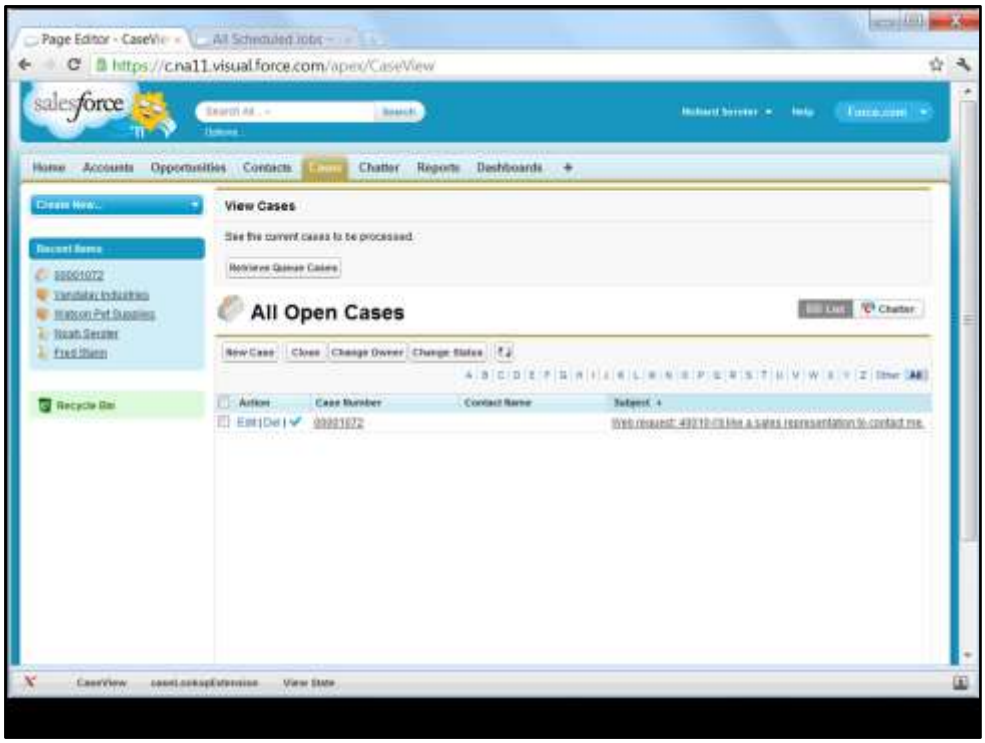
Within Salesforce.com, I have a custom page which queries the queue, and if it finds anything, it creates a new “case” in the system.



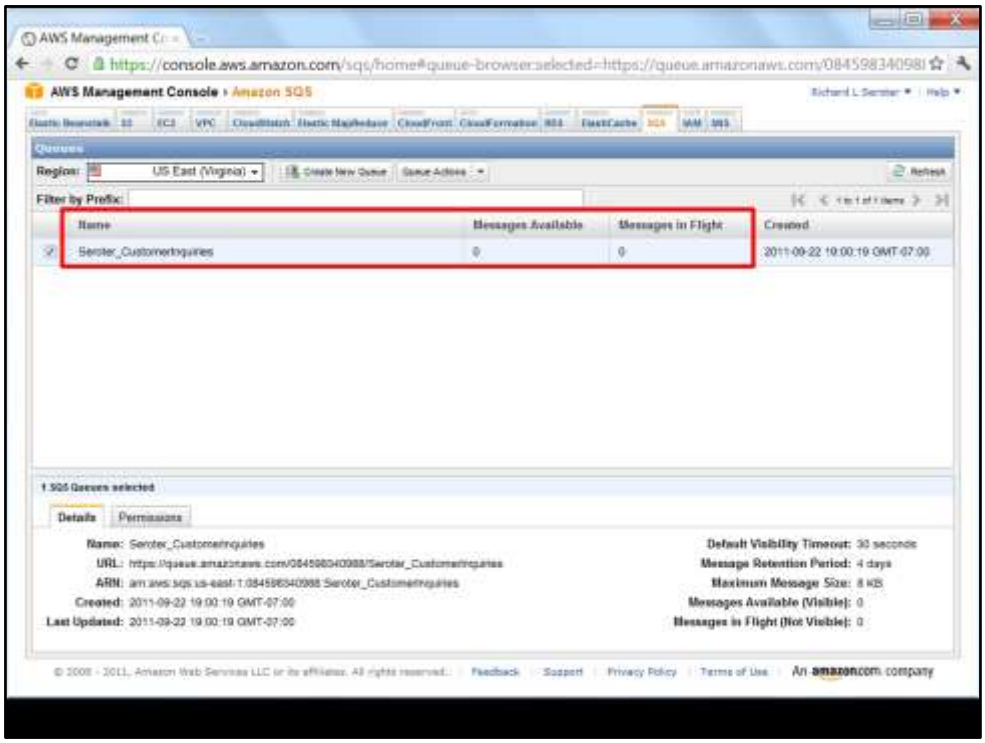
I can also schedule this “queue query” to run on scheduled intervals for a more real-time data exchange. Instead of doing bulk integration every day, we could send messages to the queue and have them loaded into Salesforce.com every hour or so.



Here is my job scheduled to run every hour.



Since I'm impatient, I can also click the button the page to look for queue messages. Here, the message we sent from Cloud Foundry shows up as a case in Salesforce.com.



The queue is now empty.

Apply proven patterns to cloud integration and look for more than just bulk transfer scenarios.

Questions?

@rseroter | seroter.wordpress.com


QCon
北京站·2012年4月18~20日
www.qconbeijing.com（11月启动）
QCon杭州站官网和资料
www.qconhangzhou.com
全球企业开发大会
INTERNATIONAL
SOFTWARE DEVELOPMENT
CONFERENCE