Introduction

- In Learning New Concepts multiple perspectives are important …
- Development **AND** Info. Management  
  - **Software Engineer** for 20+ years  
  - **Author** of 11 Technical Books  
  - **Inventor** of Patent #7299408  
  - Assisted **NMFS** on several EDM projects!
- Government **AND** Contractor  
  - Military Intelligence Officer  
  - Former **DHS Metadata Program Manager**  
  - Employee and Executive at 4 Contractors  
- Business **AND** IT  
  - Small Business Owner and IT Employee  
- GCN Columnist of “Reality Check” Column  
- Currently lead a data standardization effort for ODNI and biometrics development.
Agenda

✓ Cloud 101
✓ IM & Semantics Foundations
✓ IM & Semantics in the Cloud

Semantics & Info Mgt

InCadence Strategic Solutions – Proprietary and Confidential
Cloud 101
Cloud 101: What is the Cloud?

- Analogy for the Internet...

- Elasticity via “Computing as a Utility” (McCarthy)
  - Service Provider just like the electric company
    - Characteristics
      - On-Demand
      - Rapid Elasticity
      - Measured Service
Cloud 101: Examples of Cloud Computing

- Google Apps
- Microsoft
- Salesforce
- Amazon
- Apps.gov

My Current Use Cases:
- Software Development and Testing
  - Cloud is especially useful for provisioning test environments.
- InCadence Private Cloud – Hadoop, VMs
History of Cloud Computing

Virtualization
- Early Concepts (John McCarthy)
- IBM CP/CMS
- IBM VM/370

Mainframes
- Minicomputer
- Personal Computer
- WWW

Web

1960 | 1968 | 1972 | 1980 | 1991...

Web Hosting
- “Cloud” coined by Ramnath Chellappa
- Salesforce.com
- VM Ware

SOA
- Amazon Web Services (SOA)
- Google File System Paper

ASPs
- Google Maps
- Hadoop
- Amazon S3
- Amazon EC2

Cloud Computing...
- Google App Engine
- Windows Azure
- Apps.gov

Benefits of Cloud Computing

- For **Businesses**:
  - Low startup costs!
  - Faster Development (web deployment)

- For **Government**:
  - Data Center Consolidation – reduce wasted hardware with scalable utilization (if you can migrate apps)!!!
  - Potentially faster development (web deployment)

**Business Benefits Do Not Always Translate to Government!**
Definitions of Cloud Computing

- **NIST Definition:** Model for enabling convenient on-demand network access to a shared pool of configurable computing resources
  - Service Models
    - IaaS, PaaS, SaaS
  - Deployment Models
    - Public, Community, Private, Hybrid

- **Other Definitions**
  - Wikipedia: Cloud computing refers to the provision of computational resources on demand via a computer network.
  - Vendors. *** See my latest GCN article!
**Infrastructure As A Service (IAAS)**

- **Definition:**
  - Infrastructure as a Service (IaaS) offers compute power, storage, and networking infrastructure (such as firewalls and load balancers) as a service.

- IaaS vendors use *virtualization* technologies to provide compute power.

- **KEY USE CASE:** scalability and outsourcing of data center and web hosting functions.
Virtualization: Example Virtual Box
Platform As A Service (PaaS)

- **Definition:**
  - PaaS is a development platform, in the cloud, for building and deploying cloud-based applications.

- **KEY USE CASE:**
  - Rapid and Scalable Application Development (i.e. Build Cloud Apps!)

- **Vendors:** Google, Amazon, Microsoft, Salesforce, others.

---

InCadence Strategic Solutions
Software As A Service (SaaS)

- **Definition:**
  - Providing finished applications on-demand. The application exists in the cloud and can be consumed from any browser.

- **KEY USE CASE:**
  - Commercial (packaged) Software with no installation and metered billing (i.e. email)

- **Vendors:** Google, Salesforce, Microsoft, and many others!!!

**Buyer Beware** – per user pricing may not be a bargain!
Cloud Deployment Options

- **Public** – Services provided over the Internet and owned by an external organization.
- **Community** – Services shared by multiple organizations.
- **Private** – Internal IT owns and operates the cloud infrastructure.
- **Hybrid** – Combination of the public and private based on application sensitivity.
Cloud computing Architecture Elements

- Web Front-End (Rich Internet Applications)
- Application Programming Interfaces (API)
  - Queuing
  - Caching
  - Scalable data stores/querying
  - Parallel Processing Algorithms
    - i.e. Map/Reduce
- Virtual Machines/HyperVisor
- Disk Storage/Data Storage

Key Takeaway: The Key Cloud concept is the development of a Multi-Machine OS
Systemic Cloud Issues

- **Security/Privacy**
  - Centralized “Fort Knox” model is a target; Insider Threat, Foreign Ownership
  - Poll: 68% voted this #1 concern.

- **Reliability**: outages, lock-in

- **Interoperability**
  - Data Portability
  - Application Portability

- **Where’s the Savings?**
  - **Labor** is the real cost sink… yet another REWRITE?
  - If you started a SOA, you are ahead of the curve!

---

Where’s the Beef?!
IM & Semantics
Data is Not Information

- **So, what’s the difference?**
  - **Data** - A collection of unprocessed (or raw) facts
  - **Information** — is “derived from the word *inform*, which means ‘to give shape to’”. Data that is “shaped” (or processed) to inform a user.
    - Use a Physical Product Analogy!
  - Does your IT architecture have an “Information Layer”??
    - i.e. metadata catalog
    - How do you do metadata in the Cloud?

- **So, the goal is data that is useable by consumers …**
Outcome-Based Information Management

Resources -> Activities -> Outcomes

Spiral Implementation

Comm./Strategy -> Analytics -> Integration/MDM

Governance

Modeling/Metadata -> Tools

Decision Spt/Perf.
Single Version of the Truth
Data Quality
Awareness/Availability

Physical Data Sources

For more details, grab a copy of the OBIM Whitepaper!
Smart Data Continuum

The trend is to put the “smarts” in the data, not in the applications.
**From Information To Knowledge**

- **Data Optimization Pyramid**
  - **Products**
    - Oracle, Calais, iLog (IBM), Top Quadrant, Powerset (Microsoft)
    - Easy Knowledge Base (EZKB)
    - Many Others...
  - **TRENDS:**
    - Chicken and the Egg
    - Knowledge Bases for Everyone
The Semantic Solution

- Ontology (and associated processing tools) offer the capabilities of a:
  - Database,
  - Metadata repository,
  - Reference Data Store,
  - Link Analysis Tool,
  - Master Data Management (MDM) Hub,
  - Business Rules Engine, and
  - an Inference Engine

- All rolled into one!
Logic, Inference and Axioms

- **Logic**: a system of reasoning. Began with Aristotle:
  - Example of an Aristotelian Syllogism:
    - Major Premise: All Men are Mortal.
    - Minor Premise: Socrates is a man.
    - Conclusion: Socrates is Mortal.
  - Many types of logic: Predicate Logic, Modal Logic, Propositional Logic

- **Inference**: deriving new knowledge from existing knowledge via various techniques (i.e. deduction, induction, etc.)
  - Modus Ponens (If \( p \) then \( q \); \( p \); therefore \( q \))
  - Modus Tollens (If \( p \) then \( q \); not \( q \); therefore not \( p \))
  - Transitive (Chain Rule).
    - If \( A \rightarrow B \) and \( B \rightarrow C \) then \( A \rightarrow C \)

- **Axioms**:
  - Example: A Triangle has 3 sides.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>q</td>
<td>( p \rightarrow q )</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
</tbody>
</table>
Semantic Success Stories

- A Semantic Wikipedia
  - Powerset bought by Microsoft (Now part of Bing)
- A Question & Answering Champ
  - Watson by IBM
  - Now targeting Healthcare Expert System Market
- Open Source Knowledge Base
  - Metaweb (creators of FreeBase) bought by Google
- Question and Answering On the Web
  - Wolfram Alpha
- Data.gov/semantic
- Apple’s iPhone 4S – “Siri”
The biometrics system MARS uses a declarative objects system called “Coalesce”.

Biometrics and Identity Management are rich grounds for semantics.

- Person, Relationships, Organizations, Actions, etc.
- Facebook’s Social Graph is semantics!

Coalesce is declarative, XML

Coalesce is NOT OWL, yet…
The Tactial Rapid Exploitation (T-REX) portal also leverages Coalesce entity-graphs.

Key foundational underpinnings of semantics are becoming commonplace:


I am working on Coalesce 2.0: security, semantics and the cloud.
What is Cloud Data like?

- Data Storage is different
  - No-SQL movement, key-value pairs.
  - Map/Reduce processing (Parallel algorithm)

- Examples of Big Data:
  - Geo-location data from cell-phones; all wikipedia text; census data; Shipment data; Computer logs; **Sensor Data**, social media, surveillance…

- A new Data Type? Stream data
Ramifications of Big Data

- Clouds enable Big Data
  - Thus, big data may increase
- Streaming Data may not be record-oriented
- Streaming Data may not be document oriented
- Streaming Data may be transient (for specific analysis)
- Streaming Data may be aggregated from multiple sources.
Cloud models and IM

- **SaaS**
  - IM Tools moving to the cloud
  - SaaS Apps afford Little or no customization
  - Standardization could be a problem.

- **PaaS**
  - Non-Relational Data, esp. for Big Data
  - ***Biggest potential IM gain*** when migrating applications
  - Implement distributed metadata; keep locality.

- **IaaS**
  - Stovepipes remain, traditional IM challenges.
  - Some applications can be engineered at the “server level” (or may already be scalable)
Cloud Deployment and IM

- **Public**
  - Do you own the data?
  - Is your data portable?
  - Is your data secure?

- **Community**
  - Governance is key
  - Shared burden

- **Hybrid**
  - Same concerns as private.

- **Private**
  - Most secure option, no data ownership issues

"No, it's MY data!"
Where are Semantics in the Cloud?

- Semantic Markup is a natural fit for the cloud – why?
  - Semantic Web is Distributed and the Cloud is Distributed

- Ok, HOW?
  - Hadoop = streaming data
  - Graph Traversal -> streaming data

- In other words,
  "Stream the Graph"

- Mark Logic Hadoop Connector
  - See: Session by MarkLogic and Hortonworks on Thursday, 11am.

© Microsoft
Your Cloud Strategies

- Data Ownership is issue #1
- For PaaS, Data Asset registration, improve discovery and information sharing during migration.
- Cloud Data Strategy is different than your Cloud Application Strategy
  - App Strategy: Understand the benefits of elasticity and the potential for rapid innovation
  - Data Strategy: like transparency, big data does not change the basic disciplines (apply the DRM to Cloud!) Centralization is an opportunity.
Conclusion

- **Cloud 101**
  - You should have a good understanding of the basic concepts.
  - Remember: Cloud has pros and cons… so go slow.

- **OBIM & Semantics**
  - Information Management can focus on tangible Outcomes first!
  - Ontologies, if scoped properly, can be practical.

- **IM & Semantics in the Cloud**
  - Big Data is a challenge and opportunity!
  - Now is the time to be raising semantic requirements for the PAAS APIs.
  - Stream the Graph!

You NOW have the skills to craft a Semantic Cloud Strategy!