

# InCadence Strategic Solutions

## Information Management and Semantics in the Cloud

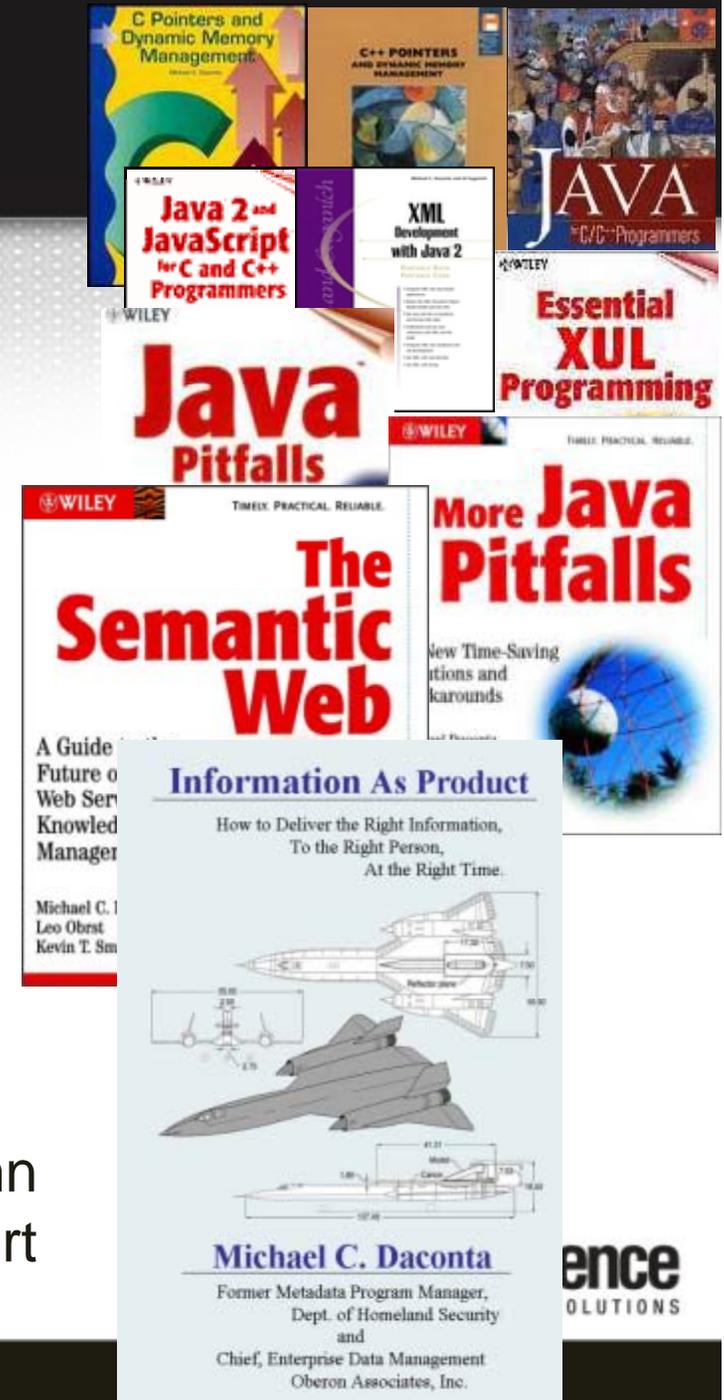


Michael C. Daconta  
VP of Advanced Technology  
[mdaconta@incadencecorp.com](mailto:mdaconta@incadencecorp.com)



# 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



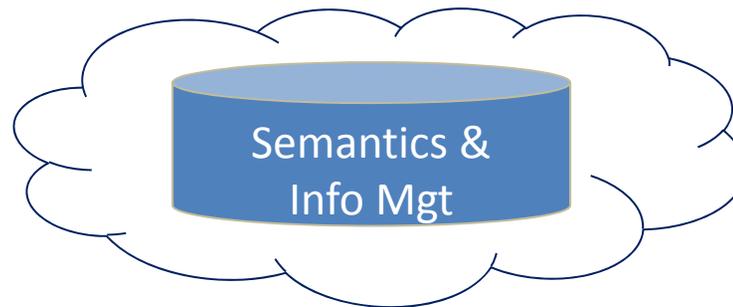
**Cloud 101**



**IM & Semantics Foundations**



**IM & Semantics in the  
Cloud**



# 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

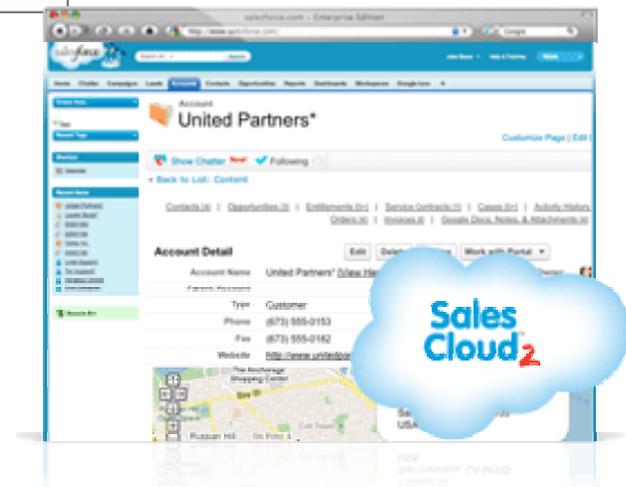


# Cloud 101: Examples of Cloud Computing

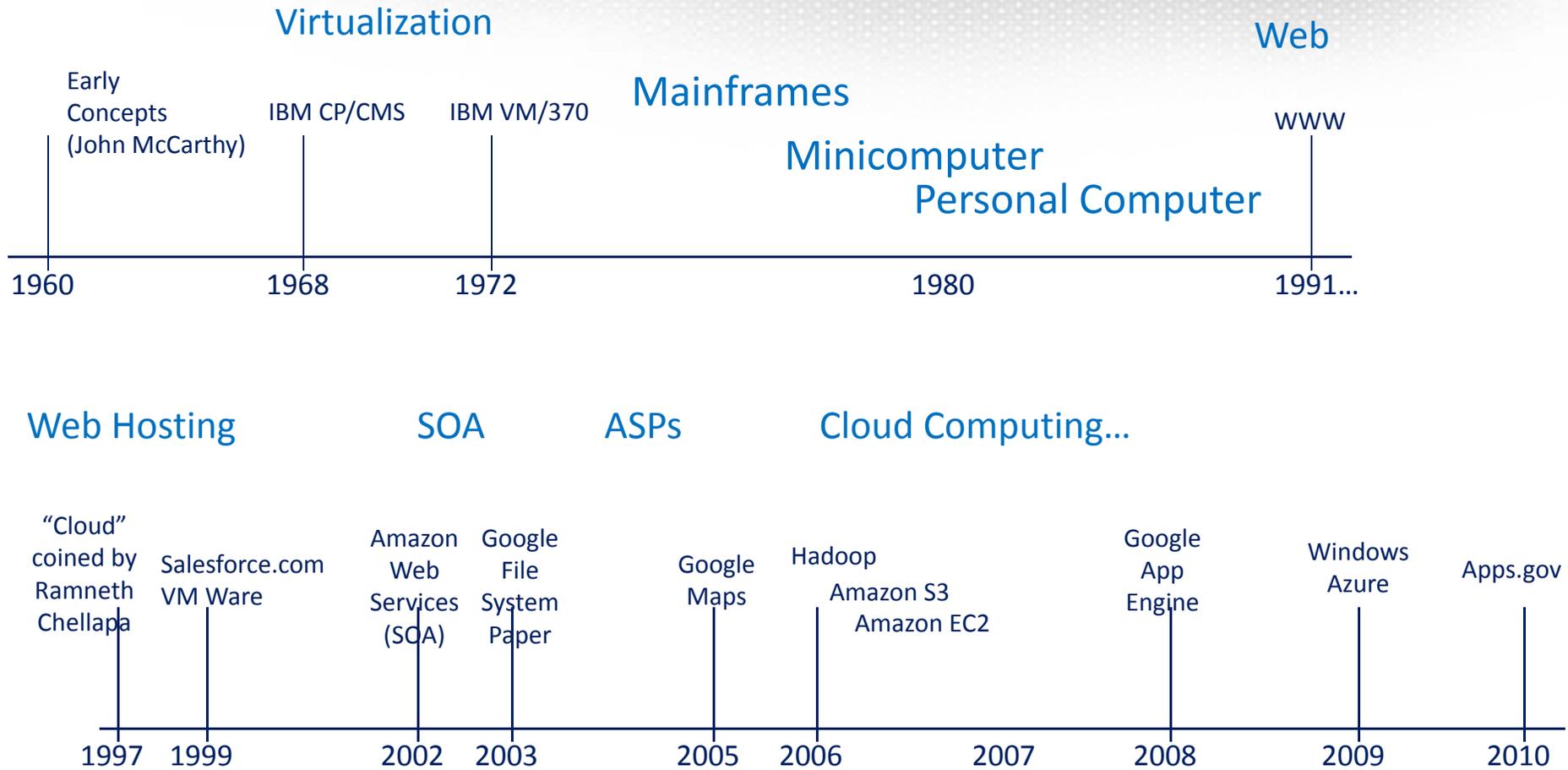
- Google Apps
- Microsoft
- Salesforce
- Amazon
- Apps.gov



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

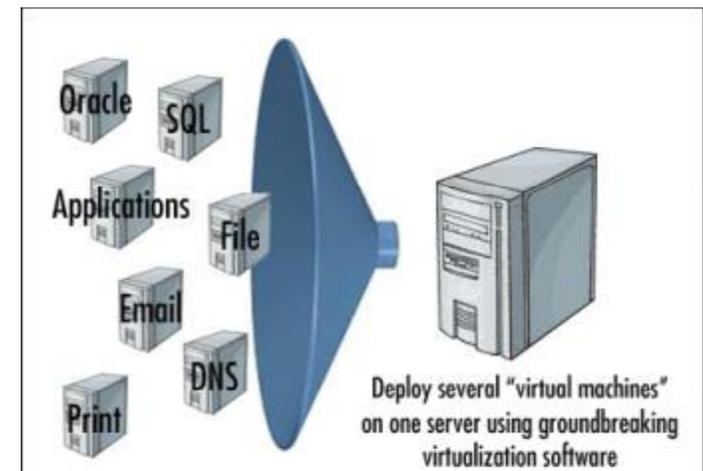


# History of Cloud Computing



# 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

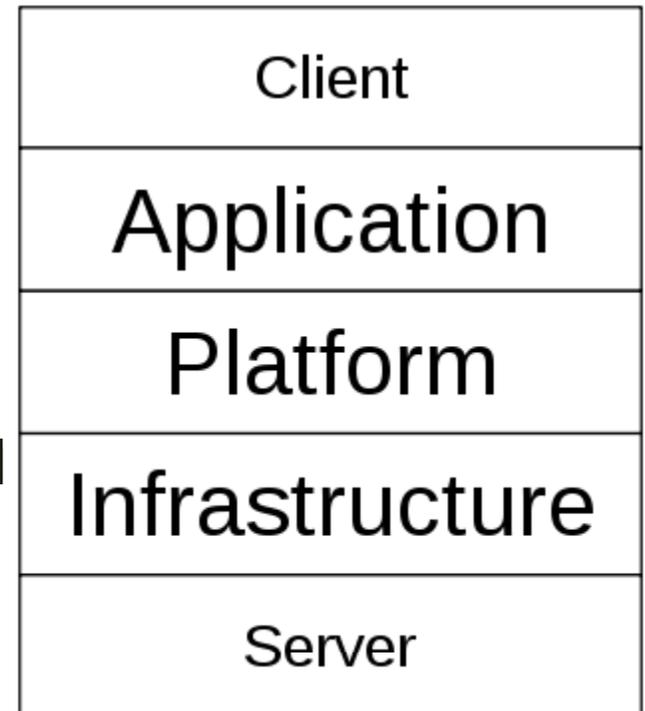
- Public, Community, Private, Hybrid

- **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.

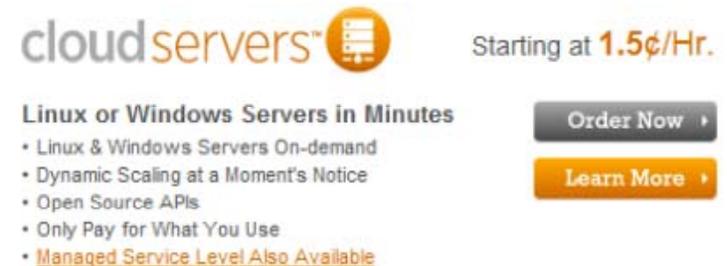


cloudfiles™  Starting at **15¢/GB**

Unlimited, On-demand Storage...

- Secure, Scalable Online Storage
- Only Pay for what You Use
- Easy to Access Via File Manager or API
- Serve Content at Blazing Speeds with Akamai CDN
- Powered by OpenStack

[Order Now](#) [Learn More](#)



cloudservers™  Starting at **1.5¢/Hr.**

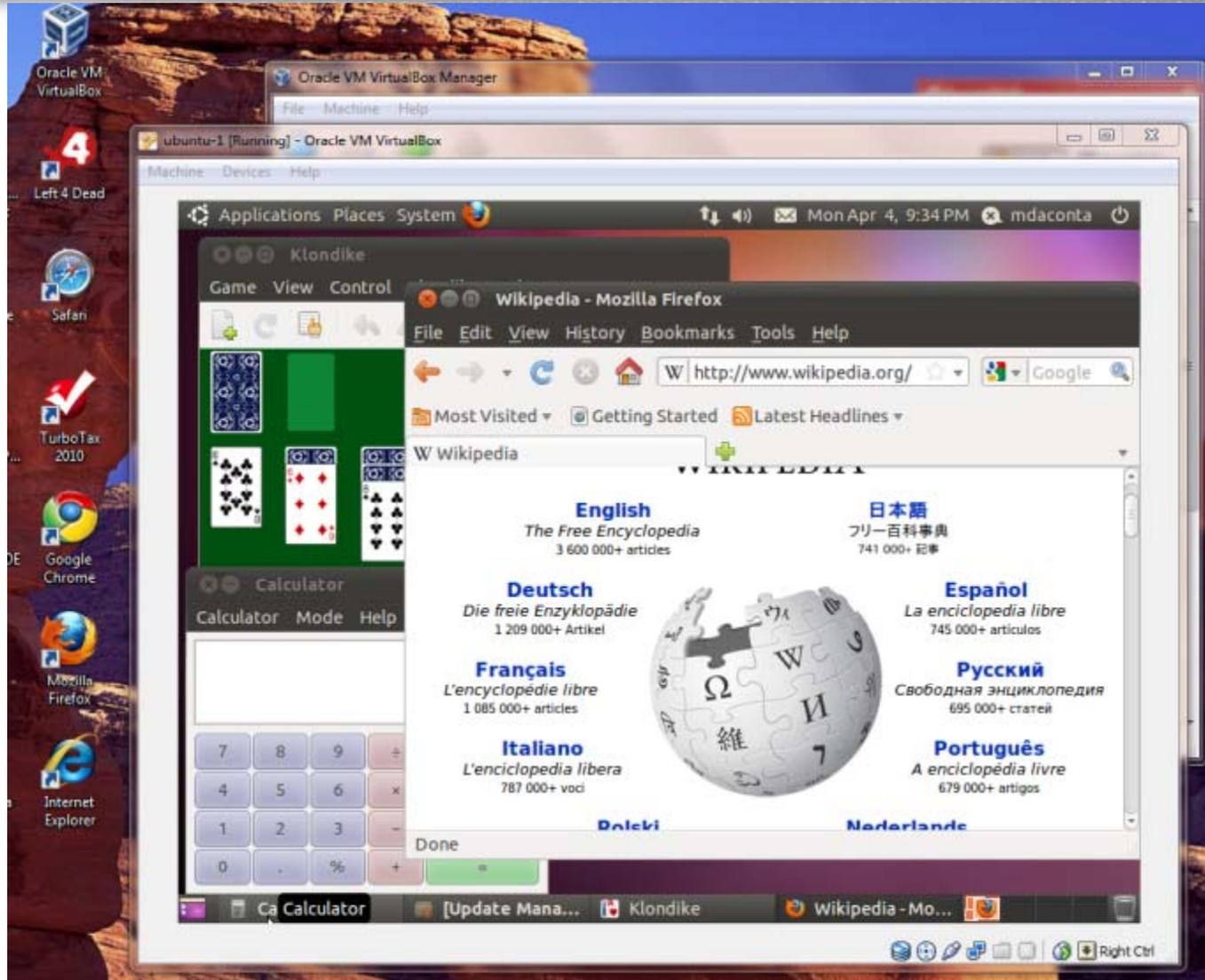
Linux or Windows Servers in Minutes

- Linux & Windows Servers On-demand
- Dynamic Scaling at a Moment's Notice
- Open Source APIs
- Only Pay for What You Use
- [Managed Service Level Also Available](#)

[Order Now](#) [Learn More](#)

Rackspace® Pricing

# 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.



Bungee Connect Offering

# 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!



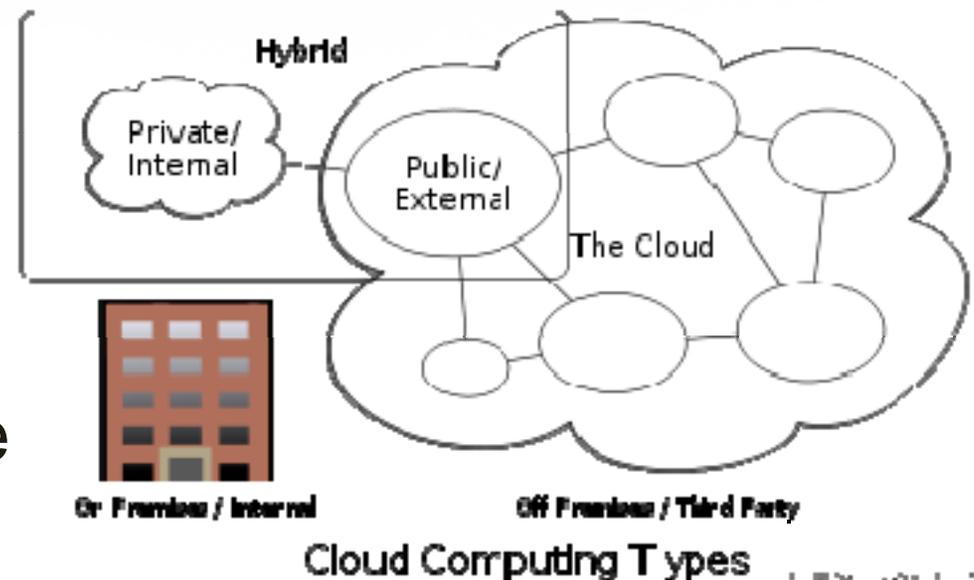
The screenshot displays the GSA Apps.Gov website, which is a service provided by GSA. The page is titled "Business Apps" and features four application categories, each with a representative icon, a brief description, and a "VIEW ALL" button:

- Analytics:** Data analyzing software that helps your agency make decisions and improve processes.
- Asset management:** Software that may include financial analysis, asset selection, and other financial software.
- CRM:** Track and organize info about your current and prospective customers to improve service.
- Communication:** Software that helps facilitate communication between various groups or systems.

In the bottom right corner of the screenshot, the logo for InCadence Strategic Solutions is visible, featuring a stylized compass rose icon.

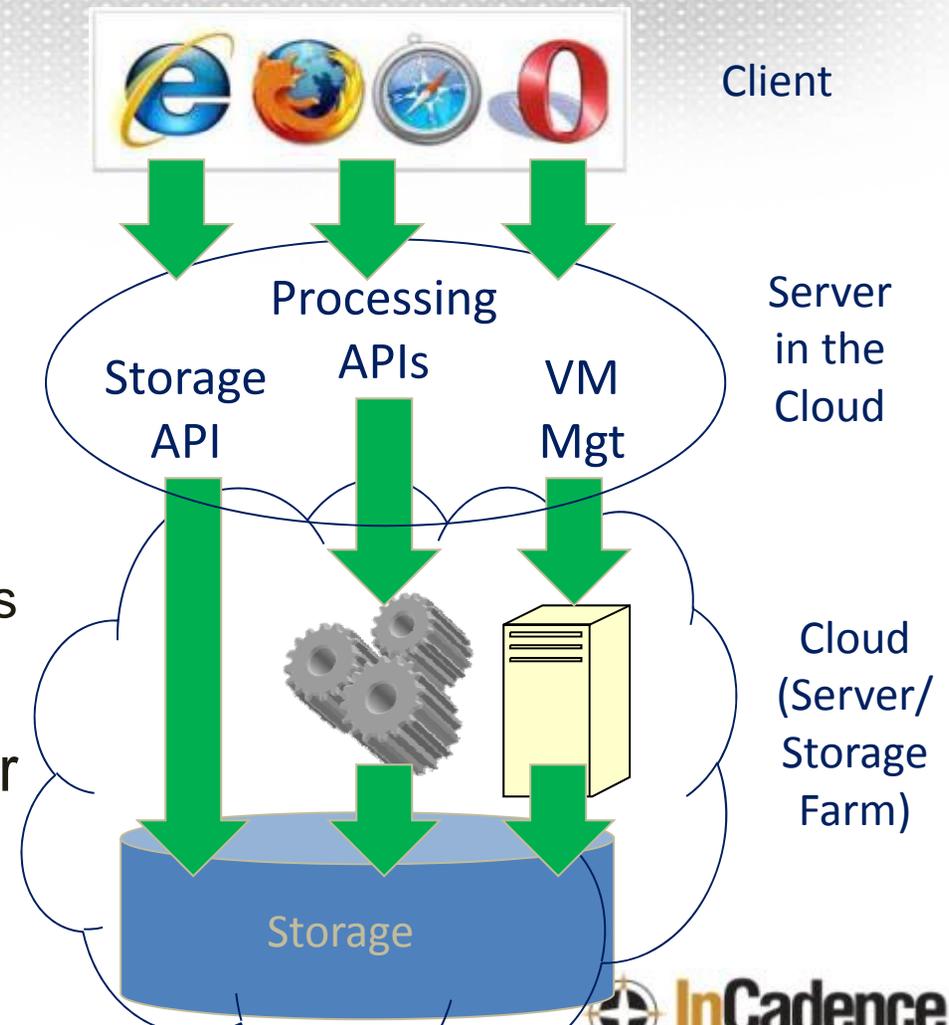
# Cloud Deployment Options

- 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?

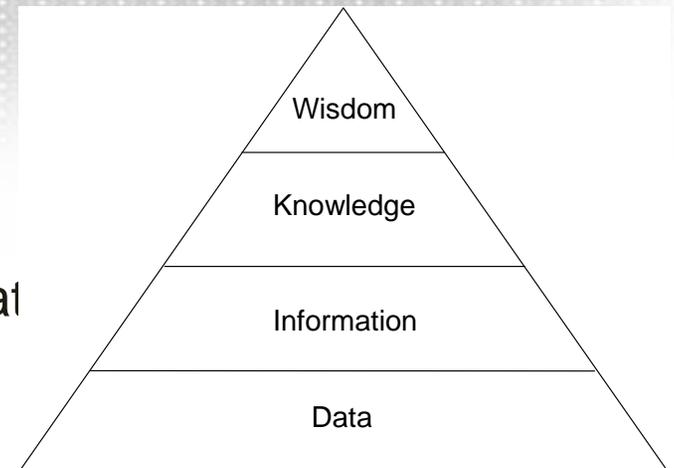
- **Labor** is the real cost sink... yet another REWRITE?
- If you started a SOA, you are ahead of the curve!



# IM & Semantics

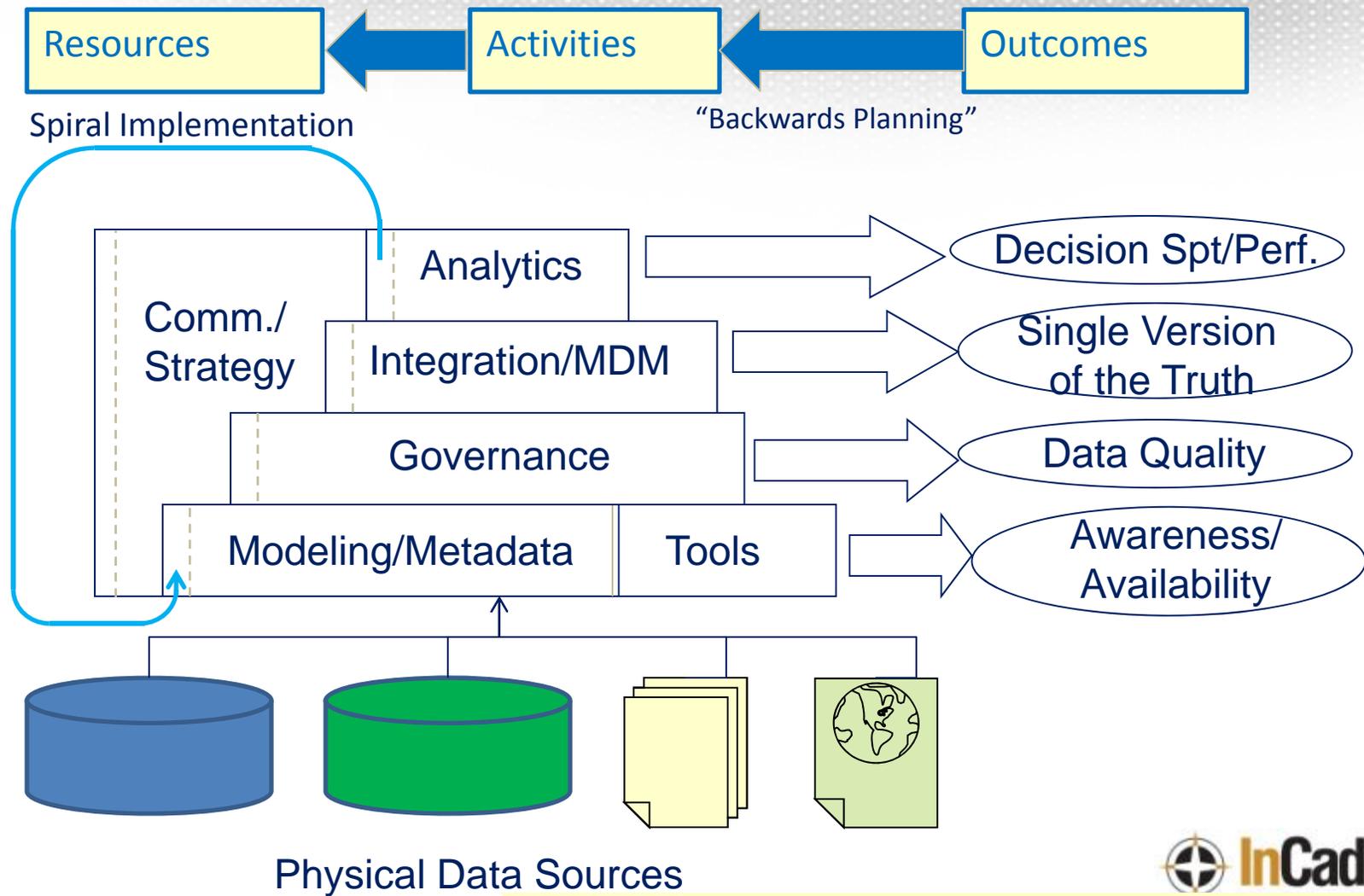
# Data is Not Information

- 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 ...

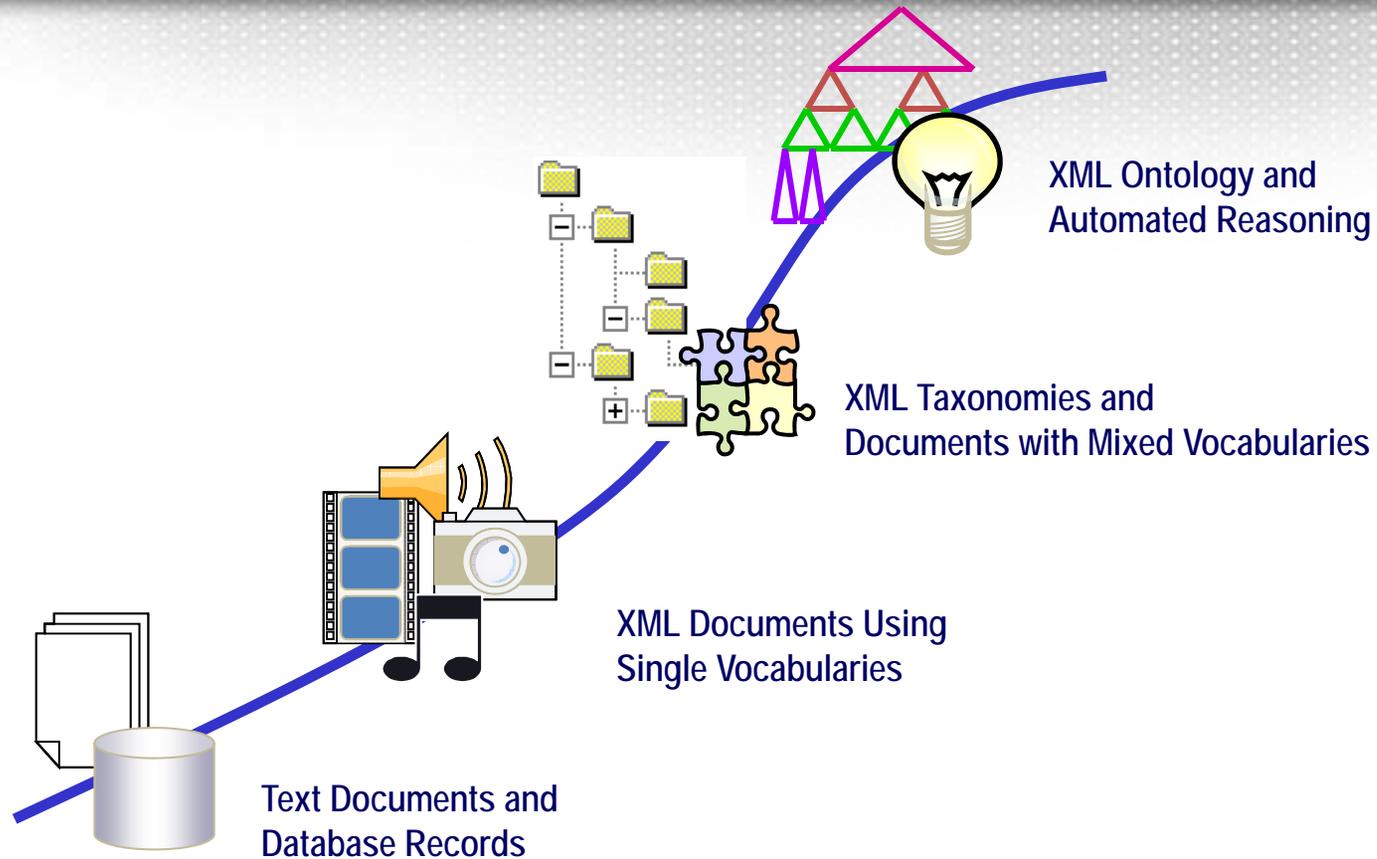


STRATEGIC SOLUTIONS

# Outcome-Based Information Management



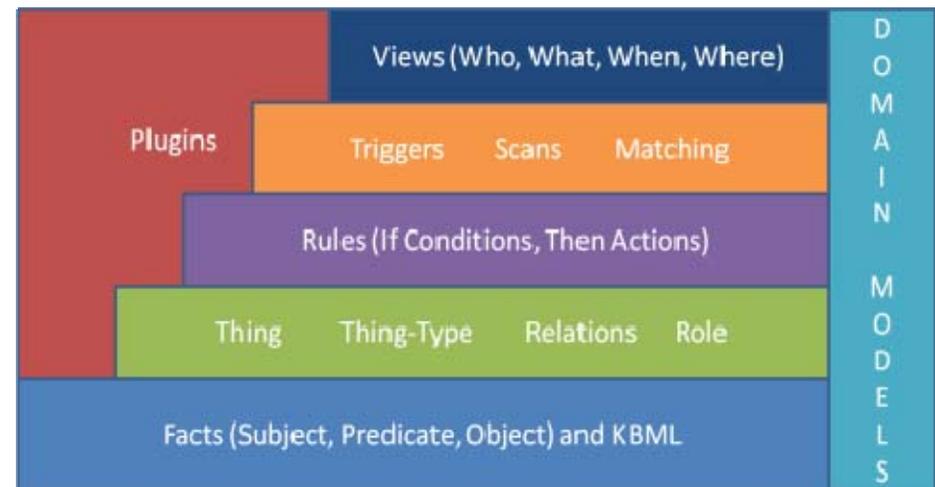
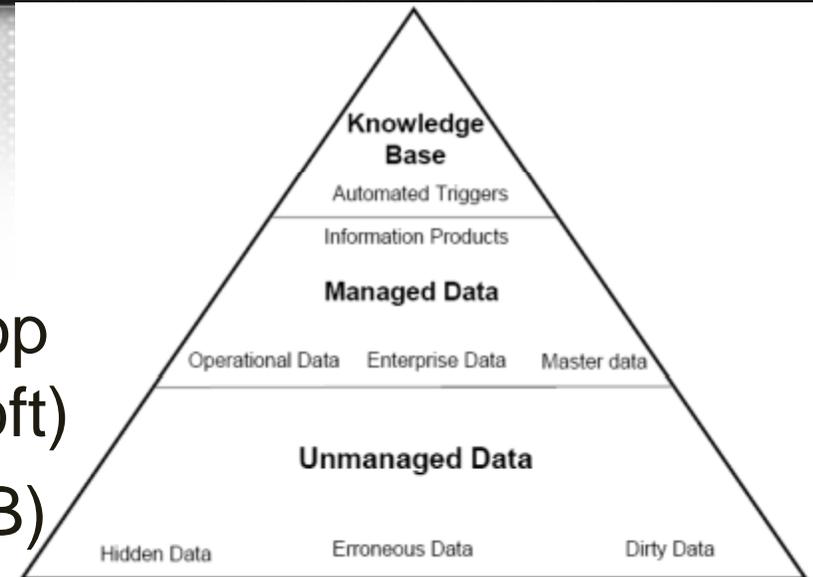
# Smart Data Continuum



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

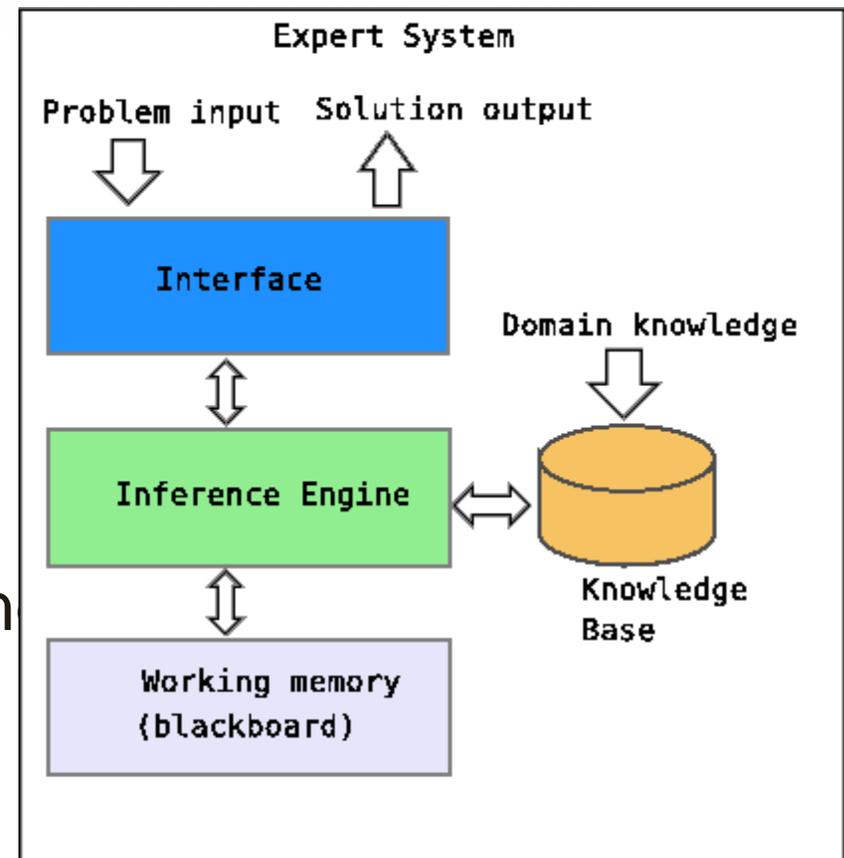
# From Information To Knowledge

- Data Optimizaton 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



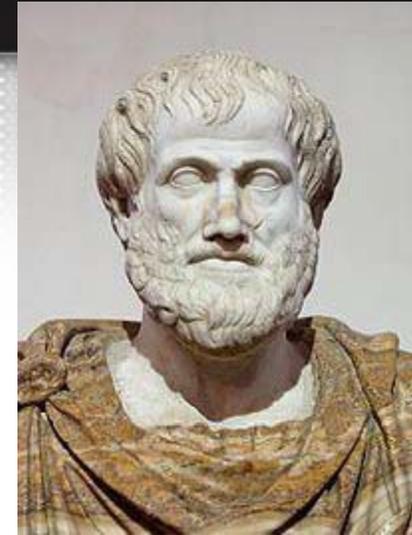
# The Semantic Solution

- 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

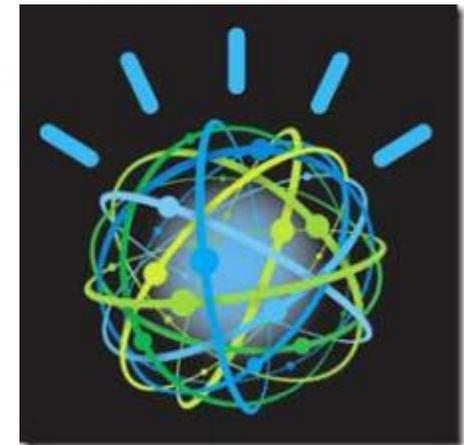
- **Logic:** a system of reasoning. Began with Aristotle:
  - Example of an Aristotelian Syllogism:
    - Major Premise: All Men are Mortal.
    - Minor Premise: Socrates is a man.
    - Conclusion: Socrates is Mortal.
  - Many types of logic: Predicate Logic, Modal Logic, Propositional Logic
- **Inference:**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



$p$	$q$	$p \rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

# 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”

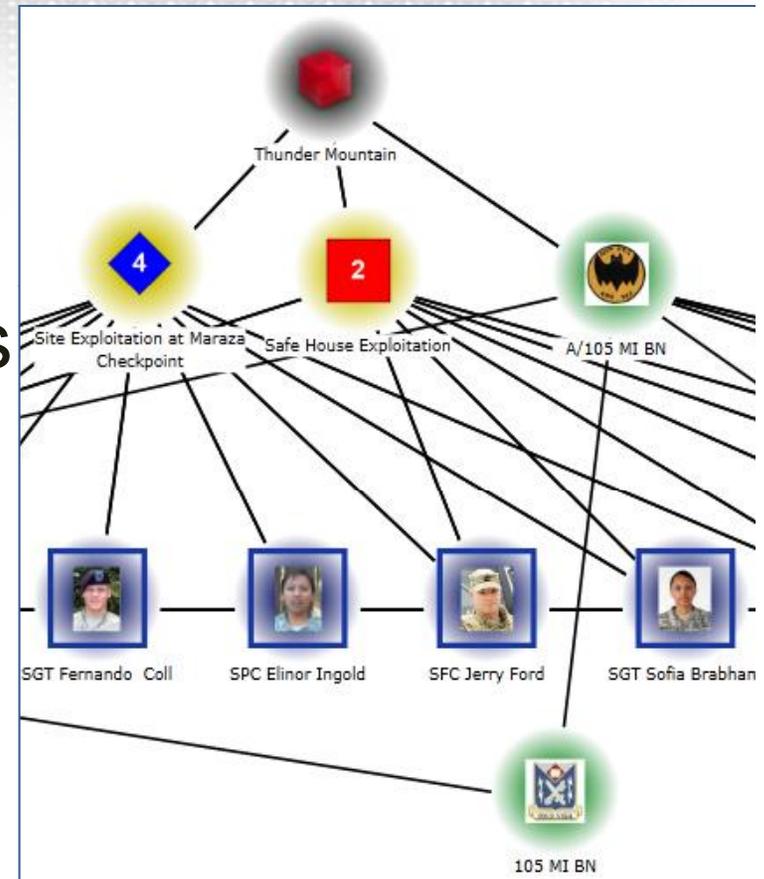


Google Refine



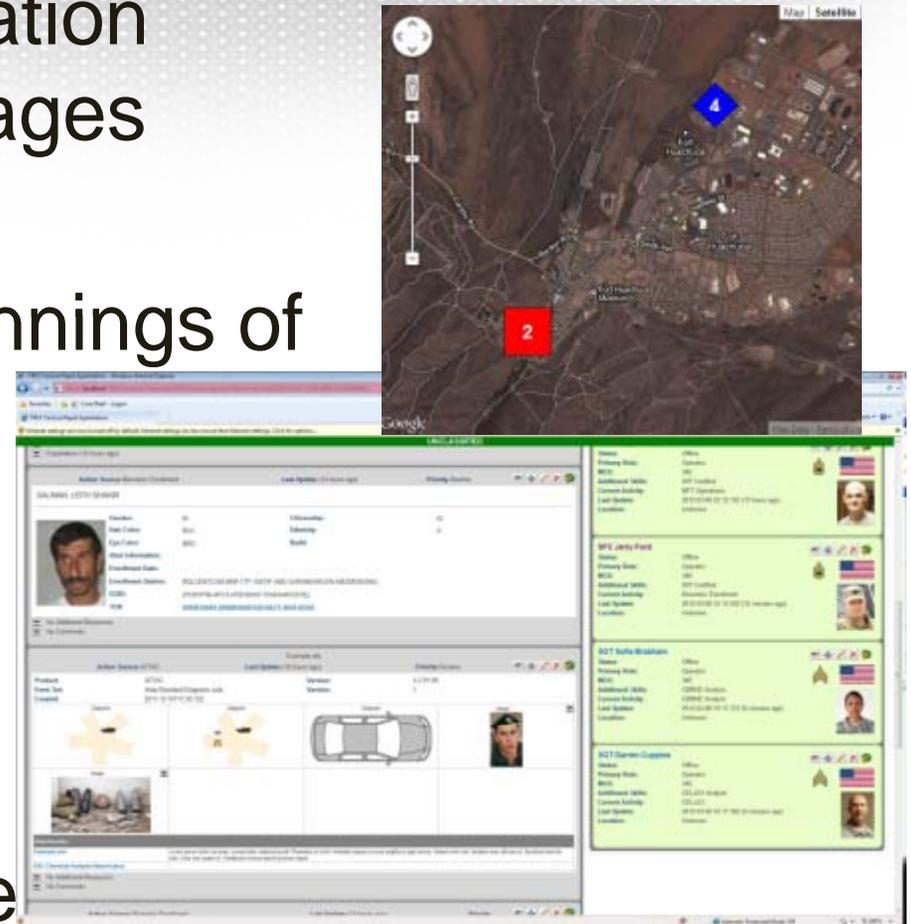
## Semantics in Biometrics (1 of 2)

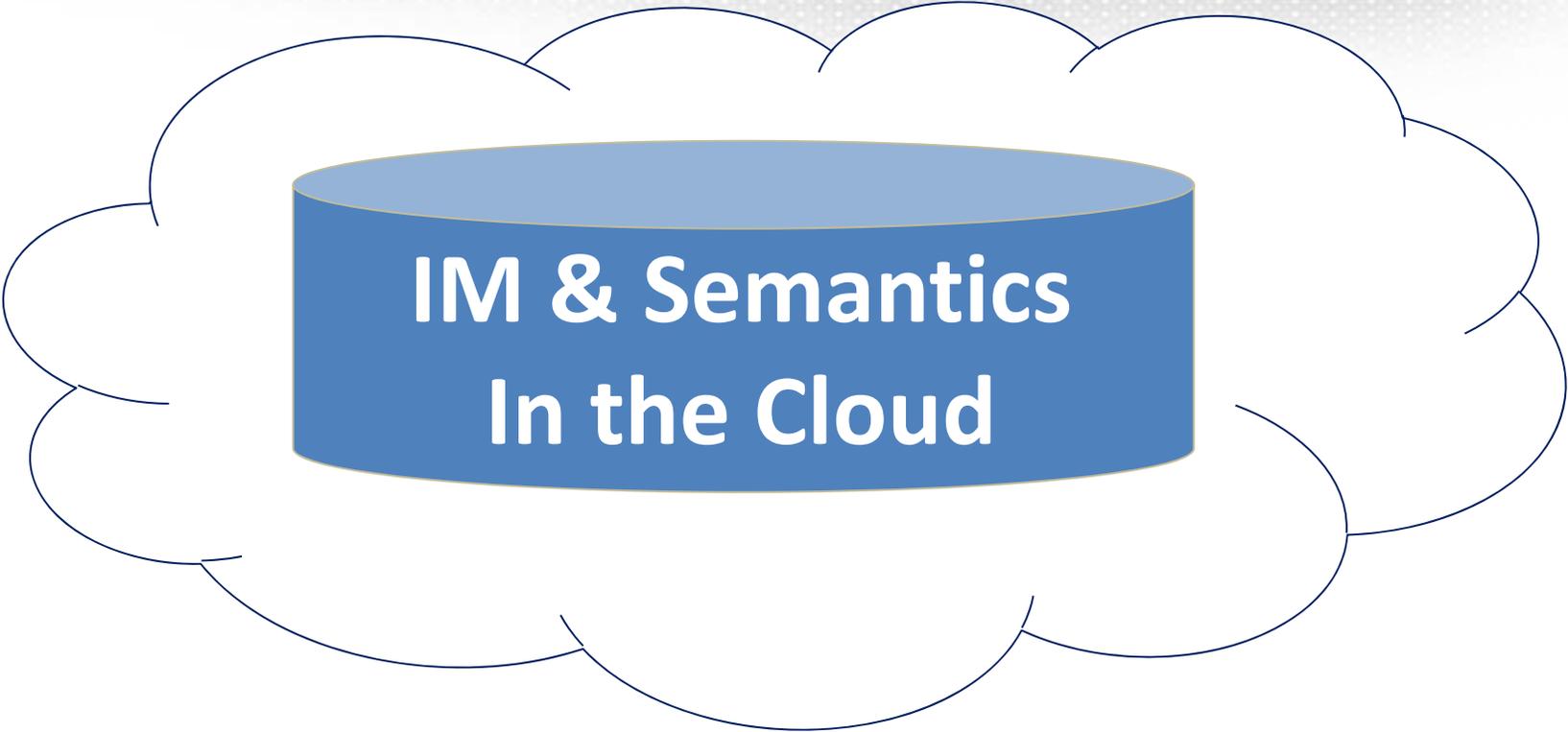
- The biometrics system MARS uses a declarative objects system called “Coalesce”.
- Biometrics and Identity Management are rich grounds for semantics.
  - Person, Relationships, Organizations, Actions, etc.
  - Facebook’s Social Graph is semantics!
- Coalesce is declarative, XML
- Coalesce is NOT OWL, yet...



## Semantics in Biometrics (2 of 2)

- The Tactial Rapid Exploitation (T-REX) portal also leverages Coalesce entity-graphs.
- Key foundational underpinnings of semantics are becoming commonplace:
  - Rich metadata, Rich 5W, Rich relationships, Declarative business rules
- I am working on Coalesce 2.0: security, semantics and the cloud.





# IM & Semantics In the Cloud

# What is Cloud Data like?

- Data Storage is different

- No-SQL movement, key-value pairs.

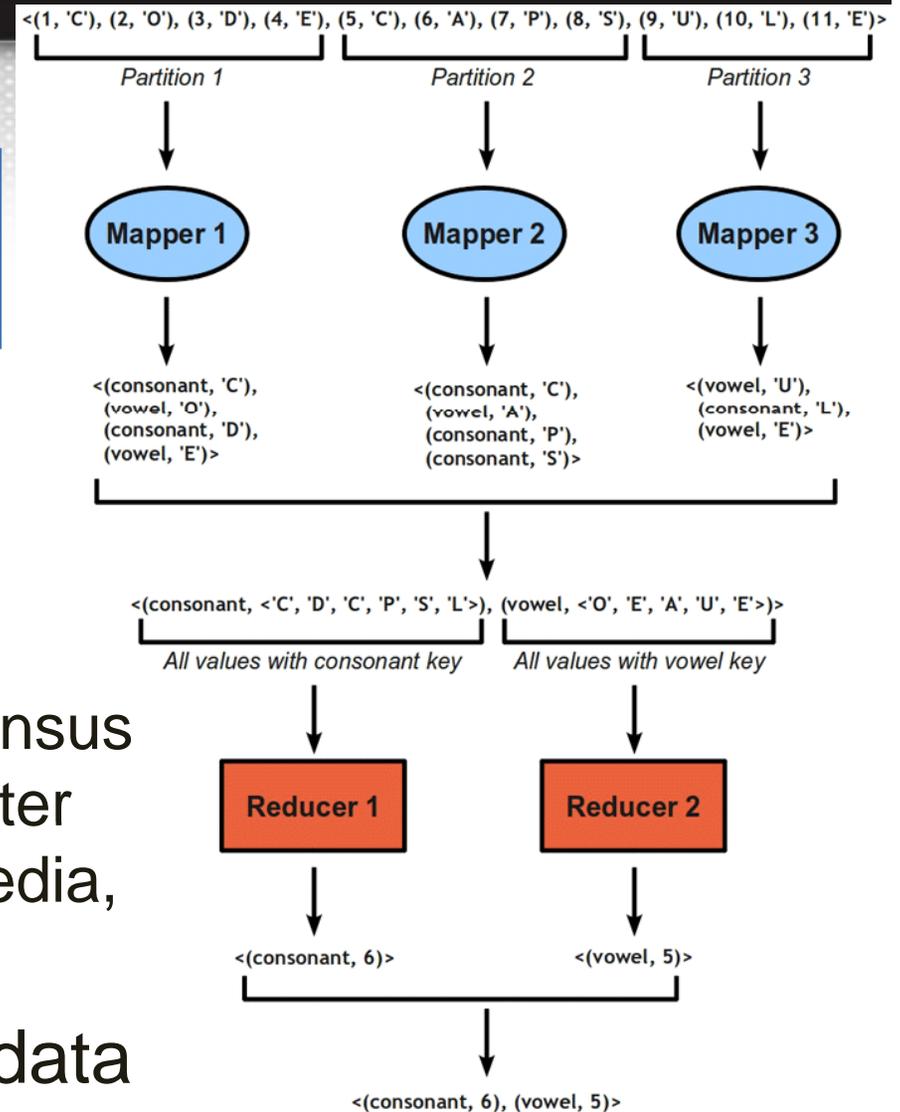
```
< Hello, 1>  
< World, 1>  
< Bye, 1>  
< World, 1>
```

- 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...

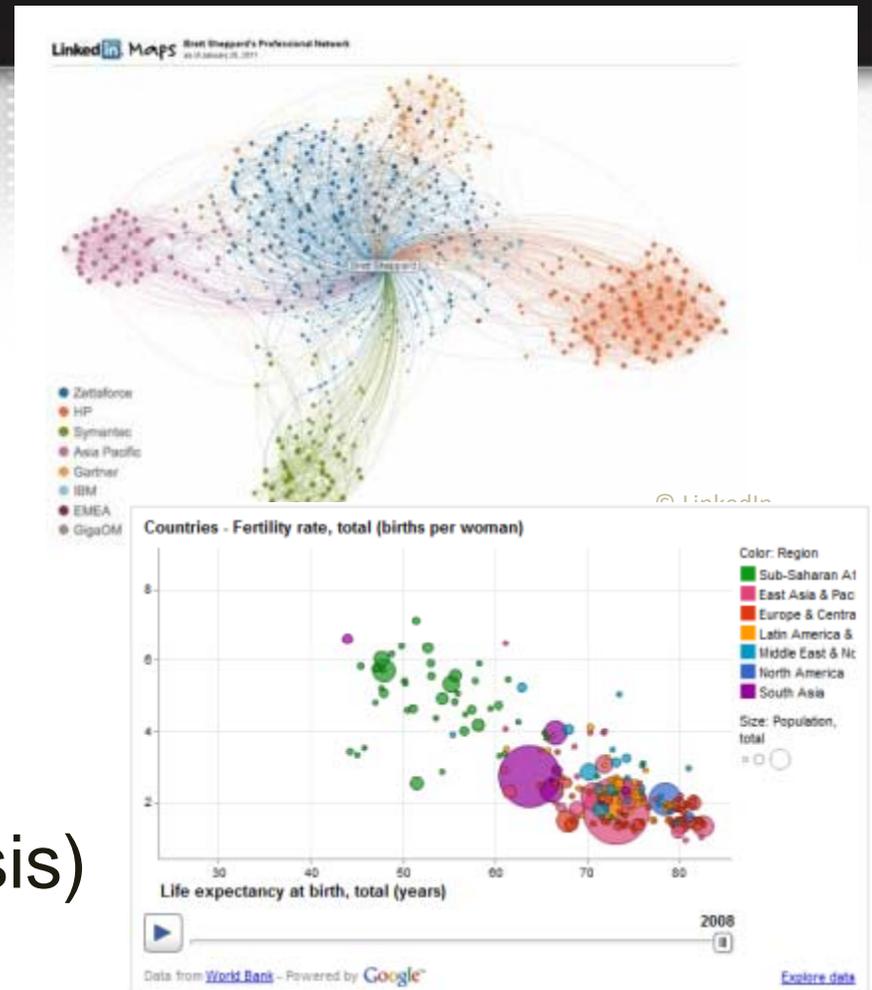
- A new Data Type? Stream data



STRATEGIC SOLUTIONS

# 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)

INFORMATICA | CLOUD



## 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.

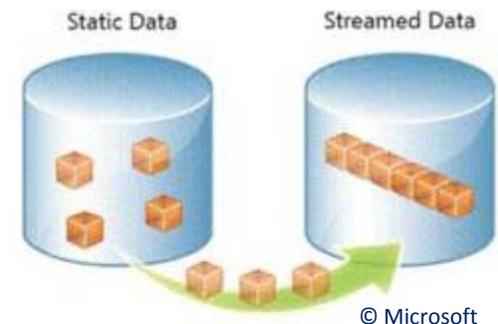
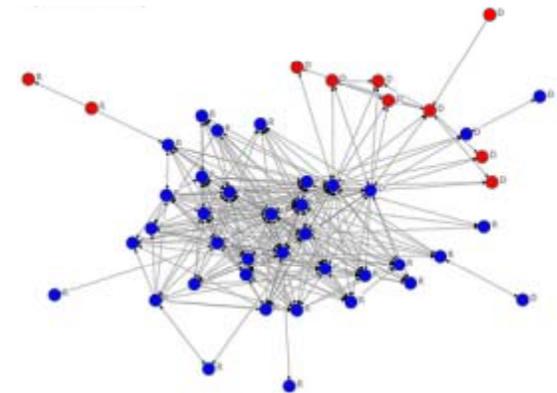
### ■ Private

- Most secure option, no data ownership issues



# Where are Semantics in the Cloud?

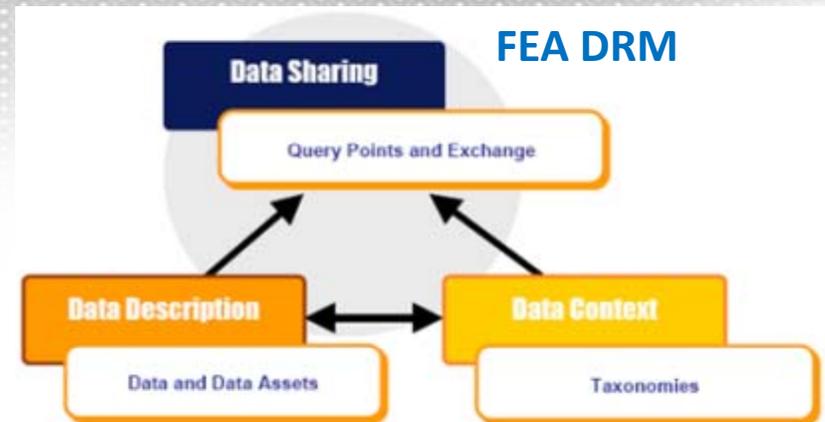
- Semantic Markup is a natural fit for the cloud – why?
  - Semantic Web is Distributed and the Cloud is Distributed
- Ok, HOW?
  - Hadoop = streaming data
  - Graph Traversal -> streaming data
- In other words,  
“Stream the Graph”
- Mark Logic Hadoop Connector
  - See: Session by MarkLogic and Hortonworks 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!

