UNCLASSIFIED

Naval Center for Cost Analysis

Estimating Information Technology Costs

July 2011
Version 1.0

POC: Corinne Wallshein
IT Technical Advisor, NCCA
corinne.wallshein@navy.mil

Overview

1. Terms

2. Measurement impacts

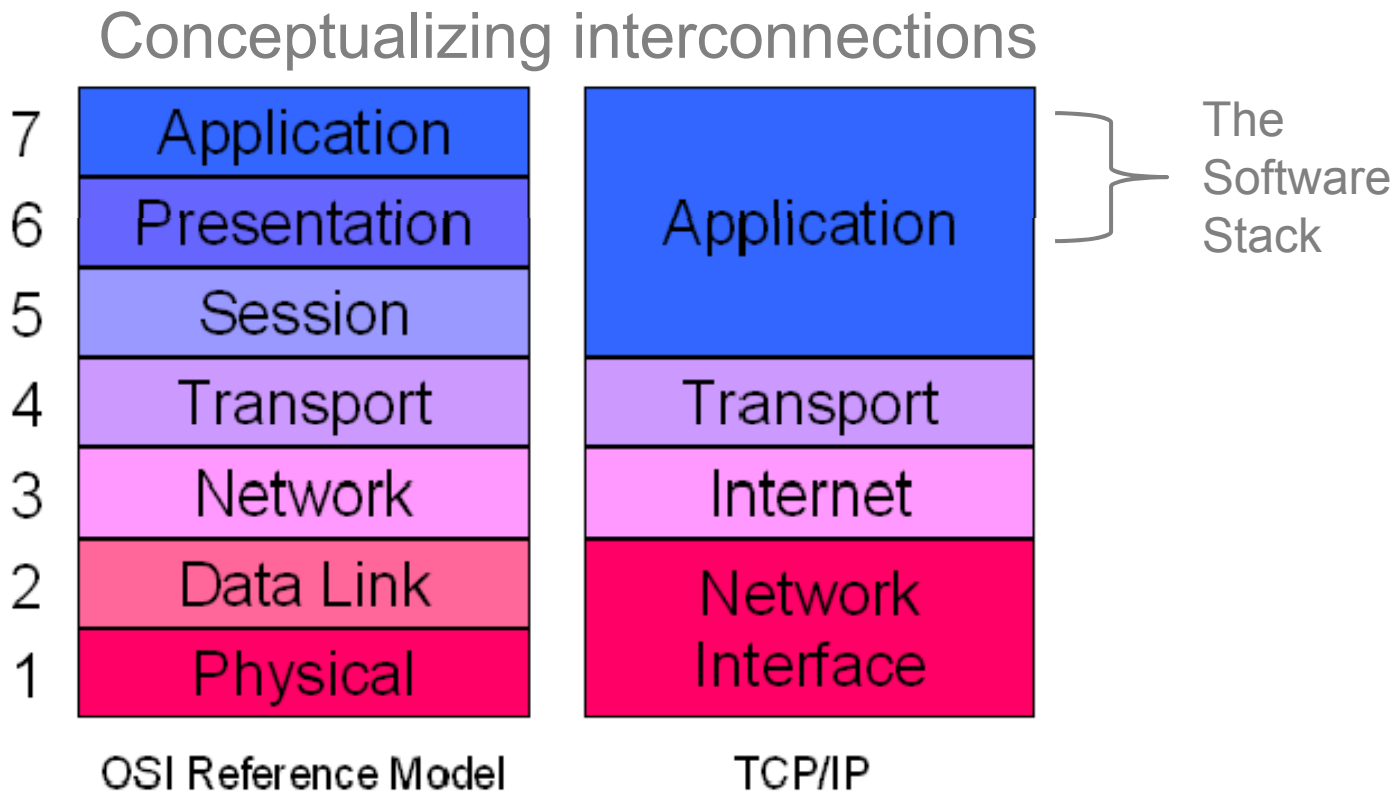
3. Relationship to costs

Where did ‘cloud’ come from?

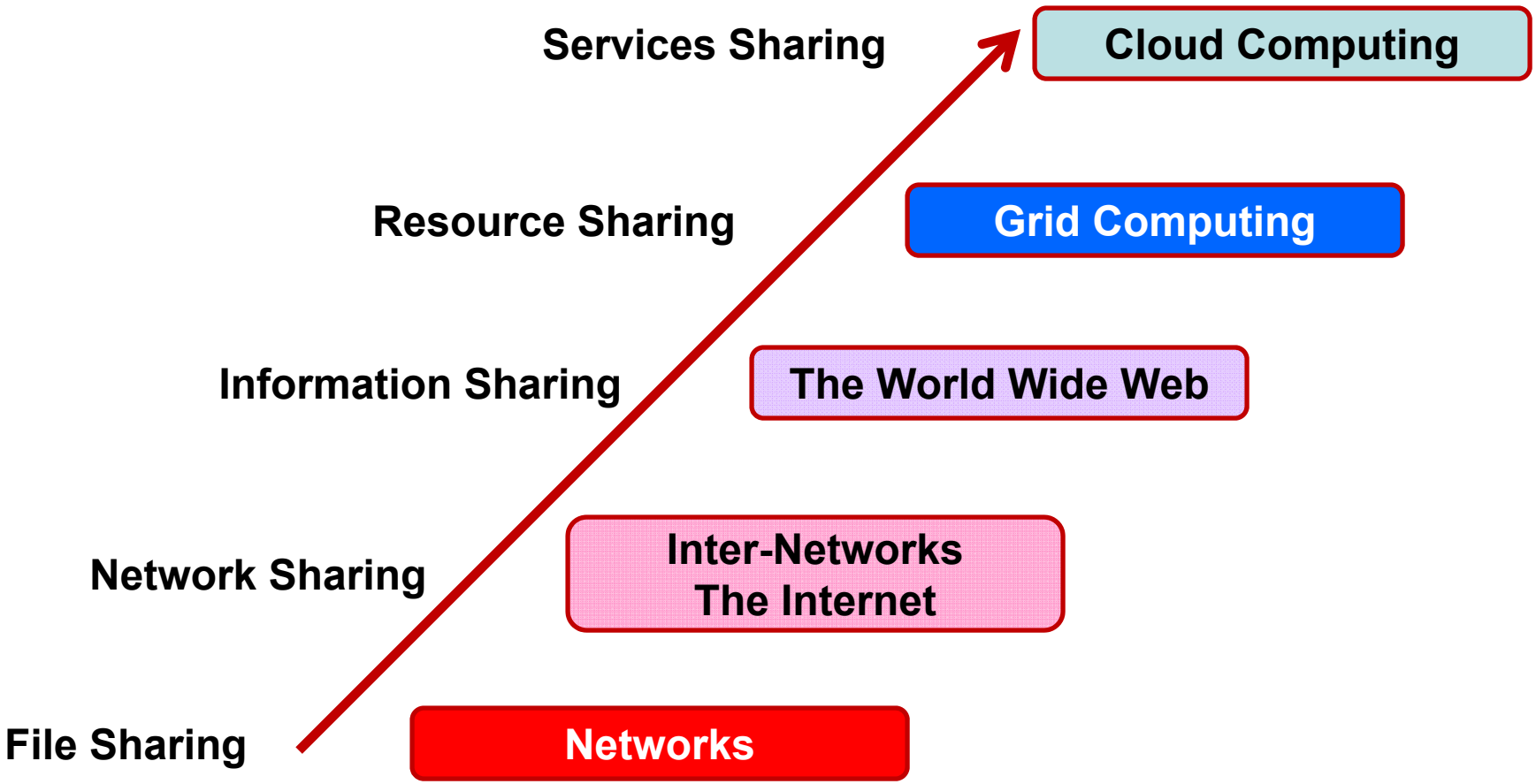
- The term “cloud” is from the 1990’s
- Users store and transfer software and data
- Cloud security responsibility questions and legal issues (e.g., e-discovery, regulatory compliance, international boundaries, auditing…)*
- In March 2011, a “hot topic” for Defense Acquisition University (DAU)

* *Data Security in the World of Cloud Computing*, Harauz, Kaufman and Potter, 2009, IEEE

Seven OSI layers and four TCP/IP



Evolution based on OSI and TCP



Concepts and models for clouds

Essential Characteristics

On-demand Self-Service
Broad Network Access
Resource Pooling
Rapid Elasticity
Measured Service

Cloud Service Models

Software as a Service (SaaS)
Platform as a Service (PaaS)
Infrastructure as a Service (IaaS)

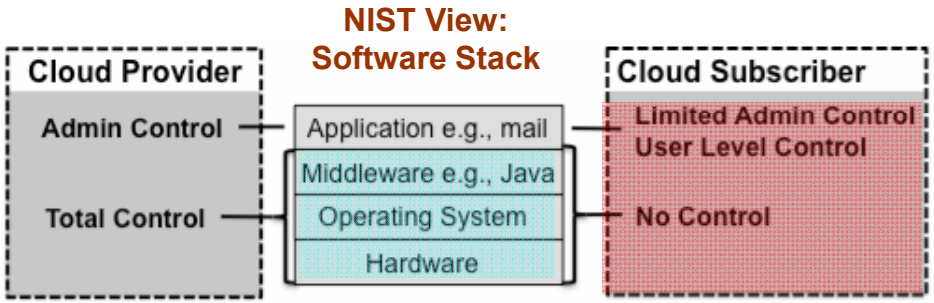
Cloud Deployment Models

Cloud - Service Level Agreements

SLA Area	Measurement	Description
Availability	Percent of Downtime	% of infrastructure downtime
Availability	Mean Time to Repair	Average time required to repair a failed component or device
Availability	Response Time	Time to react to a given input
Availability	Percent of Up time	% of infrastructure uptime
Portability	Level of portability	Use of standard vs. custom interfaces
Performance	CPU Utilization	The amount of time in use (not idle)
Performance	Disk Performance	Total job completion time
Quality of Service	Latency	Time delay experienced in a system
Configuration	Basic Configuration	Time to deploy preconfigured OS or application
Configuration	New OS or Applications	Time to deploy new OS or application

Network service metrics now applied to cloud service models

Software as a Service (SaaS)



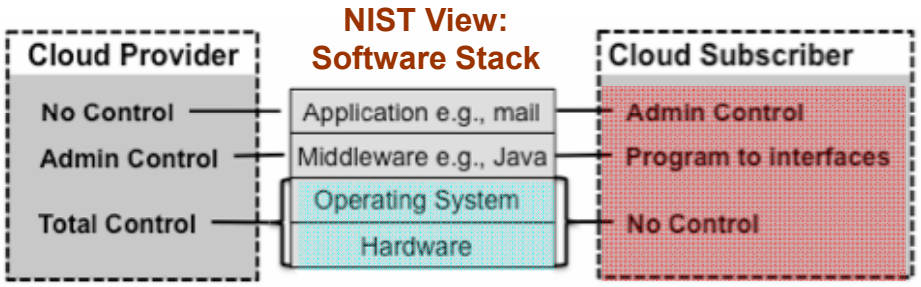
- Measurements to collect
  - Number of users
  - Time in use
    - Per-execution
    - Per-record-processed
  - Amount of network bandwidth consumed
  - Quantity/duration of data stored
- Examples
  - Google Apps
  - Microsoft Exchange
  - Cisco WebEx WebOffice
  - Oracle CRM On Demand
  - SalesForce.com
  - Yahoo Mail

Subscriber able to use provider's software applications

Diagram from NIST draft synopsis, page 5-3

Examples from Bojanova & Samba, 2011, IEEE

Platform as a Service (PaaS)



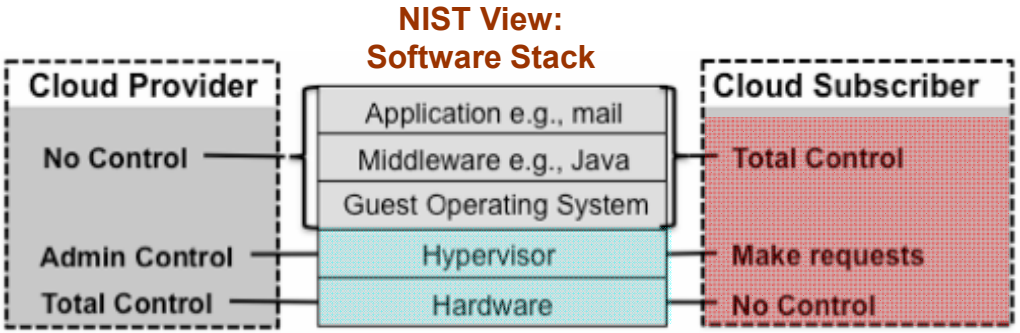
- Measurements to collect
  - Number of subscribers by kind
  - Amount of storage
  - Amount of processing
  - Amount of network resources consumed by the platform
  - Number of requests serviced
  - Amount of time platform is in use
- Examples
  - Oracle Fusion Middleware
  - Google AppEngine
  - Amazon Web Services
  - Facebook
  - Microsoft Azure

Subscriber able to build and deploy software with provider tools

Diagram from NIST draft synopsis, page 6-3

Examples from Bojanova & Samba, 2011, IEEE

Infrastructure as a Service (IaaS)



- Measurements to collect
  - Number and types of subscribers acting as system administrators
  - Proprietary cloud provider metrics
  - Providers' tracking of resources for billing purposes (e.g., SLA meters and rates)
  - W4H3
- Examples
  - IBM
  - Amazon Elastic Compute Cloud (EC2)
  - Rackspace Cloud
  - Microsoft Azure
  - Sun (Oracle)

Subscriber able to deploy and run any compatible software

Diagram from NIST draft synopsis, page 7-2

Examples from Bojanova & Samba, 2011, IEEE

Relating to building software…

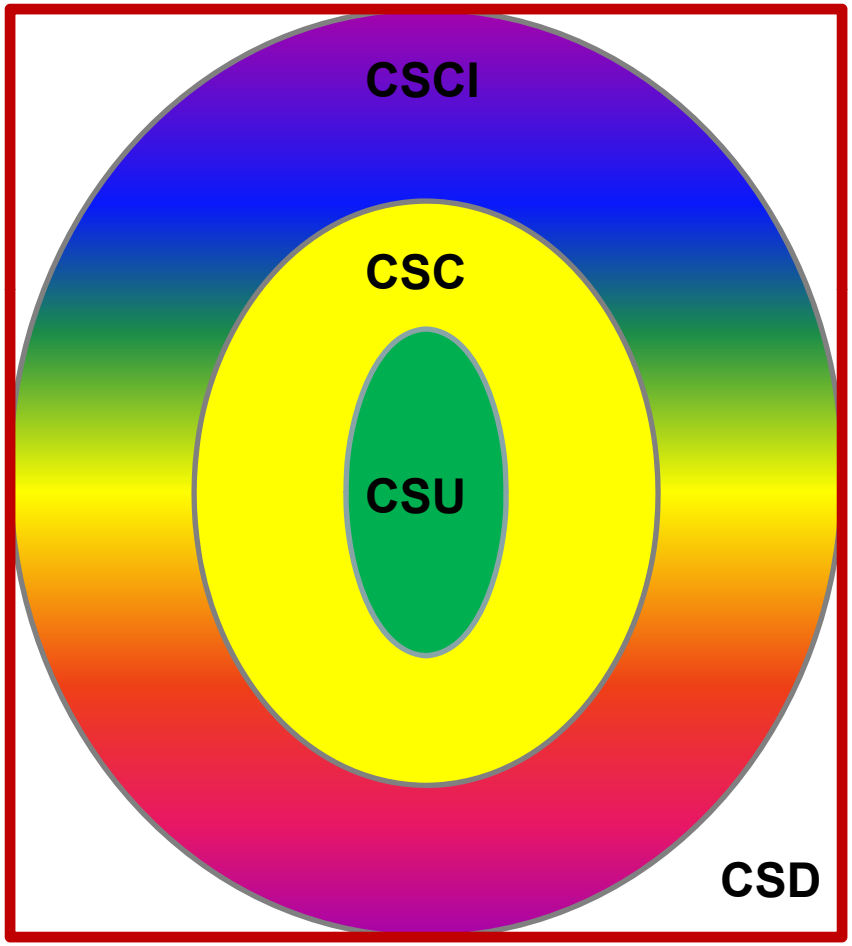
- Software Cost Estimation (SCE) involves prediction, often early in a program, of resources to produce software

- Resources of human effort translate to costs because people are paid to produce software

DAU terms for software building

Software

Computer Software



http://www.dau.mil/pubscats/PubsCats/13th_Edition_Glossary.pdf accessed on 5-23-2011

Fundamental SCE Effort Equation

$$E=c(Size)^b$$

E = Development Effort

c = Environment Calibration Constant

S = Size of Source Code (includes reuse effects)

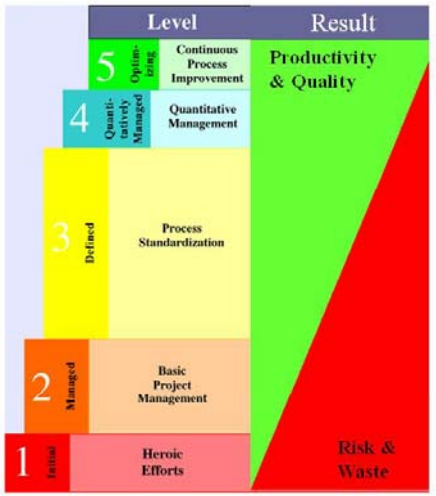
b = Entropy Constant (e.g., complexity, productivity…)

Adapted from *'What Model is Right for Me?'* presented by Joe Dean at DoDCAS, 2010

DoD Data Repositories

- **Program Data in DAMIR**
  - EVMS data…total contract
  - Not by delivery order

- **Cost and Software Data in DACIMS**
  - SRDR software data
  - Two snapshots: beginning and end
  - No quality data: CMMI as proxy



Earned Value Management System (EVMS) Data

- By contracted work ‘control package’ progress can (sometimes) be determined
  - **Schedule Performance Index: SPI = BCWP/BCWS**
  - **Cost Performance Index: CPI = BCWP/ACWP**

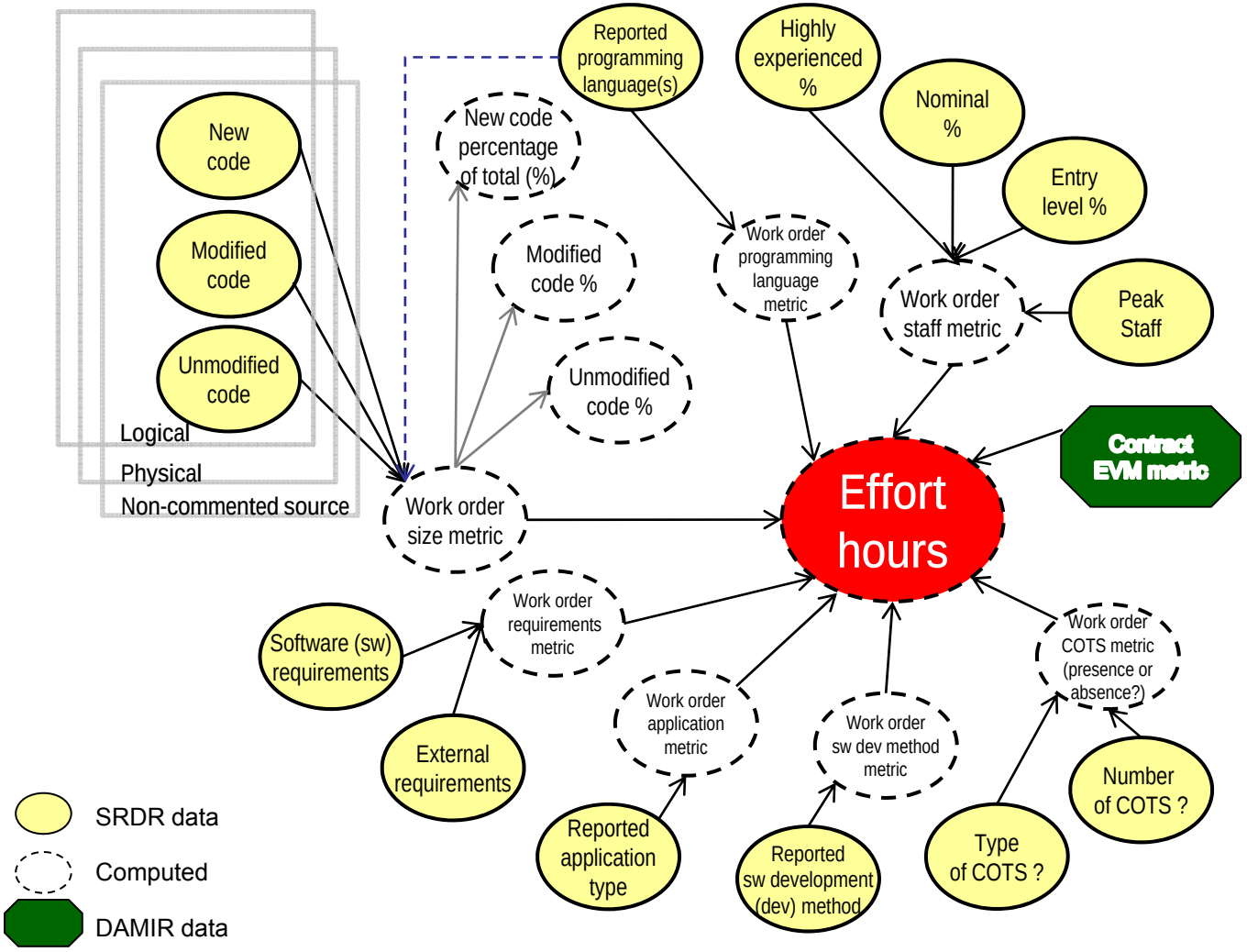
- SPI and CPI have similar values and ranges

- No Technical Performance Index

A new way to process the data

- Match initial and final SRDR data
  - Assume CMMI Level 4 or 5 in final SRDR ≈ 'High Quality'
  - Analyze each level separately
  - Normalize software size to Logical Statements (LS)
- Extract DAMIR EVMS metrics by contract number
- Experiment with subsets: User, System, Support
- Analyze data set

Variables Identified



Relationships in the data set

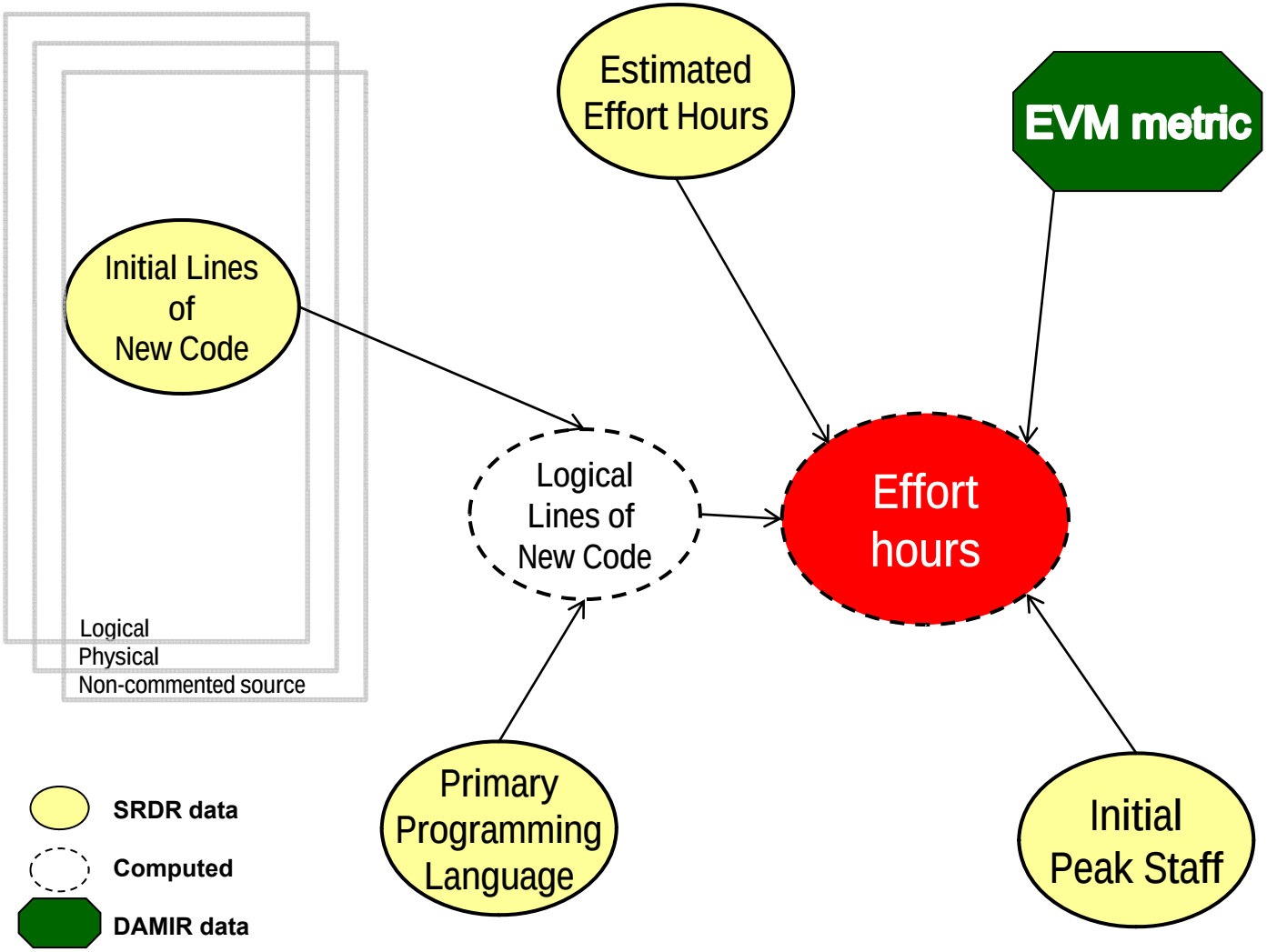
- Linear relationships

Pearson (Linear) Correlation to Final Effort Hours in Thousands (K)	CMMI Level 5	CMMI Level 4
Initial Effort Hours (K)	0.65	0.97
Initial Peak Staff	0.79	0.98
Initial New Lines of Code (KLOC)	0.51	0.93

- Monotonic relationships

Spearman (Non-Linear) Correlation to Final Effort Hours in Thousands (K)	CMMI Level 5	CMMI Level 4
Initial Effort Hours (K)	0.79	0.93
Initial Peak Staff	0.82	0.65
Initial New Lines of Code (KLOC)	0.77	0.59

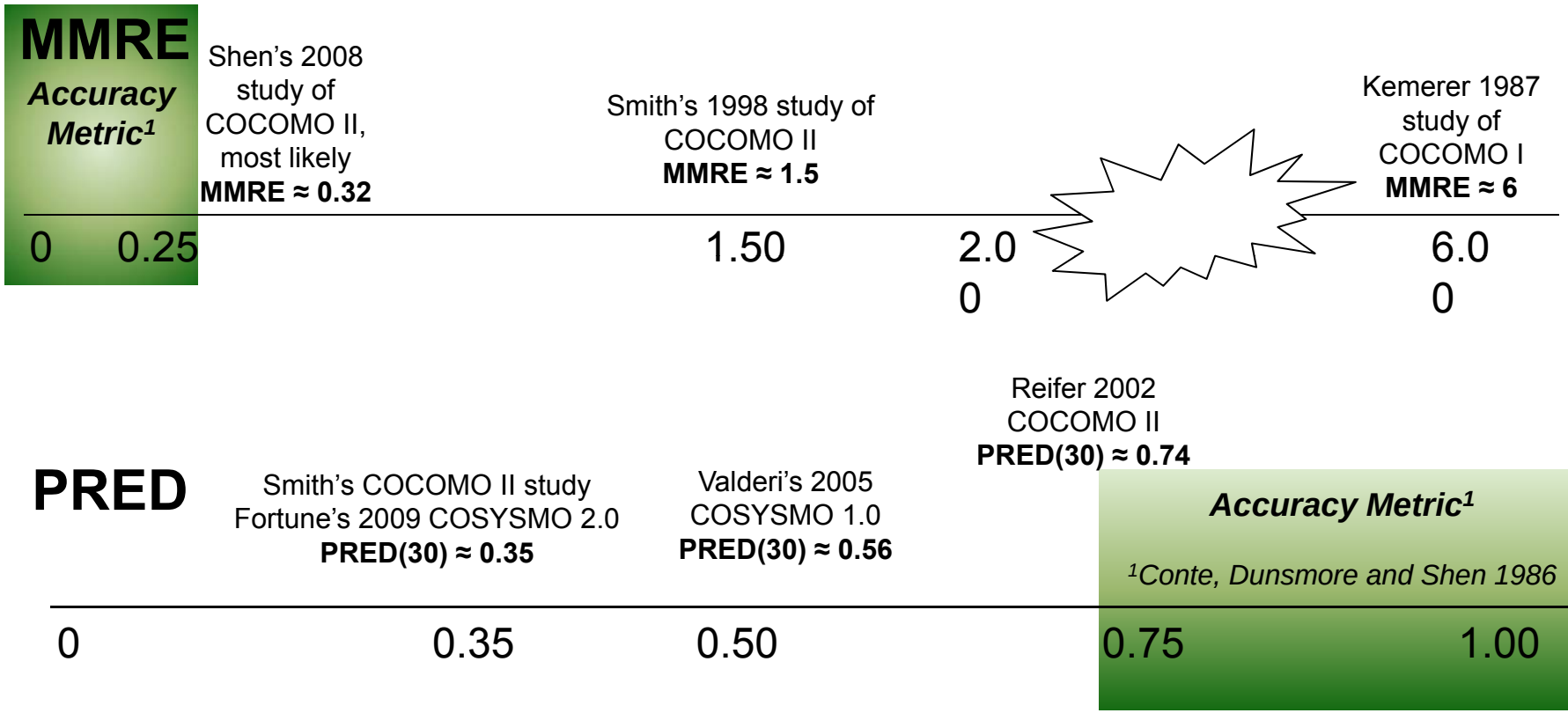
Diagram of Variables



Generate CER Equations

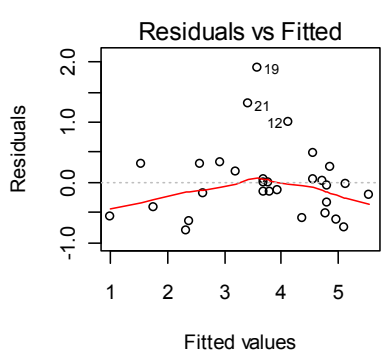
- By CMMI level
- Independent variables to predict dependent variable (effort hours)
- Compute accuracy metrics
- Analyze residuals

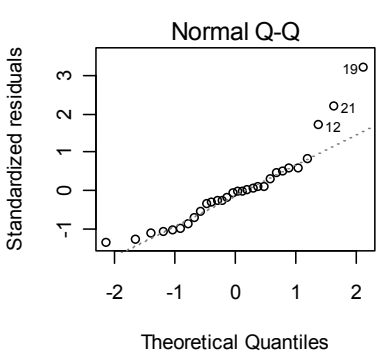
Wide-spread Accuracy Metrics

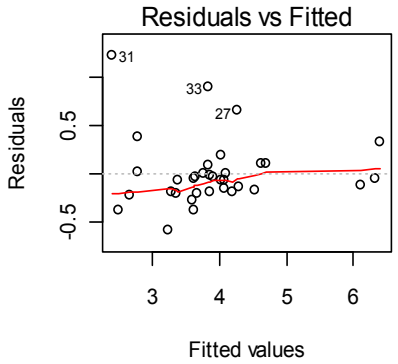


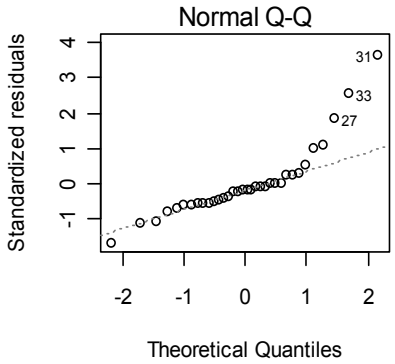
MMRE is the Mean Magnitude of Relative Error (MRE)
PRED(L) = X where if L = 30, means the individual MRE is ≤ 0.30
X is between 0 and 100 as the percentage of data meeting this condition

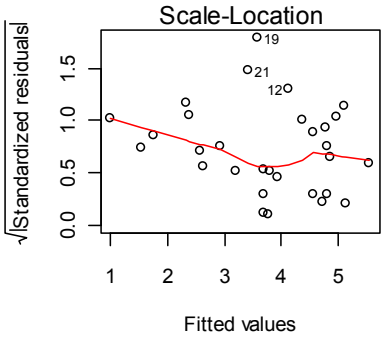
Log-Log (Est. Effort Hours)

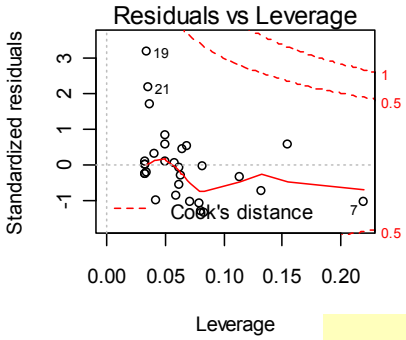




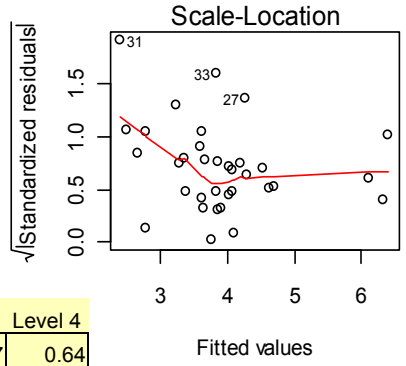


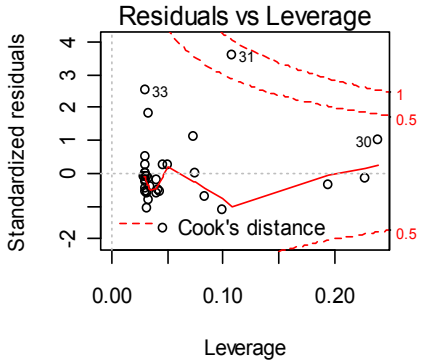






	Level 5	Level 4
Intercept	0.47	0.64
Coefficient	0.89	0.88
Adjusted R2	0.79	0.87
p-value	< 0.001	< 0.001
MMRE	0.69	0.36
PRED(30)	0.27	0.50





CMMI Level 5

CMMI Level 4

Same procedure, different variables

Independent Variable	*Transform*	*CMMI Level*	*Intercept*	*Coefficient*	*Adjusted R2*	*p-value*	*MMRE*	*PRED(30)*
Est. Effort Hrs	None	Level 5	N/A	N/A	**Too Low**	N/A	N/A	N/A
Est. Effort Hrs	None	Level 4	10.95	0.94	0.94	< 0.001	0.28	0.71
Est. New Code	Log - Log	Level 5	1.65	0.68	0.70	< 0.001	0.98	0.27
Est. New Code	Log - Log	Level 4	2.86	0.48	0.64	< 0.001	0.77	0.26
Est. New Code	None	Level 5	N/A	N/A	**Too Low**	N/A	N/A	N/A
Est. New Code	None	Level 4	11.74	2.66	0.86	< 0.001	0.42	0.5
Est. Peak Staff	Log - Log	Level 5	1.32	0.96	0.74	< 0.001	0.87	0.40
Est. Peak Staff	Log - Log	Level 4	2.27	0.64	0.68	< 0.001	0.75	0.53
Est. Peak Staff	None	Level 5	22.1	2.44	0.60	< 0.001	1.53	0.50
Est. Peak Staff	None	Level 4	25.29	1.67	0.96	< 0.001	0.62	0.59

Partitioning by Applications

- For Level 5 Application Area:
  - ***User***: Accuracy metrics are improved for Peak Staff

Peak Staff	
MMRE for full set, CMMI Level 5	1.53
MMRE for User Subset, CMMI Level 5	0.44
MMRE for full set, CMMI Level 4	0.62
MMRE for User Subset, CMMI Level 4	0.43
PRED(30) for full set, CMMI Level 5	0.50
PRED(30) for User subset, CMMI Level 5	0.67

- For Level 4 Application Areas:
  - ***User***: MMRE improved for Peak Staff and Estimated (Est.) Hours
  - ***User & Support***: PRED(30) improved for Est. Hours

Estimated Hours	
MMRE for full set, CMMI Level 4	0.28
MMRE for User Subset, CMMI Level 4	0.14
PRED(30) for full set, CMMI Level 4	0.71
PRED(30) for User Subset, CMMI Level 4	1.00
PRED(30) for Support Subset, CMMI Level 4	0.82

Adding EVMS Metrics

- SPI boosts accuracy
  - Improved MMRE and PRED (30) for transformed New Code
  - Improved PRED(30) for Peak Staff

CMMI Level 5

Transformed New Code	
MMRE, no EVMS metrics	0.98
MMRE, with SPI	0.77
PRED (30), no EVMS metrics	0.27
PRED(30), with SPI	0.43
Peak Staff	
PRED(30), no EVMS metrics	0.50
PRED(30), with SPI	0.57

- CPI or SPI boosts accuracy
  - For Est. Hours
  - For Peak Staff
  - For transformed Est. Hours
  - For transformed New Code

CMMI Level 4

Estimated Hours		Peak Staff	
MMRE, no EVMS	0.28		
MMRE, with CPI	0.23		
PRED(30), no EVMS	0.71	PRED(30), no EVMS	0.59
PRED(30), with SPI	0.76	PRED(30), with SPI	0.76
PRED(30), with CPI	0.74	PRED(30), with CPI	0.74
Transformed Est. Hrs.		**Transformed New Code**	
MMRE, no EVMS metrics	0.36	MMRE, no EVMS	0.77
MMRE, with SPI	0.31		
MMRE, with CPI	0.30	MMRE, with CPI	0.69
		PRED(30), no EVMS metrics	0.26
		PRED(30), with CPI	0.38

Metric meets recommended accuracy level

Revised Effort Equation

$$E = c(A)^b(M)^d \text{ or}$$

$$E = c + bA + dM$$

c = Calibration Constant

A = Estimated New Code, Peak Staff, or Hours

b, d = Entropy Constant (where $b \neq 0$)

M = Earned Value Metric (SPI or CPI)

Summary

- Clouds evolving from existing and emerging technologies
- Cloud measurements need to be established, collected, stored and analyzed
- Software measurements need improvement
- EVMS by task would be more useful than by contract
- Ongoing cloud security issue resolution

Conclusion

- Need to identify measurements to collect and use for cloud
- Need to identify measurements to collect and use for software progress and quality
- Need to hypothesize, evaluate and report cost relationships for decision making

QUESTIONS?

Clusters, Grids and Clouds

Characteristics	Clusters	Grids	Clouds
Population	Commodity computers	High-end computers (servers, clusters)	Commodity computers and high-end servers and network attached storage
Size/scalability	100s	1000s	100s to 1000s
Node Operating System (OS)	One of the standard OS's (Linux, Windows)	Any standard OS (dominated by Unix)	A hypervisor (VM) on which multiple OS run
Ownership	Single	Multiple	Single
Interconnection network/speed	Dedicated, high-end with low latency and high bandwidth	Mostly Internet with high latency and low bandwidth	Dedicated, high-end with low latency and high bandwidth
Security/privacy	Traditional login/password-based, medium level of privacy - depends on user privileges	Public/private key based authentication and mapping a user to an account. Limited support for privacy.	Each user/application is provided with a VM. High security/privacy is guaranteed. Support for setting per-file access control list (ACL)
Discovery	Membership services	Centralized indexing and decentralized information services	Membership services
Service negotiation	Limited	Yes, SLA based	Yes, SLA based
User management	Centralized	Decentralized & virtual organization (VO)-based	Centralized (or can be delegated to a third party)
Resource management	Centralized	Distributed	Centralized/Distributed
Allocation/scheduling	Centralized	Decentralized	Both centralized and decentralized
Standards/interoperability	Virtual Interface Architecture (VIA)-based	Some Open Grid Forum standards	Web Services (SOAP and REST)
Single system image	Yes	No	Optional
Capacity	Stable and guaranteed	Varies, but high	Provisioned on demand
Failure management (self-healing)	Limited (often failed tasks/applications are restarted)	Limited (often failed tasks/applications are restarted)	Strong support for failover and content replication. VMs can be easily migrated from one node to another
Pricing of services	Limited, not open market	Dominated by public good or privately assigned	Utility pricing, discounted for large customers
Internet Working	Multi-clustering within an organization	Limtied adoption, but being explored through research efforts such as Gridbus InterGrid	High potential, third party solution providers can loosely tie services of different Clouds
Application drivers	Science, business, enterprise computing, data centers	Collaborative scientific and high throughput computing applications	Dynamically provisioned legacy and web applications with content delivery
Potential for building third party or value-added solutions	Limted due to rigid architecture	Limited due to strong orientation for scientific computing	High potential - can create new services by dynamically provisioning of compute, storage, and application services and offer as their own isolated or composite Cloud services to users

Cloud Service Model Metrics

Cloud Service Model	Performance to Measure
Software as a Service (SaaS)	Security of the application Access authority and processes Connectivity of the service Performance of the application
Platform as a Service (PaaS)	Integrity Stability of the platform
Infrastructure as a Service (IaaS)	Performance of server consolidation Security of the (VM) hypervisor environment Load balance of the running workloads

Workload Evaluation and Analysis on Virtual Systems, Chen, Liang & Yang, 2010, IEEE

Cloud Comparison

Property	Amazon Elastic compute cloud (EC2)	Google App engine	Microsoft Azure	Sun Network.com (Sun Grid)
Focus	Infrastructure (IaaS)	Platform (PaaS)	Platform (PaaS)	Infrastructure (IaaS)
Service type	Compute, storage (Amazon S3)	Web application	Web and non-Web applications	Compute
Virtualization	OS level running on a Xen hypervisor	Application container	OS level through fabric controller	Job management system (Sun Grid Engine)
User access interface	Tools	Administration console	Microsoft windows Azure portal	Job submission scripts, Sun Grid web portal
Web APIs	Yes	Yes	Yes	Yes
Value-added service providers	Yes	No	Yes	Yes
Programming framework	Customizable Linux-based Amazon Machine Image (AMI)	Python	Microsoft.NET	Solaris OS, Java, C, C++, FORTRAN

Cloud Performance Metrics

Workload	Metric
Mail server	Actions per minute
Java server	New orders per second
Standby server	None
Web server	Accesses per second
Database server	Commits per second
File server	Megabytes per second

Workload Benchmarks

Workload	Representative Benchmarks
Mail server	SPECmail2008
Application server	SPECjAppServer2004
Standby server	None
Web server	SPECweb2005
Database server	Sysbench using MYSQL
File server	Dbench

Cloud Security

- Three buzz-words with the acronym CIA
  - Confidentiality
    - Stringent access controls to prevent unauthorized access to the data
  - Integrity
    - A tested encryption schema to ensure that the shared storage environment safeguards all data
  - Availability
    - Scheduled data backup and safe storage of the backup media

Potential impacts with cloud security

Security Objective	Description [44 U.S.C., SEC. 3542]	Potential Impact		
		Low	**Moderate**	**High**
Confidentiality	*Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information.*	The unauthorized disclosure of information could be expected to have a **limited adverse effect** on organizational operations, organizational assets, or individuals.	The unauthorized disclosure of information could be expected to have a **serious adverse effect** on organizational operations, organizational assets, or individuals.	The unauthorized disclosure of information could be expected to have a **severe or catastrophic adverse effect** on organizational operations, organizational assets, or individuals.
Integrity	*Guarding against improper information modification or destruction, and includes ensuring information non-repudiation and authenticity.*	The unauthorized modification or destruction of information could be expected to have a **limited adverse effect** on organizational operations, organizational assets, or individuals.	The unauthorized modification or destruction of information could be expected to have a **serious adverse effect** on organizational operations, organizational assets, or individuals.	The unauthorized modification or destruction of information could be expected to have a **severe or catastrophic adverse effect** on organizational operations, organizational assets, or individuals.
Availability	*Ensuring timely and reliable access to and use of information.*	The disruption of access to or use of information or an information system could be expected to have a **limited adverse effect** on organizational operations, organizational assets, or individuals.	The disruption of access to or use of information or an information system could be expected to have a **serious adverse effect** on organizational operations, organizational assets, or individuals.	The disruption of access to or use of information or an information system could be expected to have a **severe or catastrophic adverse effect** on organizational operations, organizational assets, or individuals.