Cloud Survival Kit
How to estimate and measure “the Cloud”

PSM

August 2, 2012

David Seaver
OPS Consulting
davidseaver@opsconsulting.com

Some Perspective

- The DoD Cloud Computing Strategy introduces an approach to move the Department from the current state of a duplicative, cumbersome, and costly set of application silos to an end state which is an agile, secure, and cost effective service environment that can rapidly respond to changing mission needs.

Some Data

- Amazon touted its Web services in a new self-commissioned study, which found organizations receive huge return on their investments in the company's cloud computing service.
- The report -- conducted by IDC, an industry analyst -- comes on the heels of a study by CSC, a major government IT vendor, that found a majority of organizations that transitioned to the cloud saved little or no money. The CSC study itself came after federal officials claimed the transition would save the government $5 billion annually.
  - Analysts at IDC said 11 small, medium and large organizations at various stages in their cloud transition to Amazon Web Services spent an average of 70 percent less than it would have cost them to deploy the same resources on their premises or in a hosted environment.
  - IDC reported organizations with AWS saved even more money over time. Those using the service for three years saved $3.50 for every $1 invested, while those using it for five years saved $8.40 per $1 invested, the study found. The latter figure marks a 626 percent return on investment, according to the report.
  - While a majority of the savings come from reduced costs in infrastructure and services, part of that return is a result of increased productivity, the analysts found, as end users had fewer service disruptions and therefore saw 72 percent reduction in downtime.

Efficiency	
Cloud Benefits	**Current Environment**
• Improved asset utilization (server utilization > 60-70%) • Aggregated demand and accelerated system consolidation (e.g., Federal Data center Consolidation initiative) • Improved productivity in application development, application management, network, and end-user devices	• Low asset utilization (server utilization < 30% typical) • Fragmented demand and duplicative systems • Difficult to manage systems
Agility	
Cloud Benefits	**Current Environment**
• Purchase "as-a-Service" from trusted cloud providers • Near-instantaneous increases and reductions in capacity • More responsive to urgent agency needs	• Years required to build data centers for new services • Months required to increase capacity of existing services
Innovation	
Cloud Benefits	**Current Environment**
• Shift focus from asset ownership to service management • Tap into private sector innovation • Encourages entrepreneurial culture • Better linked to emerging technologies (e.g., devices)	• Burdened by asset management • De-coupled from private sector innovation engines • Risk-averse culture

PART 1

CLOUD COMPUTING DEFINED

Cloud Computing Defined

- The National Institute of Standards and Technology (NIST) defines cloud computing as:
  - *"A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."*
- The details of the NIST cloud computing definitions provide a simple and unambiguous taxonomy of three service models available to cloud consumers that are the core of cloud computing:
  - Software as a Service (SaaS)
  - Platform as a Service (PaaS)
  - Infrastructure as a Service (IaaS)

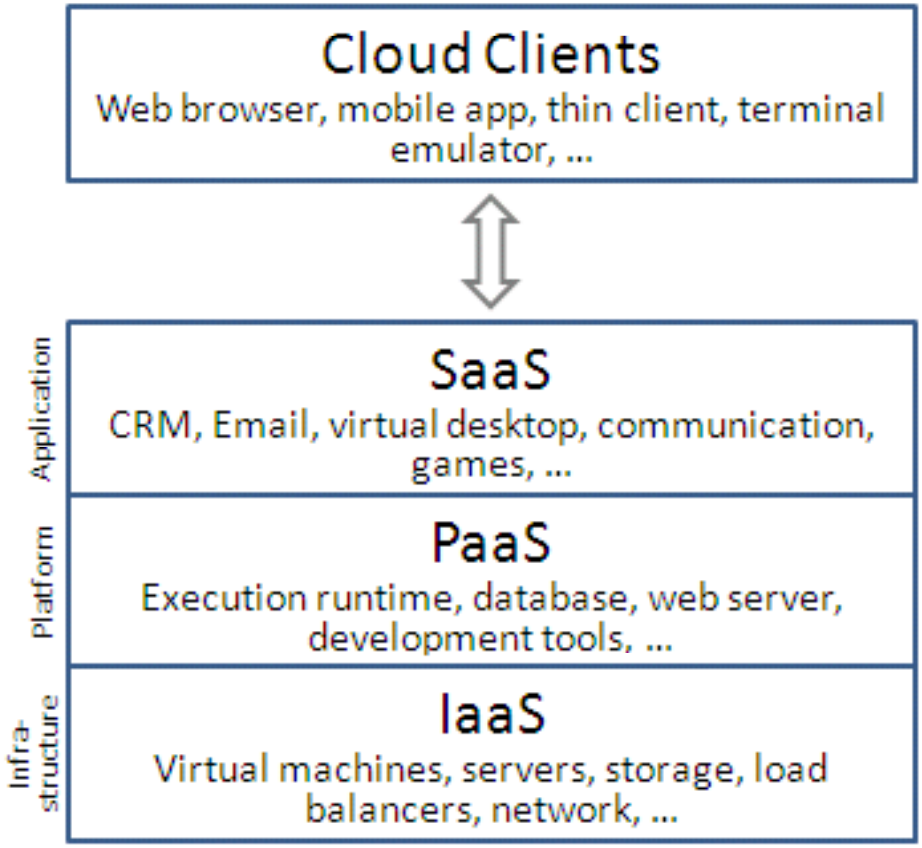
Cloud Computing Defined

- **Software as a service (SaaS)** sometimes referred to as "on-demand software, is a software delivery model in which software and associated data are centrally hosted on the cloud
  - Cloud providers install and operate applications software in the cloud and cloud users access the software from cloud clients.
  - The cloud users do not manage the cloud infrastructure and platform on which the application is running. This eliminates the need to install and run the application on the cloud user's own computers simplifying maintenance and support

- Platform as a service (PaaS) is a category of cloud computing services that provide a computing platform and a solution stack as a service
  - Cloud providers deliver a computing platform typically including operating system, programming language execution environment, database, and web server.
  - Application developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers.
  - With some PaaS offers, the underlying computer and storage resources scale automatically to match application demand such that cloud user does not have to allocate resources manually.

- Infrastructure as a service (IaaS)In this cloud service model, cloud providers offer computers, as physical or more often as virtual machines, and other resources.

Cloud Computing Defined



PART 2

WHAT MAKES THE CLOUD DIFFERENT

What's Different with Cloud?
Big Data

- In information technology, big data is a loosely-defined term used to describe data sets so large and complex that they become awkward to work with using on-hand database management tools.

- Difficulties include capture, storage, search, sharing, analysis, and visualization.

- The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, determine quality of research, prevent diseases, combat crime, and determine real-time roadway traffic conditions

What's Different with Cloud?
Map Reduce

- MapReduce is a programming model for processing large data sets, and the name of an implementation of the model by Google. MapReduce is typically used to do distributed computing on clusters of computers.
- The model is inspired by the map and reduce functions commonly used in functional programming although their purpose in the MapReduce framework is not the same as their original forms.
- MapReduce libraries have been written in many programming languages. A popular free implementation is Apache Hadoop.

What's Different with Cloud?
Map Reduce

- MapReduce is a programming model for processing large data sets, and the name of an implementation of the model by Google. MapReduce is typically used to do distributed computing on clusters of computers.
- The model is inspired by the map and reduce functions commonly used in functional programming although their purpose in the MapReduce framework is not the same as their original forms.
- MapReduce libraries have been written in many programming languages. A popular free implementation is Apache Hadoop.

What's Different with Cloud?
Apache Hadoop

- Apache Hadoop is an open source software framework that supports data-intensive distributed applications licensed under the Apache v2 license.[1] It enables applications to work with thousands of computational independent computers and petabytes of data. Hadoop was derived from Google's MapReduce and Google File System (GFS) papers.
- Hadoop is a top-level Apache project being built and used by a global community of contributors,[2] written in the Java programming language. Yahoo! has been the largest contributor[3] to the project, and uses Hadoop extensively across its businesses
- Apache Accumulo is a sorted, distributed key/value store based on Google's BigTable design. It is a system built on top of Apache Hadoop, Apache ZooKeeper, and Apache Thrift. Written in Java, Accumulo has cell-level access labels and a server-side programming mechanisms.

What's Different with Cloud?
Apache PIG

- Apache Pig is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets.
- At the present time, Pig's infrastructure layer consists of a compiler that produces sequences of Map-Reduce programs, for which large-scale parallel implementations already exist (e.g., the Hadoop subproject). Pig's language layer currently consists of a textual language called Pig Latin, which has the following key properties:
  - Ease of programming. It is trivial to achieve parallel execution of simple, "embarrassingly parallel" data analysis tasks. Complex tasks comprised of multiple interrelated data transformations are explicitly encoded as data flow sequences, making them easy to write, understand, and maintain.
  - Optimization opportunities. The way in which tasks are encoded permits the system to optimize their execution automatically, allowing the user to focus on semantics rather than efficiency.
  - Extensibility. Users can create their own functions to do special-purpose processing.

What's Different with Cloud?

Manage Data Center Configuration at Router level

- Traditional data centers are managed by IP address
- This limits the size of a data center
- New technology had moved management of the data center configuration up to the router level
- Bigger data centers possible….economies of scale which were not feasible before
- This combined with free and open source (FOSS) software, more effective virtualization technology is what distinguishes todays cloud from yesterdays data center

PART 3

CLOUD COST DRIVERS

Cloud Cost Drivers

Cost Drivers

- Public Cloud
- Private Cloud
- How many users
- How many ports/connections
- System Administrators
- Power Space and Cooling
- # Physical Servers
- Virtual Server to Physical Server ratio

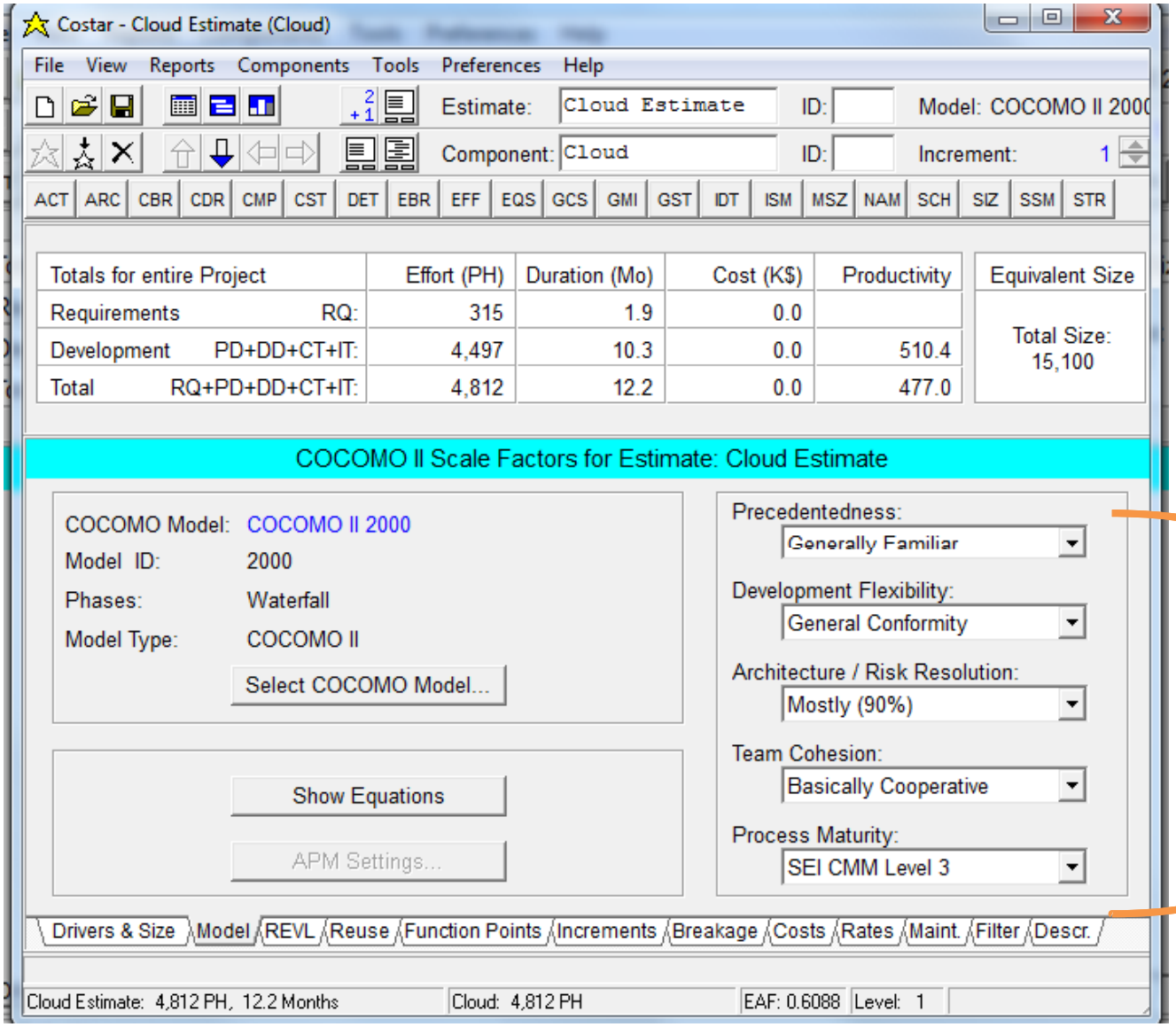
Cost Drivers

- Governance
- Free and Opens Source
- Hardwar Refresh
- Network Bandwidth
- Migration of applications to the cloud
- Re-engineering applications for parallel processing
- COTS License before and after
    - DBMS
    - Virtualization Software
    - Operating Systems
    - Applications

What Concerns Me

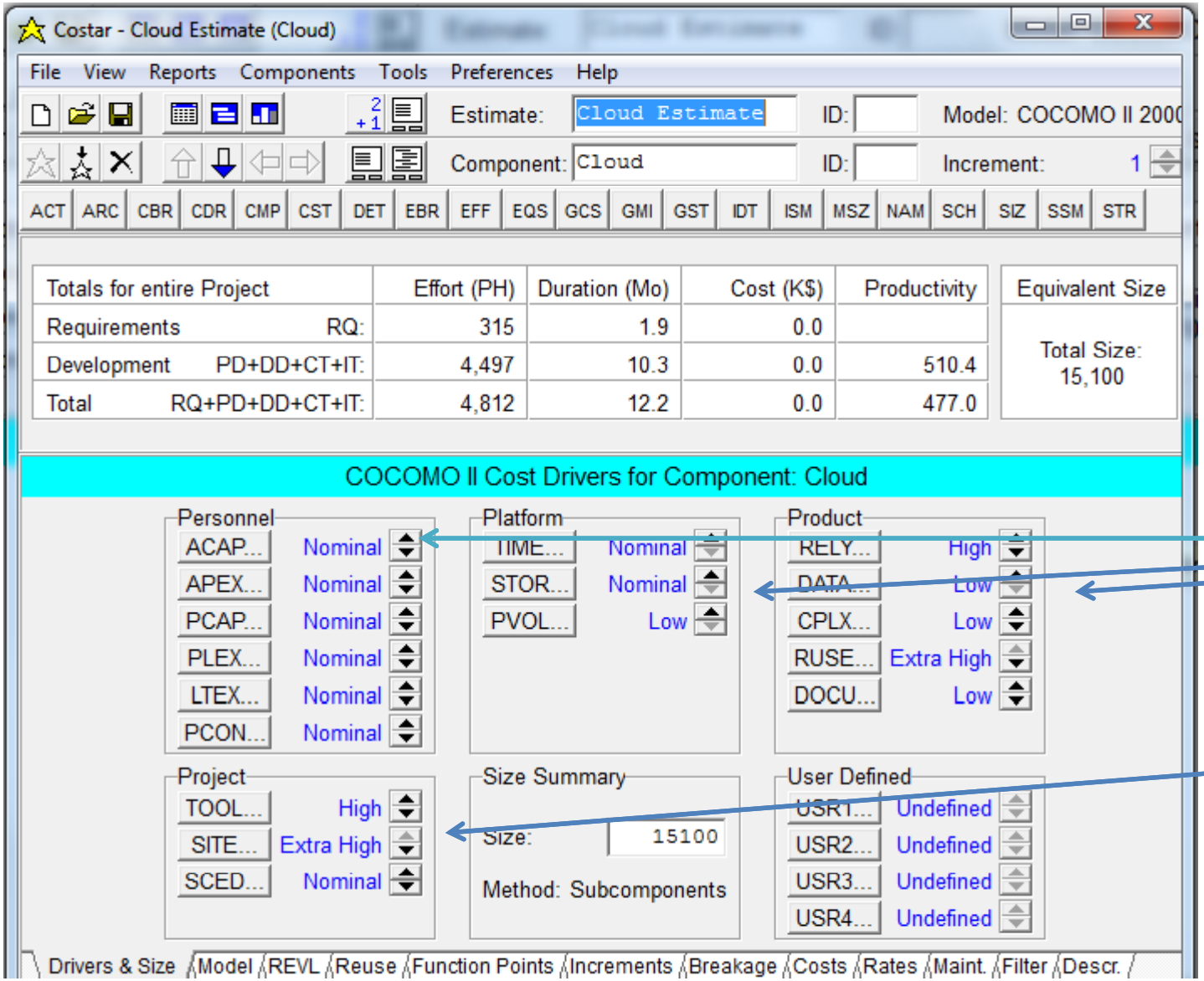
- Application migration: moving applications from standalone configuration to parallel processing
- Data transition
  - Meta data headers
    - One new input to process the header
    - One new data set to save the header
    - At least 15 function points
    - Or at least 755 SLOC (assume JAVA)
  - Tagging source data
    - One new input to process the data
    - One new data set to save the data
    - 15 function points/755 SLOC
- 10 data transitions would cost 4812 hours over 1 year
- Or $721.8K

Cloud via COCOMO



We need to talk about these

Cloud via COCOMO



We need to talk about these

What Else Concerns me

- Re-engineering the application
  - How many screens, reports, data retrievals have to be reworked
    - 5 function points per transaction
    - 265 Java SLOC per transactions
    - Assume 5 transactions per application
  - Data sources reworked
    - 10 function points
    - 300 Oracle SLOC
- 10 applications would cost 5042 hours over 1 year
- Or $756.3K

Conclusions

- Cloud is a combination of hardware and software
- The hardware part is getting easier and more efficient
- Free and open source (FOSS) is available for most of the cloud software if desired
- The transition is not free
  - Applications and data need to be migrated and re-engineered
- Current tools are workable but may need to be updated to be more efficient
- More to come ….