雲端運算簡介

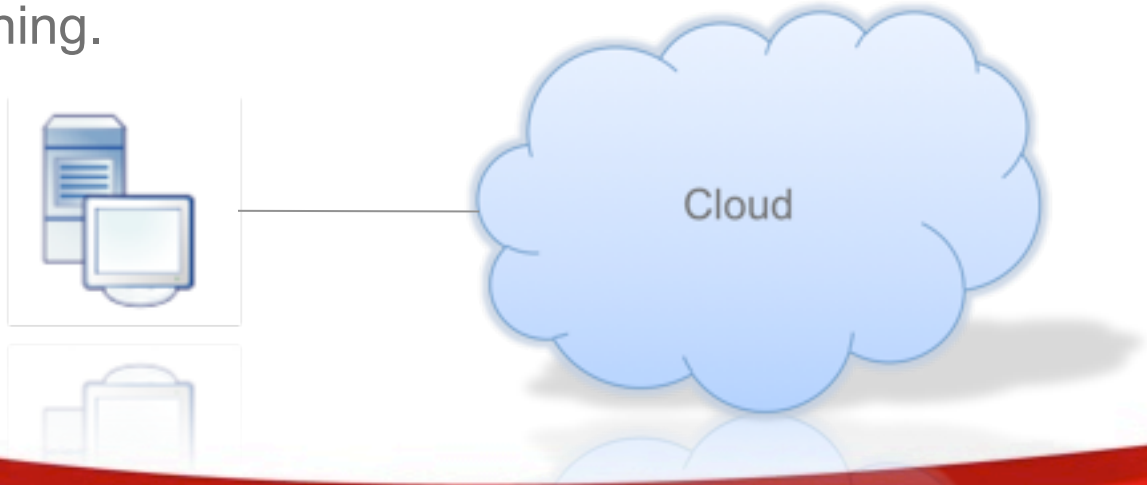
趨勢科技騰雲駕霧程式競賽

趨勢科技研發實驗室

Public
05/18/09

Why Cloud ?

- When drawing Computer Network Diagrams, we used to represent Internet as a “Cloud”.
- Connect a computer to Cloud by a line, meaning that the computer is connected to Internet, and accesses Internet services or content.
- Web Mail、Instant Messaging、Web Pages、Online virus scanning.



Copyright 2009 - Trend Micro Inc.

Super Computing
Cluster Computing
Grid Computing

Public
05/18/09

- The underlying concept of “Cloud Computing” dates back to 1960 and is attributed to computer scientist, “John McCarthy”.

- John McCarthy proposed the concept of “Computation being organized and delivered as a public utility”.

- The characteristics of the concept are similar to the “Service bureaus” in 1960s. (A service bureaus is a company which provides business services for a fee.)

Copyright 2009 - Trend Micro Inc.

Super Computing

- Supercomputers were introduced in the 1960s and primarily designed by Seymour Cray, called “the father of supercomputing”.
- Supercomputer is a massive computer with hundreds or thousands of CPUs, sharing common memory and I/O.
- IBM Roadrunner is currently the fastest supercomputer in the world, achieving 1.026 petaflops on May 25, 2008.

Copyright 2009 - Trend Micro Inc.

Cluster Computing

- Another approach to build a supercomputer, as adopted by Google.
- A cluster (group) of hundreds of thousands of COTS low-cost computers interconnected via fast LAN, configured so that they appear as a single machine.
- Achieve high availability and scalability.
- Typically cluster computing is much more cost-effective than a single supercomputer.

Copyright 2009 - Trend Micro Inc.

Cluster Computing – Programming Tools

- Message Passing Interface (MPI)
  - API for sending/receiving messages
  - Designed for high performance computing on both massively parallel machines and on workstation clusters.
  - Group communication, scalable file I/O, dynamic process management, synchronization.

- Parallel Virtual Machine (PVM)
  - Emulate a general purpose heterogeneous computing framework on interconnected computers.
  - Present a view of virtual processing elements
  - Create and manage processing tasks, basic message passing.

Copyright 2009 - Trend Micro Inc.

Cluster Computing – Google example

- Estimated 450,000 low-cost commodity servers in 2006.
- A Google's rack contains twenty 2U or forty 1U servers on each side.
- The servers on each side of a rack are interconnected via a 100-Mbps Ethernet switch.
- Each rack Ethernet switch has one or two Gigabit uplinks to a core Gigabit switch that connects all racks together.

Copyright 2009 - Trend Micro Inc.

Cluster Computing – Google example

Copyright 2009 - Trend Micro Inc.

Grid Computing

- Early 1990s, Ian Foster and Carl Kesselman came up with a new concept of “The Grid: Blueprint for a new computing infrastructure”.

- Expand the techniques of cluster computing where multiple independent computer clusters act like a grid due to their nature of not being located in a single administrative domain, as distributed and large-scale cluster computing, as well as a form of network-distributed parallel processing.

- What distinguishes grid computing from conventional cluster computing systems is that grids tend to be more loosely coupled, heterogeneous, and geographically dispersed.

Copyright 2009 - Trend Micro Inc.

Grid Computing

- Applied to computationally intensive scientific, mathematical, and academic problems as volunteer computing.

- Used in commercial enterprises for diverse applications and commercial transactions in support of E-commerce and Web services as utility computing.

Copyright 2009 - Trend Micro Inc.

Grid Computing

- SETI@home
  - A grid computing project using Internet-connected computers for scientific experiments in the Search for Extraterrestrial Intelligence, hosted by Space Sciences Laboratory, at the University of California, Berkeley.

- BOINC
  - Non-commercial middleware system for volunteer and grid computing.
  - For researchers to tap into the enormous processing power of personal computers around the world.
  - About 550,000 active hosts worldwide processing on average 1.5 petaflops as of March 2009.

Copyright 2009 - Trend Micro Inc.

Grid Computing

- Globus Toolkit
  - A fundamental enabling technology for the "Grid," letting people share computing power, databases, and other tools securely online across corporate, institutional, and geographic boundaries without sacrificing local autonomy.
  - Open Grid Services Architecture
    - GRIP (Grid Resource Information Protocol)
    - GRAM (Grid Resource Access and Management)
    - GridFTP

Copyright 2009 - Trend Micro Inc.

Distributed Computing

Public
05/18/09

TREND MICRO

Distributed Computing

- Multiple CPUs across multiple computers over a network, working together to solve distributed computing problems.
- A process is split up into parts that run simultaneously on multiple computers communicating over a network.
- A form of parallel programming. Parallel programming usually means vector processing of data or multi-tasked programming.

Copyright 2009 - Trend Micro Inc.

Distributed Computing

- Distributed Problem Domains:
  - Large amounts of data
  - Lots of Computation
  - Parallelizable applications
  - Many computers

- Indexing the World Wide Web (Google)

- Analyzing the Internet threats (Trend Micro)

Copyright 2009 - Trend Micro Inc.

Distributed Computing

- Distributed System
  - An application that executes a collection of protocols to coordinate the actions of multiple processes on a network, such that all components cooperate together to perform a single or small set of related tasks.
  - Fault-Tolerance
  - High Availability
  - Recoverability
  - Consistency
  - Scalability
  - Performance Predictability
  - Security

Copyright 2009 - Trend Micro Inc.

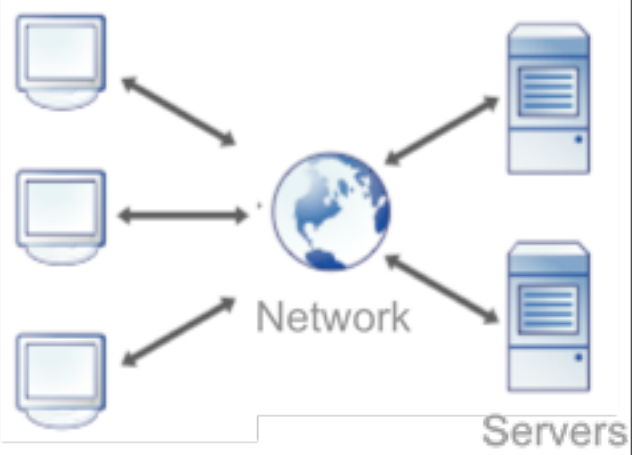
Distributed Computing

- Expectation and handling of failure on both hardware and software.
- 8 Fallacies:
  - 1. The network is reliable.
  - 2. Latency is Zero.
  - 3. Bandwidth is infinite.
  - 4. The network is secure.
  - 5. Topology does not change.
  - 6. There is one administrator.
  - 7. Transport cost is Zero.
  - 8. The network is homogeneous.

Copyright 2009 - Trend Micro Inc.

Distributed Computing

- One of the basic architectures in distributed computing: Client-Server
- Clients request services from servers over a computer network.
- Services are provided by a group of servers of a particular type.
- DNS services, Web services.



Copyright 2009 - Trend Micro Inc.

Distributed Computing

- Actual service server processing client requests may base on locality or load balance.

- Locality
  - Local machine -> Local server -> Remote server

- Load balance
  - Spread request works among service servers.
  - Optimal resource utilization, maximize performance, minimize response time.

Copyright 2009 - Trend Micro Inc.

Distributed Computing

- Data replication
  - Multiple copies of service data at multiple servers
  - Permit access at multiple locations
  - Increase availability
  - Avoid SPOF (Single Point of Failure)

- Caching
  - Local data copy for quick access
  - Validate cached data before use, commonly TTL.
  - Browser and web proxy cache frequently accessed web pages.

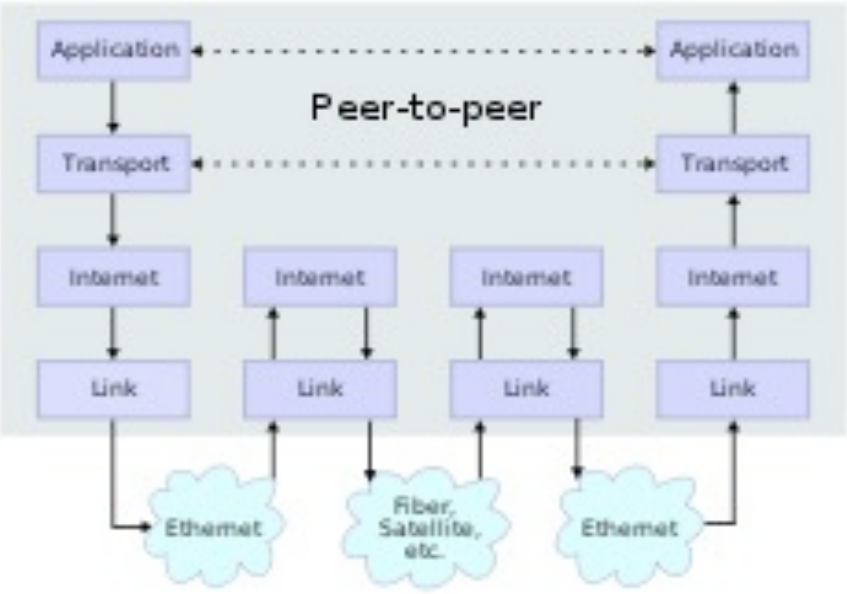
Copyright 2009 - Trend Micro Inc.

Distributed Computing

- Reliable communication via TCP/IP protocols
- Internet Protocol (IP)
  - Protocol for communication data across a packet-switched internetwork.
  - Addressing methods: IPv4 and IPv6.
- Transmission Control Protocol (TCP)
  - A reliable stream delivery service that guarantees in-order delivery of data stream between sender and receiver.
  - Socket programming: socket(), bind(), connect(), accept(), send(), recv(), ...

Copyright 2009 - Trend Micro Inc.

Distributed Computing

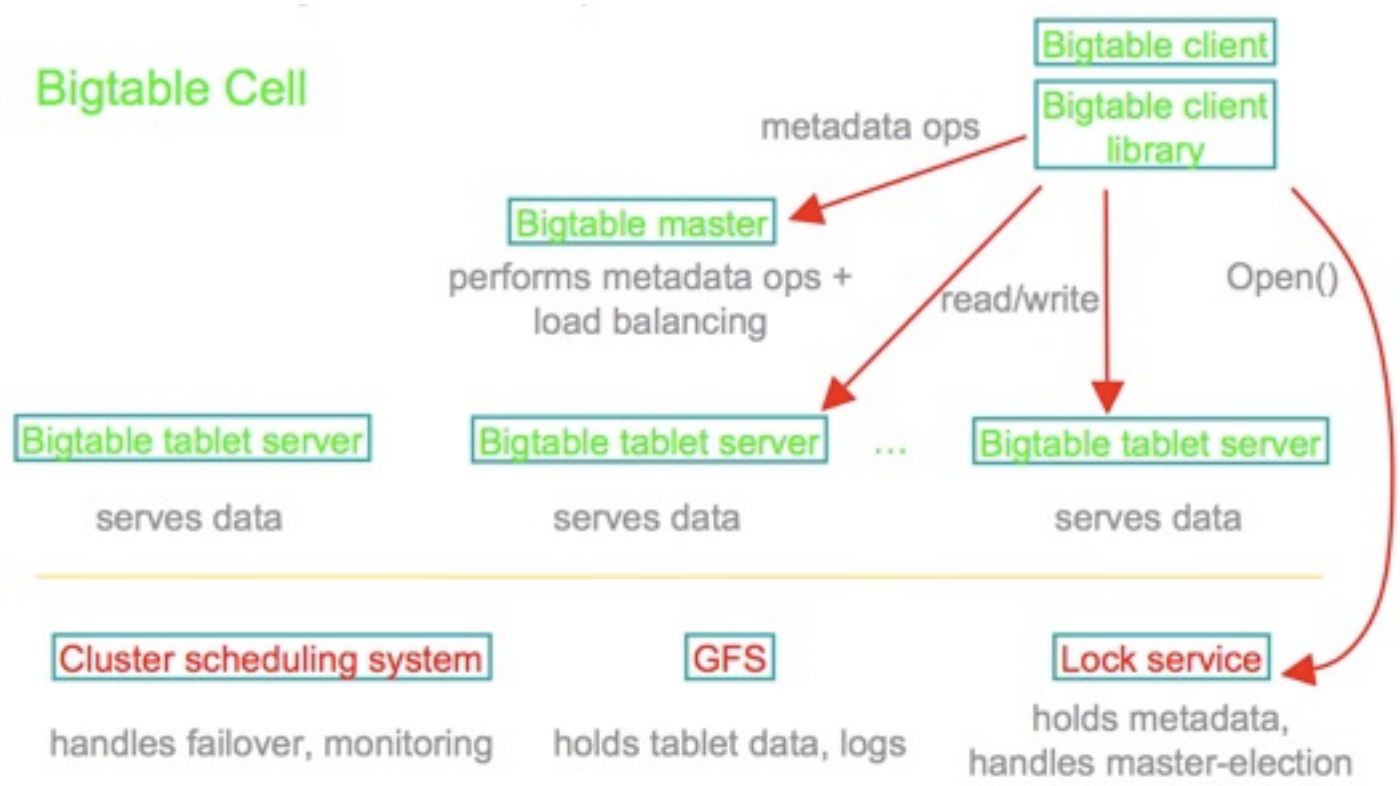


- Application layer
  - HTTP, FTP, RPC, SSH
- Transport layer
  - TCP, UDP
- Internet layer
  - IP, ICMP
- Link layer
  - Physical transmission of data
  - Ethernet, MAC

Copyright 2009 - Trend Micro Inc.

Distributed Computing

- Google Big Table
  - Large scale and distributed structure storage system



Copyright 2009 - Trend Micro Inc.

Distributed Computing

- Hadoop Distributed File System (HDFS)
  - Distributed file system designed to run on COTS hardware.
  - Detection of failures and automatic recovery
  - High throughput of streaming data access
  - Batch processing, rather interactive
  - Large data sets, favor large files
  - Write-once-read-many coherency model
  - Moving computation is cheaper than moving data
  - Portability (Java)

Copyright 2009 - Trend Micro Inc.

MapReduce

Public
05/18/09

MapReduce

- A simple programming model that applies to many large-scale computing problems and processes large data sets.
- Automatic parallelization and distribution of large-scale computations on a large cluster of commodity machines.
- It allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system.

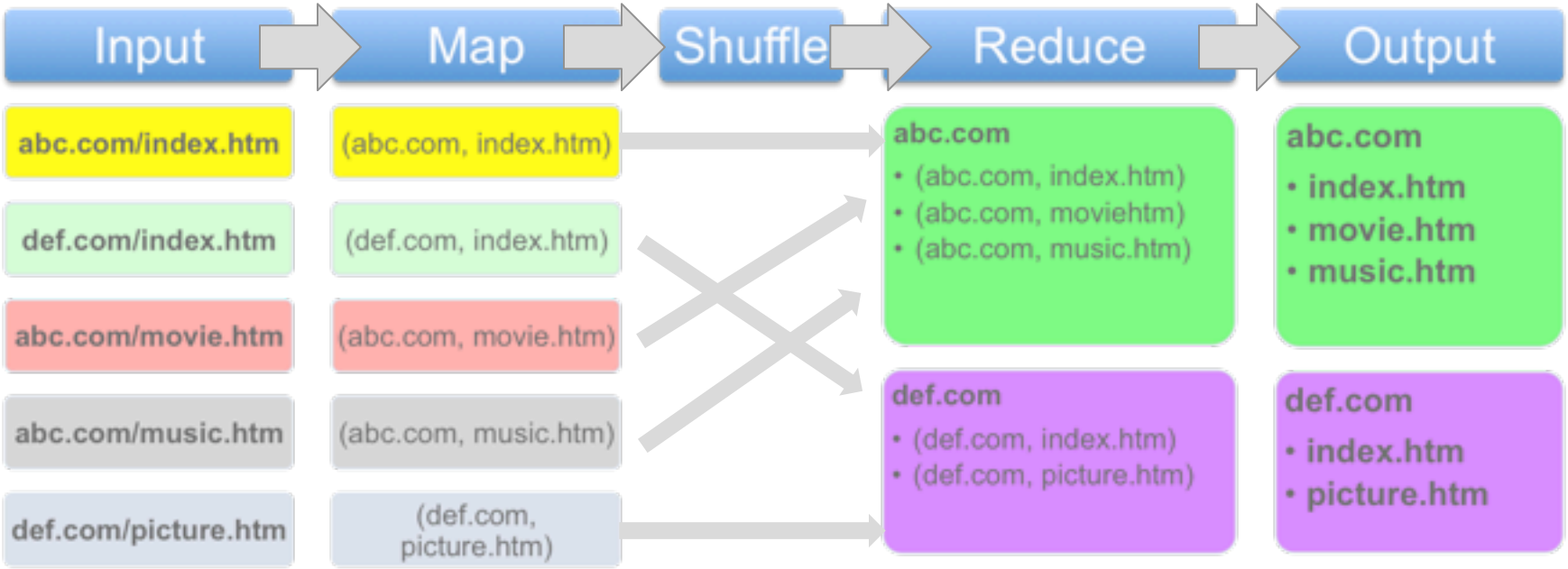
Copyright 2009 - Trend Micro Inc.

MapReduce

- Automatic parallelization and distribution
- Fault-tolerance via re-execution
- I/O scheduling
- Status and Monitoring

Copyright 2009 - Trend Micro Inc.

MapReduce



Copyright 2009 - Trend Micro Inc.

Concept & Service Models

Public
05/18/09

TREND MICRO

Web 2.0

- The Web as a Platform
  - Communication
  - Information sharing
  - Collaboration
  - Interaction
  - Content

- Delivery of web sites to users as services.

- Allow users to run software applications via browsers.

Copyright 2009 - Trend Micro Inc.

Web 2.0 網站

Copyright 2009 - Trend Micro Inc.

Software as a Service

- A model of software deployment whereby a provider licenses an application to customers for use as a service on demand.
- Globally and remotely access and manage commercial software via Web.
- No client-side software/hardware installation, maintenance, update. Reduce total cost of ownership.
- Integration of interconnected software services, Mashups.

Copyright 2009 - Trend Micro Inc.

Software-as-a-Service Providers

- Trend Micro Hosted Security
  - http://us.trendmicro.com/us/solutions/software-as-a-service/index.html

- Google Apps
  - http://www.google.com/a/help/intl/zh-TW/index.html

- Salesforce.com Customer Relationship Management(CRM)
  - http://www.salesforce.com/tw/

Copyright 2009 - Trend Micro Inc.

Platform as a Service

- By Salesforce.com. Delivery of a platform as a service for building and hosting web applications.
- Cloud-based web application development, testing, deployment, hosting lifecycle.
- http://www.youtube.com/watch?v=EbnCUqAL6UM

Copyright 2009 - Trend Micro Inc.

Platform-as-a-Service Providers

- Force.com (from Salesforce.com)
  - http://www.salesforce.com/platform/
- Google App Engine
  - http://code.google.com/appengine/
- Microsoft Azure Services Platform
  - http://www.microsoft.com/azure/default.mspx
- Amazon Web Services
  - http://aws.amazon.com/

Copyright 2009 - Trend Micro Inc.

Infrastructure as a Service

- (Originally Hardware as a Service, HaaS) Delivery of computing infrastructure as a service.
- Platform / hardware virtualization environment.
- Servers, Network equipment, RAM, Disk, CPU, etc.
- Dynamic resource allocation based on the needs of your applications.
- Only pay for what you use.

Copyright 2009 - Trend Micro Inc.

Infrastructure-as-a-Service Providers

- Amazon EC2 (Elastic Compute Cloud)
  "a web service that provides resizable compute capacity in the cloud. ..."
  - http://aws.amazon.com/ec2/

United States | Europe

Standard On-Demand Instances	Linux/UNIX Usage	Windows Usage
Small (Default)	$0.10 per hour	$0.125 per hour
Large	$0.40 per hour	$0.50 per hour
Extra Large	$0.80 per hour	$1.00 per hour
High CPU On-Demand Instances	**Linux/UNIX Usage**	**Windows Usage**
Medium	$0.20 per hour	$0.30 per hour
Extra Large	$0.80 per hour	$1.20 per hour

- Small Instance (Default) 1.7 GB of memory, 1 EC2 Compute Unit (1 virtual core with 1 EC2 Compute Unit), 160 GB of instance storage, 32-bit platform

Copyright 2009 - Trend Micro Inc.

Cloud Computing

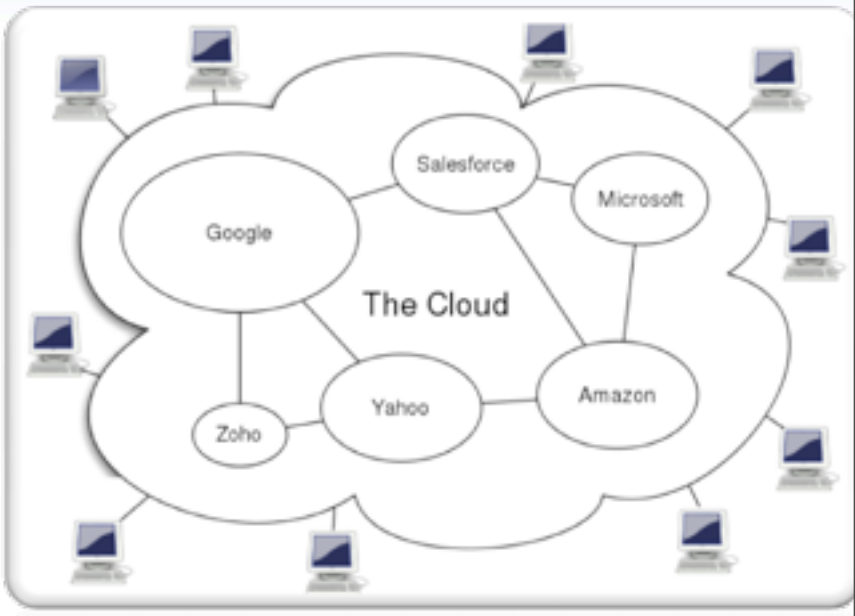
Public
05/18/09

TREND MICRO

Cloud Computing

- Integrate the concepts of IaaS、PaaS、SaaS、Web 2.0、and related technologies (ex, MapReduce、Ajax、Virtualization) .

- Based on the Internet Cloud, to satisfy the computing needs of the users on demand.



Copyright 2009 - Trend Micro Inc.

Cloud Computing

- A distributed computing model
- In the Internet Cloud, use dynamic, scalable, and virtualized computing resources to provide web services
- For complex computation on large-scale data, use dynamic and virtualized computer clusters to execute distributed and parallel computation in the Cloud.

Copyright 2009 - Trend Micro Inc.

Cloud Architecture

Public
05/18/09

TREND MICRO

Cloud Architecture

- The architecture of the software systems involved in the delivery of clouding computing, [...]. It typically involves multiple cloud components communicating with each other over application programming interfaces, usually web services.

Copyright 2009 - Trend Micro Inc.

Cloud Architecture

- Designs of software applications that use Internet-accessible on-demand services.

- Applications built on Cloud Architectures are such that the underlying computing infrastructure is used only when it is needed (for example to process a user request), draw the necessary resources on-demand (like compute servers or storage), perform a specific job, then relinquish the unneeded resources.

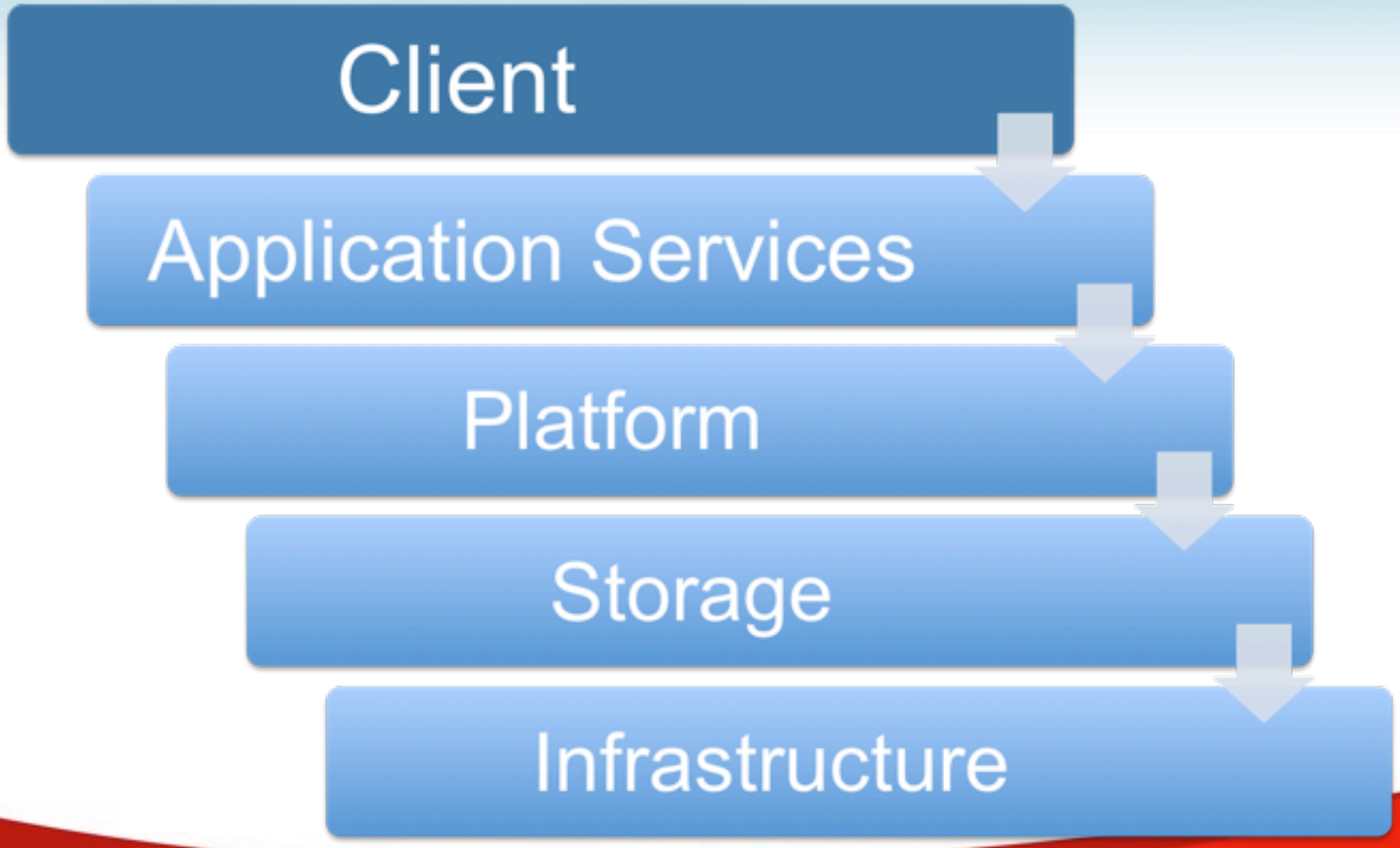
Copyright 2009 - Trend Micro Inc.

Cloud Architecture

- Address key difficulties surrounding large-scale data processing.
  - difficult to get as many machines as an application needs.
  - difficult to get the machines when one needs them.
  - difficult to distribute and co-ordinate a large-scale job on different machines, run processes on them, and provision another machine to recover if one machine fails.
  - difficult to auto-scale up and down based on dynamic workloads.
  - difficult to get rid of all those machines when the job is done.

Copyright 2009 - Trend Micro Inc.

Cloud Architecture



Copyright 2009 - Trend Micro Inc.

Cloud Architecture - Client

- Cloud client that relies on application services on cloud computing architecture.

- Browsers
  - Microsoft Internet Explorer
  - Mozilla Firefox
  - Google Chrome

- Mobile devices
  - Android
  - iPhone
  - Windows Mobile

Copyright 2009 - Trend Micro Inc.

Cloud Architecture – Platform

- Cloud platform for developing and hosting the application services.

- Google AppEngine
  - Let you run your web applications on Google infrastructure
  - Java and Python runtime environments
  - Full support for common web technologies and services
  - Distributed Datastore (BigTable)

- Hadoop MapReduce
  - Open source distributed computing platform for running MapReduce applications
  - Batch processing

Copyright 2009 - Trend Micro Inc.

Cloud Computing - Storage

- Distributed persistent data storage in the cloud.
- Unstructured data storage
  - Traditional file system
    - HDFS
  - Key-Object pair
    - Amazon S3 (Simple Storage Service)
- Structured data storage
  - Amazon SimpleDB
  - Google AppEngine DataStore

Copyright 2009 - Trend Micro Inc.

Cloud Architecture - Infrastructure

- Infrastructure as a Service
  - Pay for what you use
  - On demand
- Virtualized computing infrastructure
  - Computer Clusters
  - Hardware abstraction
  - Virtualization
    - Xen, VMware
- User configurable computing environment
  - Operation system, networking
  - Memory, disk, CPU

Copyright 2009 - Trend Micro Inc.

Smart Protection Network

Public
05/18/09

TREND MICRO

趨勢科技智慧防護網絡(Smart Protection Network)

- http://itw.trendmicro.com/smart-protection-network/index.php

- 近年，在個人電腦及網絡上耗費記憶體的 Virus Patterns 數目正不斷上升。在2005 年，一天約有50 個新 Virus Patterns 增加至 Pattern database。到2008 年，數目已經增加到5,000 個。趨勢科技把大多數的 Virus Patterns 儲存到雲端資料庫，目的是減低儲存在端點的數目，並減少佔用在網絡或端點的頻寬，而且更提供即時的網絡保障。

- 超越傳統的阻擋威脅方法，將安全保護範圍延伸至 Internet ，並採用趨勢最新的雲端運算技術將 Internet 上的威脅阻擋在網絡或電腦之外。

Copyright 2009 - Trend Micro Inc.

趨勢科技智慧防護網絡(Smart Protection Network)

- Smart Protection Network 利用 Reputation Technologies，連結來自網絡、檔案、及電子郵件的威脅，並持續更新雲端的威脅資料庫，利用 Cloud-oriented Architecture 以偵測與分析最新的網絡威脅。

- Smart Protection Network 結合雲端技術，確保用戶無論在何時何地，或是以任何方式上網，均能隨時獲得最新最強的安全保護。

Copyright 2009 - Trend Micro Inc.

趨勢科技智慧防護網絡(Smart Protection Network)

- 網站信譽服務 (Web Reputation Service)
  - 趨勢科技的網站信譽資料庫屬全世界最大之一，其網站信譽技術能夠追蹤網域的可靠程度，方法是以網站的歷史、網址的轉變、以及惡意軟件行為分析所發現的可疑活動等因素，給予一個信譽評分，並阻止使用者進入風險較高的網站。

Copyright 2009 - Trend Micro Inc.

趨勢科技智慧防護網絡(Smart Protection Network)

- 電子郵件信譽服務(Email Reputation Service)
  - 電子郵件信譽技術藉著查核已知垃圾郵件來源的信譽資料庫，以及使用時評估發信者信譽的動態服務來驗證 IP 地址。信譽評分是根據 IP 地址的行為、活動規模、與先前紀錄而制定。發信者的 IP 地址一經查明，惡意電子郵件在抵達網絡之前就可被阻擋，並有效防止如：zombies 或 botnet 的威脅進入網絡或是用戶的個人電腦。

Copyright 2009 - Trend Micro Inc.

趨勢科技智慧防護網絡(Smart Protection Network)

- 檔案信譽服務 (File Reputation Service)
  - 結合網站與電子郵件信譽技術，雲端檔案信譽技術能夠檢查每份流經端點、伺服器、或閘道的檔案，並與防毒模式系統作比較。高效能內容傳輸網絡與區域快取伺服器能確保在檢查過程中的存取時間減到最低。
  - 由於有關惡意軟件的資料已儲存在雲端網絡，所以用戶可以隨時獲得相關資訊。與下載防毒模式檔案的傳統方法比較，檔案信譽技術不但可以節省用戶時間，亦可以大幅減少下載檔案的端點儲存空間。

Copyright 2009 - Trend Micro Inc.

趨勢科技智慧防護網絡(Smart Protection Network)

- 行為分析的相關性技術（Correlation technology with behavior analysis）
  - 以行為分析來檢查一系列網絡行為的相互關係，以判斷其惡意程度。獨立的網絡行為可能看似無害，但是將若干行為串連在一起時，則可能發現當中的惡意企圖。透過不斷更新威脅資料庫和比對不同的惡意行為，趨勢科技可以即時提供自動防禦功能。

Copyright 2009 - Trend Micro Inc.

趨勢科技智慧防護網絡(Smart Protection Network)

- 即時回應及更新功能
  - 可以保持趨勢科技的產品技術和公司24x7 的網絡威脅研究中心之間的緊密聯繫，提供雙向的溝通橋樑。例如，每當在用戶的電腦中偵測到新威脅，便會自動更新公司遍佈全世界的威脅資料庫，避免有更多用戶遇到同一威脅問題。趨勢科技不斷收集及處理由honeypot、客戶、及業務夥伴所組成的環球網絡威脅資料，對最新的網絡襲擊提供自動防禦功能，並實踐「團結就是力量」的保安理念，與其他用戶共同發揮網絡保安的強大功能。

Copyright 2009 - Trend Micro Inc.

Reference

- [1] http://en.wikipedia.org/wiki/Cloud_computing
- [2] http://aws.amazon.com/
- [3] http://www.davidchappell.com/CloudPlatforms--Chappell.pdf
- [4] http://code.google.com/edu/parallel/dsd-tutorial.html
- [5] http://code.google.com/edu/submissions/rutgers/index.html
- [6] http://en.wikipedia.org/wiki/Internet_protocol_suite
- [7] http://hadoop.apache.org/core/
- [8] http://labs.google.com/papers/mapreduce.html

Copyright 2009 - Trend Micro Inc.

http://www.trend.org/fd/tabid/66/Default.aspx

Copyright 2009 - Trend Micro Inc.