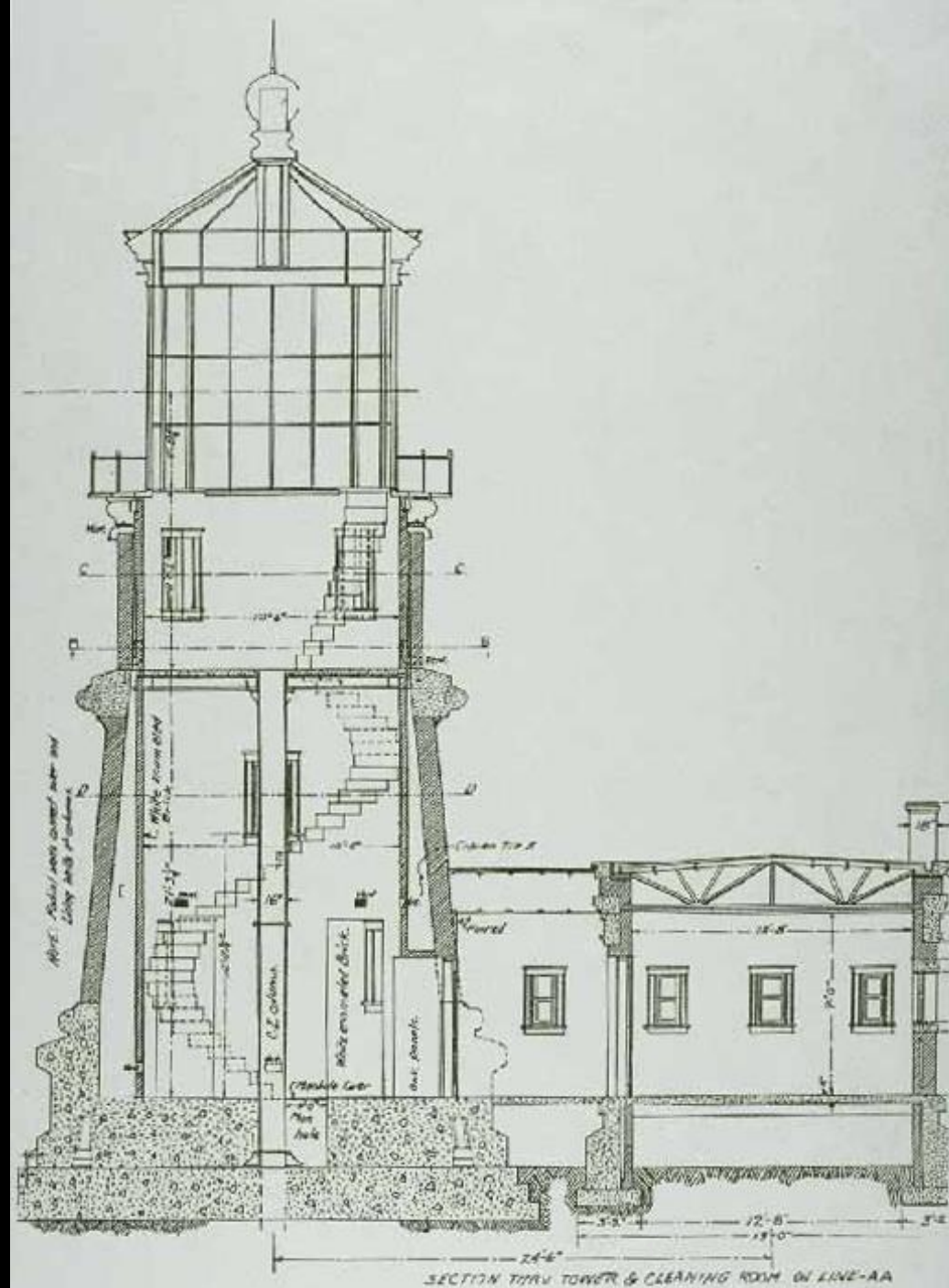


# Software Architecture & Design





# The Third Platform

Architecting for the Future – Cloud & Big Data

# Contents

- Introduction
- The Third Platform
  - Cloud
  - Mobile
  - Social
  - Big Data
- Discussion & Q&A

# The Third Platform

The Evolution of modern computing

# THE WALL STREET JOURNAL.

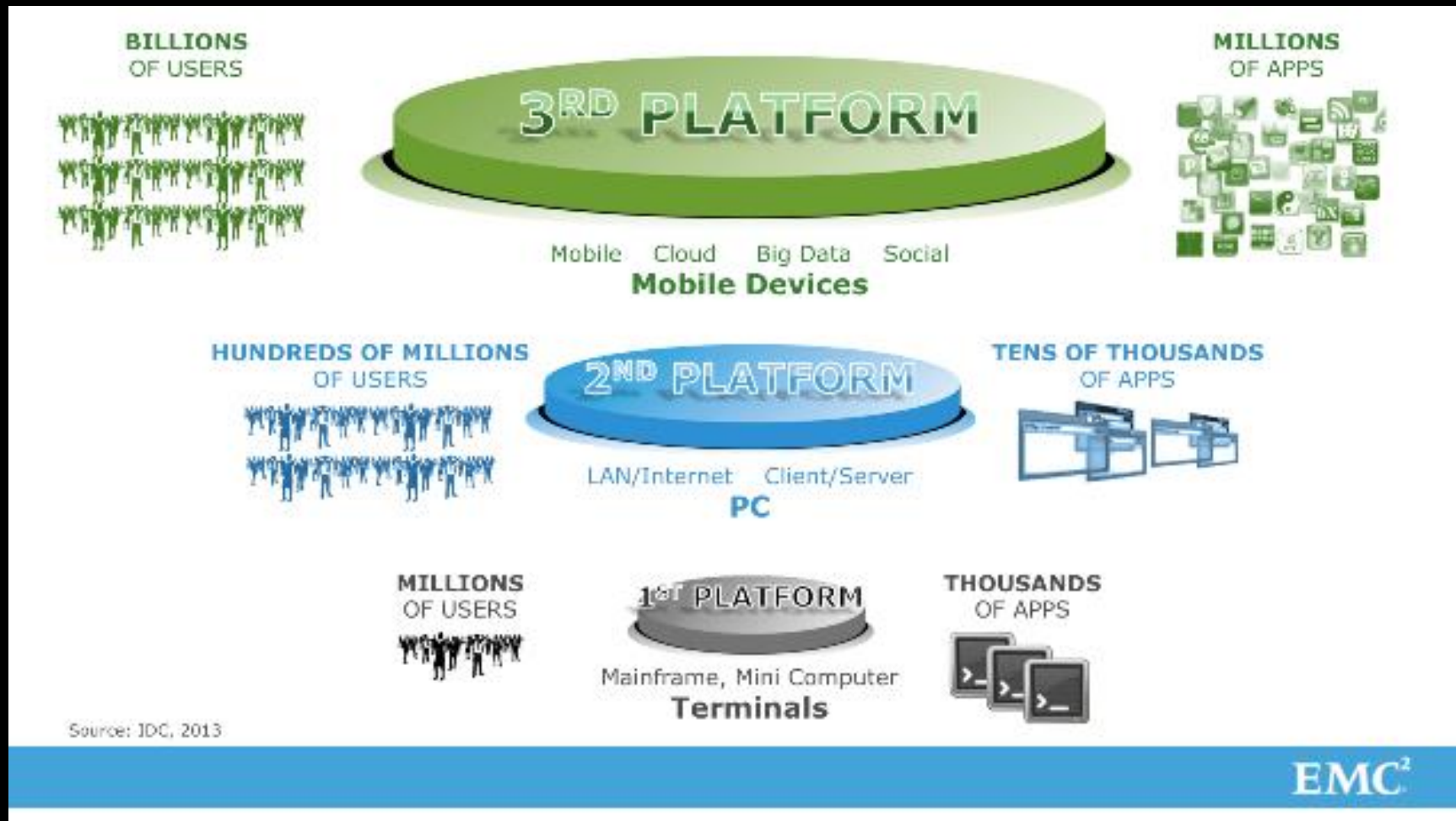
## Why Software Is Eating The World

*By Marc Andreessen*

August 20, 2011

Every business is becoming a software business.

# Characteristics Contrasted



# Architecture Implications

## 1<sup>st</sup> Platform The Mainframe

- Highly Centralized
- Administered Centrally
- Batch Processing Oriented Systems
- Few languages & databases supported
- Long development cycles

## 2<sup>nd</sup> Platform PC & Client/Server

- Highly Distributed
- Locally administered
- Online, Real-time Applications
- More languages & databases
- Shorter development cycles

## 3<sup>rd</sup> Platform Cloud & Mobile

- Both centralized (cloud) and distributed (mobile)
- Centrally administered
- Online, Realtime Apps
- Large choice of development platforms
- Very fast development cycles

# Four Aspects of 3<sup>rd</sup> Platform

- Cloud Computing
- Mobile
- Social
- Big Data



# Cloud Computing

The Virtualized Datacenter

What is the cloud?

What is the cloud?

A *collection of virtualized machines* delivering compute, storage and network facilities, independent of location, used to store, process or manage data.

# Cloud Computing

Definition



What is  
virtualization?

An abstraction layer that allows a single physical machine to simultaneously host multiple instances of possibly heterogeneous operating systems.

# Virtualization

Definition

# Virtualization - Drivers

- Hardware Improvements
  - Processing power increased to make virtualization feasible
- Independence from hardware
- Greater utilization of hardware

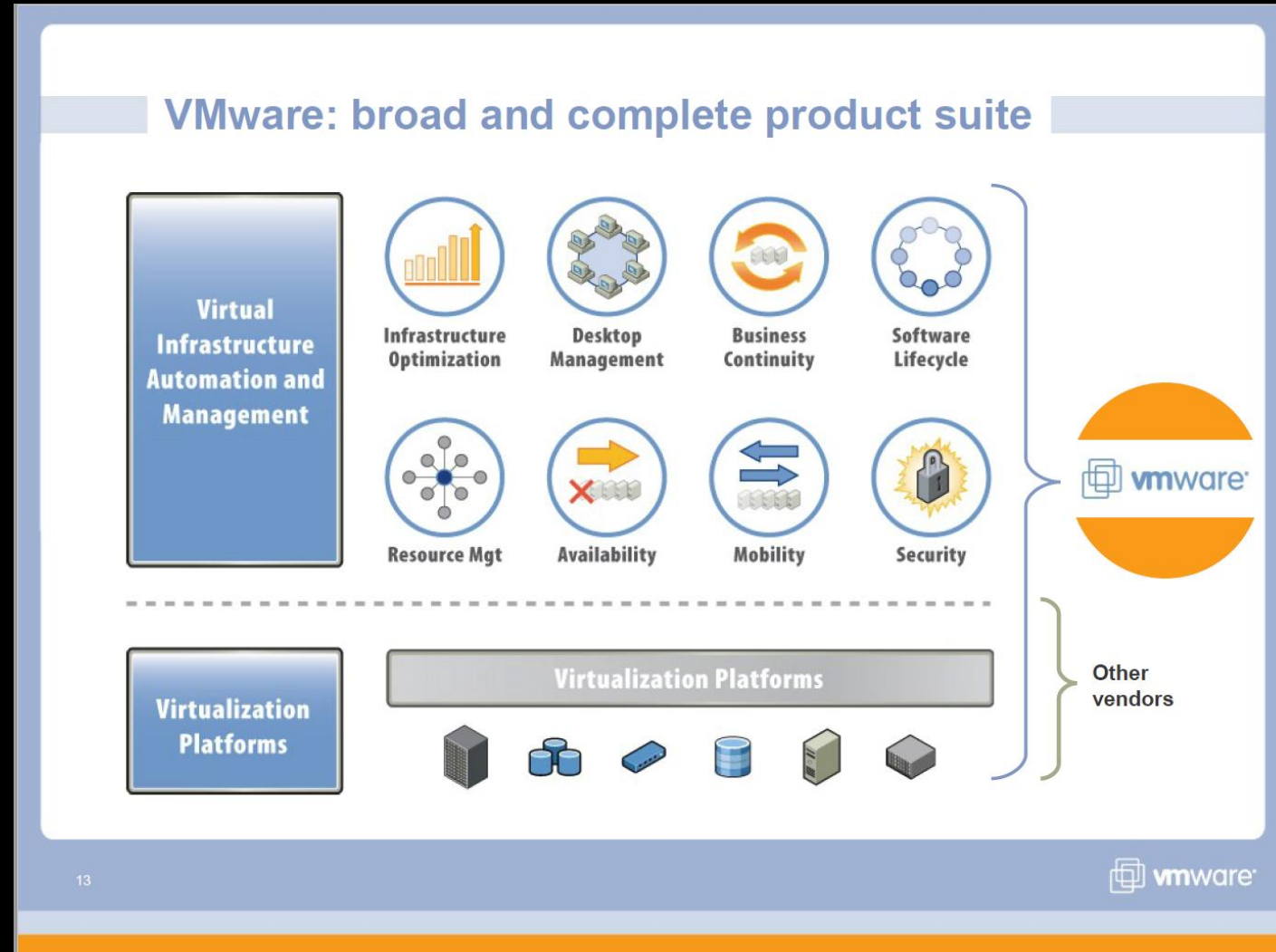


# Virtualization - Implications

- Logically tight-coupling of application + operating system as a single unit
- Inherent Isolation – Fault & Security
- Freedom of Choice – able to choose the OS to best suit the application
- Location independence – able to provision a VM anywhere

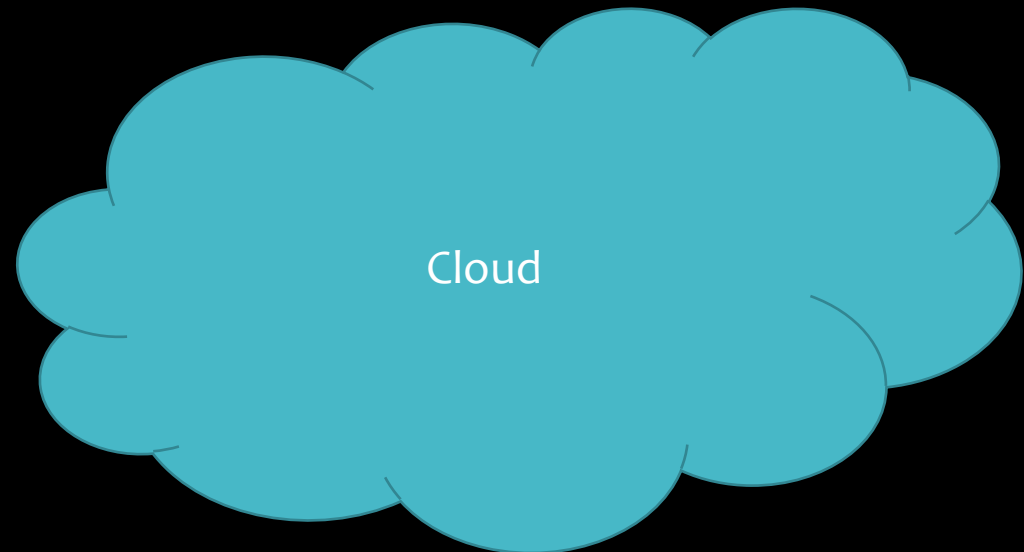
# Virtual Infrastructure Automation & Management

- IT as a Service
- Enterprise Scale Management
- Infrastructure Optimization
- Resource Management & Optimization
- DR



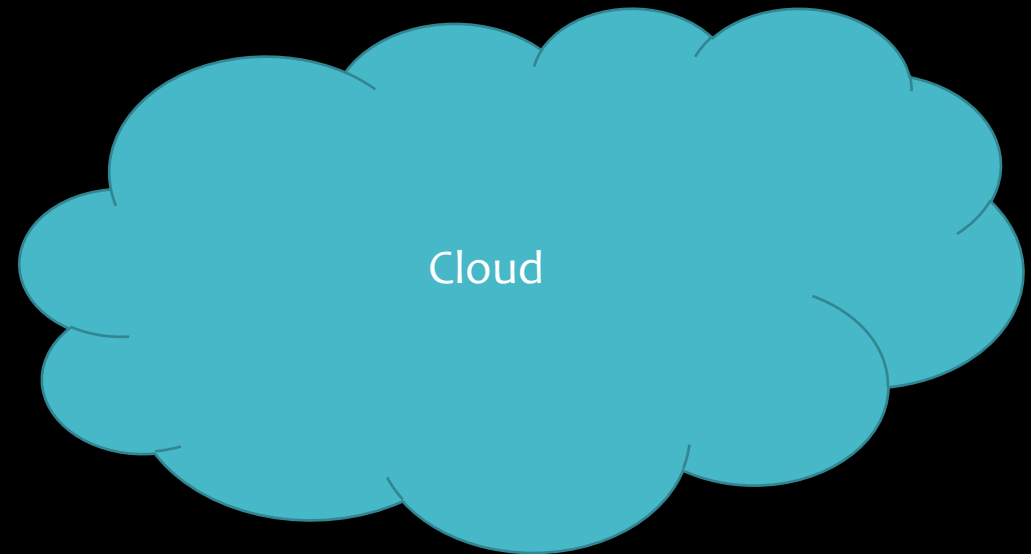
# Virtualization - Benefits

- Server Consolidation
- Rapid Provisioning
- Migration Flexibility
  - Scale up - Performance
  - Scale down - Reduced Power Consumption
  - Maintenance
  - Always up, always available
- Resource Pooling - A single pool of:
  - Compute
  - Network
  - Storage



# The case for Cloud

- Economies of Scale
- Affordability
- Extensibility
- Agility



# Pro's & Con's

## Advantages

- 'Infinite' Resources
  - Dynamically increase resources
- Ease of Administration
  - Ease & Speed of Deployment & Upgrade
  - Backup & Recovery

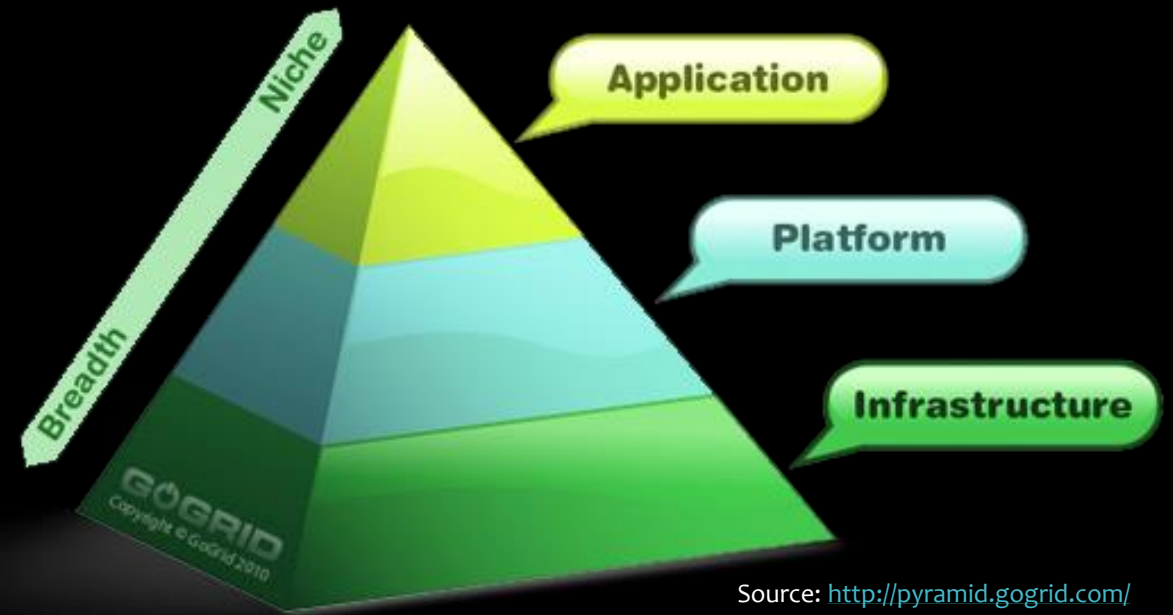
## Disadvantages

- Security Concerns → Trust
- Latency Concerns
  - Bandwidth/Responsiveness
- Outages
- Reliance on DevOps

# Types of Cloud - Abstraction

## Abstraction

- SaaS – Software as a Service
  - e.g. SAP, Workday
- PaaS – Platform as a Service
  - e.g. CloudFoundry
- IaaS – Infrastructure as a Service
  - e.g. AWS



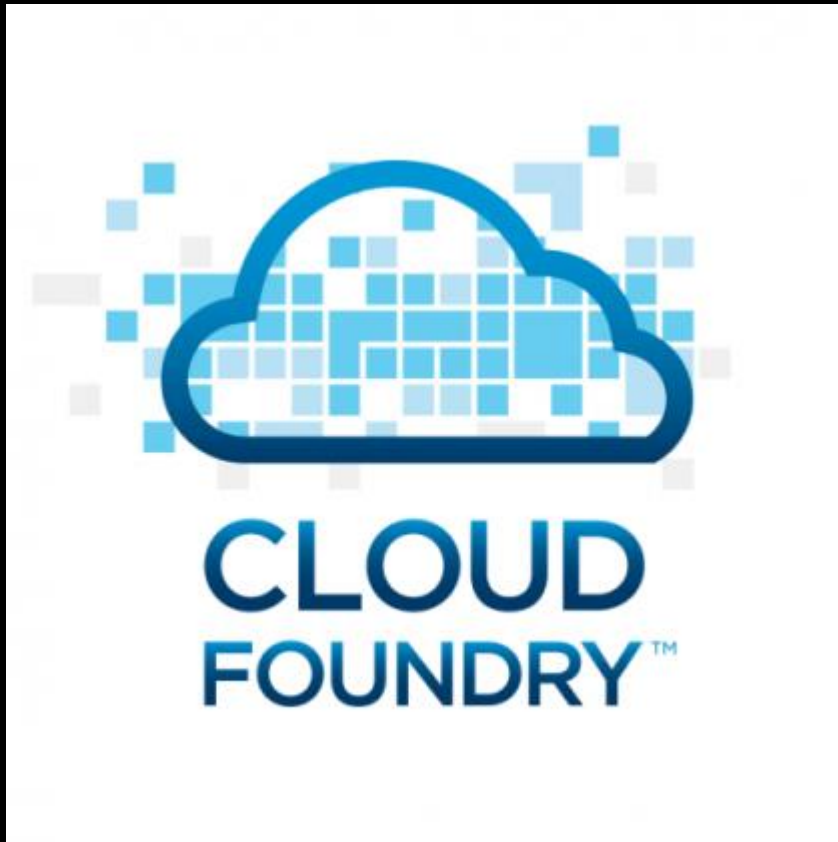
# Types of Cloud - Visibility

- Public
  - e.g. AWS, Google
- Private
  - Similar model, but internal
  - e.g. VCE
- Hybrid
  - Public/Private combination



Source: <http://emergingtechblog.emc.com/wp-content/uploads/2015/01/Cloud-Strategy-300x209.png>

# PaaS – Platform as a Service



## Influencing Factors

1. “The emergence of new application frameworks, which are primarily developer-led and open-source.”
2. “The emergence of large-scale cloud platforms as a new computing engine offering both the capability of operating at scale and a new deployment model.”

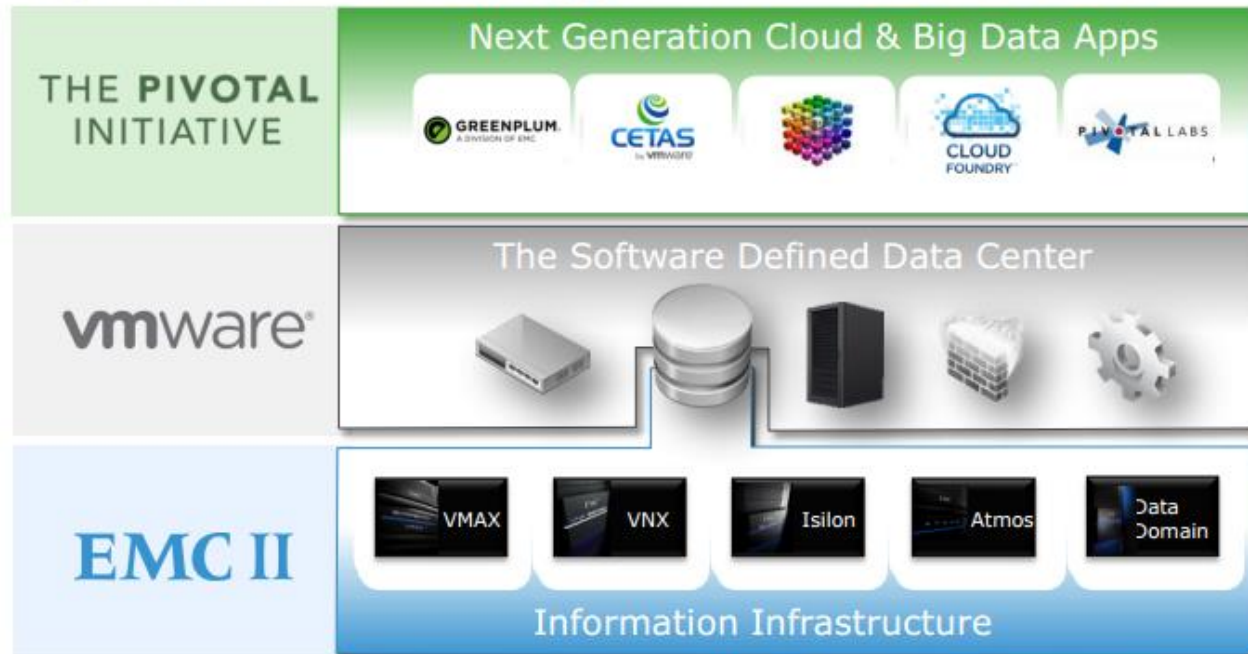
*Paul Maritz, then CEO VMware (CF Launch Event, April 2011)*



# The Pivotal Initiative

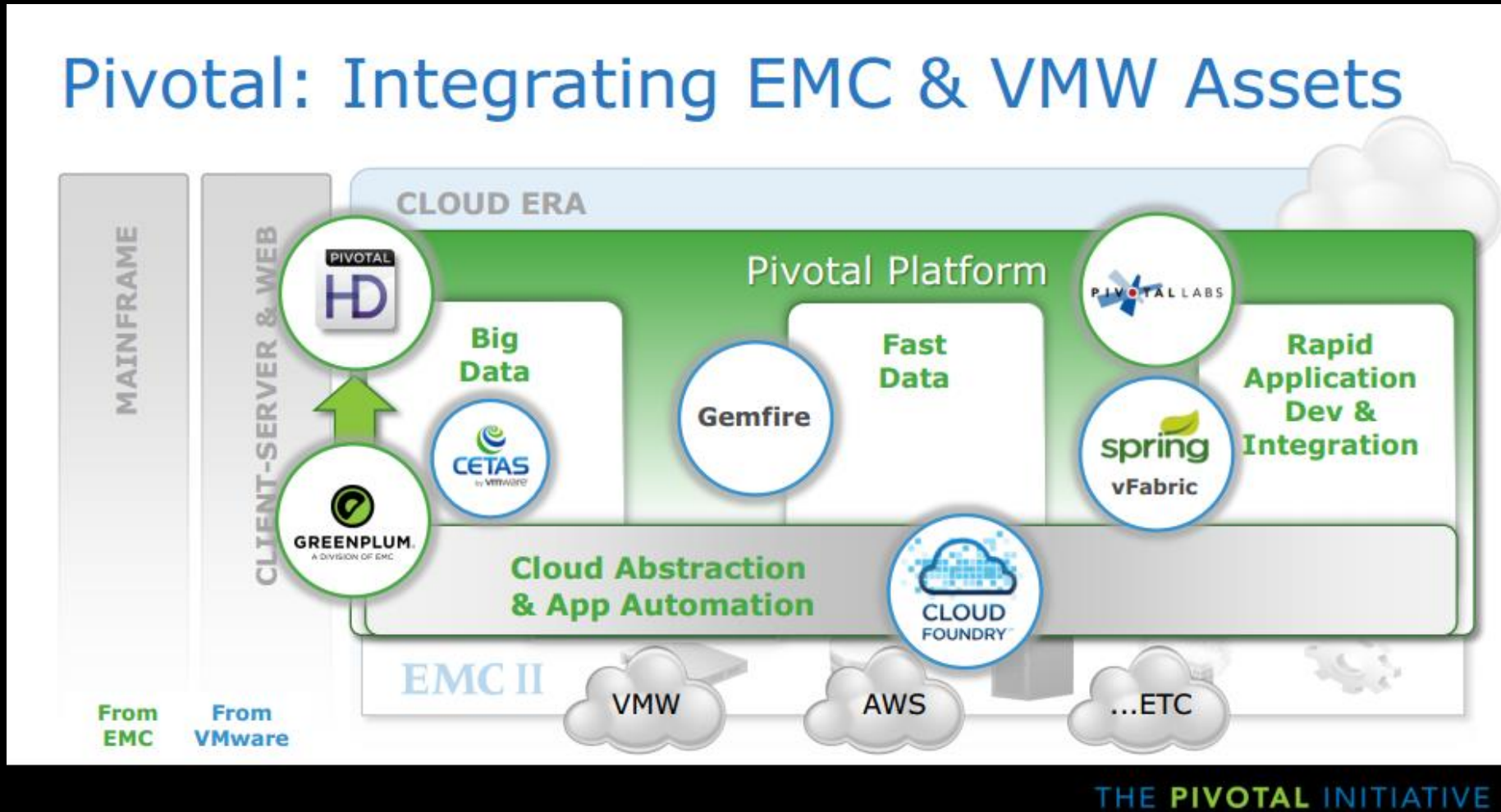
## Open Architecture Provides Choice

Clean Separation Between Application & Infrastructure Layers

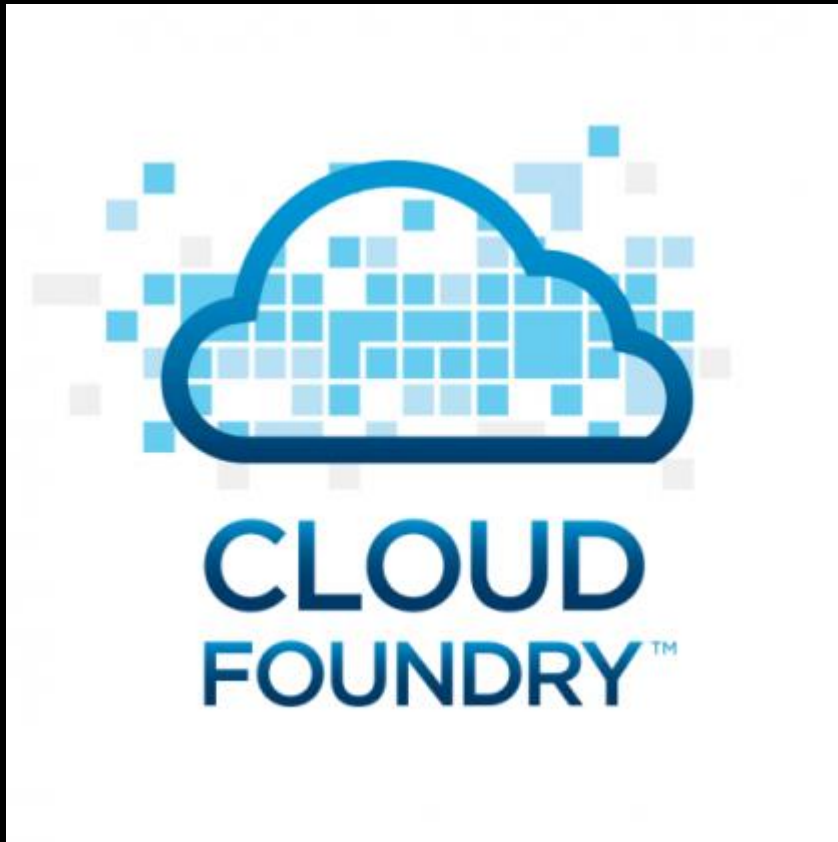


THE PIVOTAL INITIATIVE

# The Pivotal Initiative

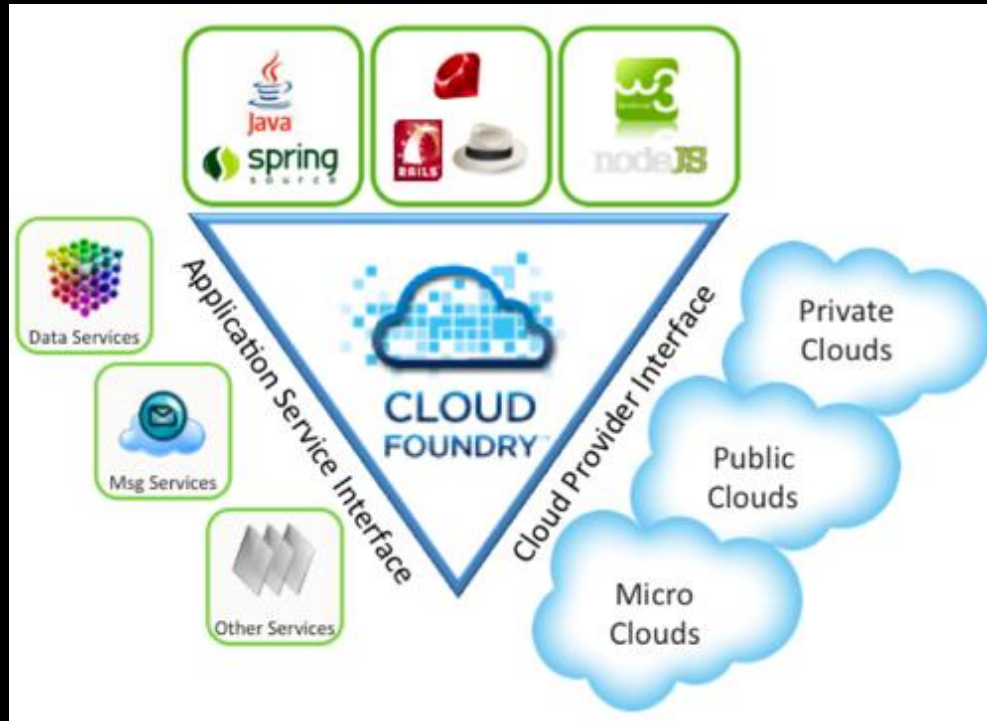


# Cloud Foundry



- Open Source
- Extensible Open PaaS
  - Developer Frameworks
  - Application Services
  - Cloud Platform
- Portability
  - Freely move to another CF provider (without modification)
  - Deploy your own CF instance
- Cloud Platform Agnostic
  - Deploy CF on any cloud platform (e.g. AWS)

# Cloud Foundry



- Developer Frameworks
- Application Services
- Cloud Environments

Source: <https://spring.io/blog/2011/04/12/launching-cloud-foundry-the-industry-s-first-open-paas>

# Implications for Engineering

1. New Deployment Model
2. Capability of Operating at Scale

Q. What does 'cloud'  
mean for  
Architects?

# A. Scalability

# Scalability

## Vertical (Scale-up)

- Increase size & resources of host

## Horizontal (Scale-out)

- Deploy more hosts (processing)



# Scale-up (Vertical Scaling)

## Advantages

- Easier
  - Increase Memory & CPU (VM)
  - Larger Host
- Multicore Processors
  - Increases in parallelism, not speed

## Disadvantages

- Increased Memory increases House-Keeping
  - Sorting/Searching
  - Garbage Collection
- Mutually Exclusive Operations
  - Limit Scaling Effect
- Finite Limits
  - e.g. JVM Limitations

# Scale-out (Horizontal Scaling)

## Advantages

- Easy to deploy more hosts
  - Can be automated/elastic
- Load spread across large number of hosts
- Relatively cheap\*

## Disadvantages

- Needs to be designed to be 'replicated'
  - Requires a 'modular' design
  - Need to achieve 'statelessness'
- Additional Internal Management
  - Greater reliance on Middleware
  - More communication/coordination
- Complexity
  - Violates 1<sup>st</sup> Law of Distributed Systems

Q. What does 'cloud'  
mean for  
Architects?

# A. Scalability

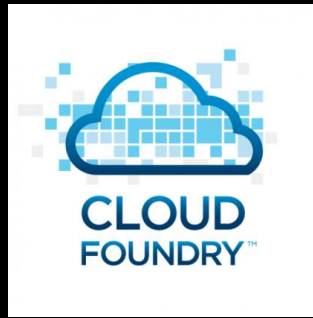
A. Scalability  
Horizontal

*“Cloud computing does not automatically solve scalability challenges – it merely facilitates the possibility of solving them.”*

October 2013

## Donnacha's Law of Cloud Computing & Scalability

Definition



# Achieving scale-out with Cloud Foundry

## Application Tier

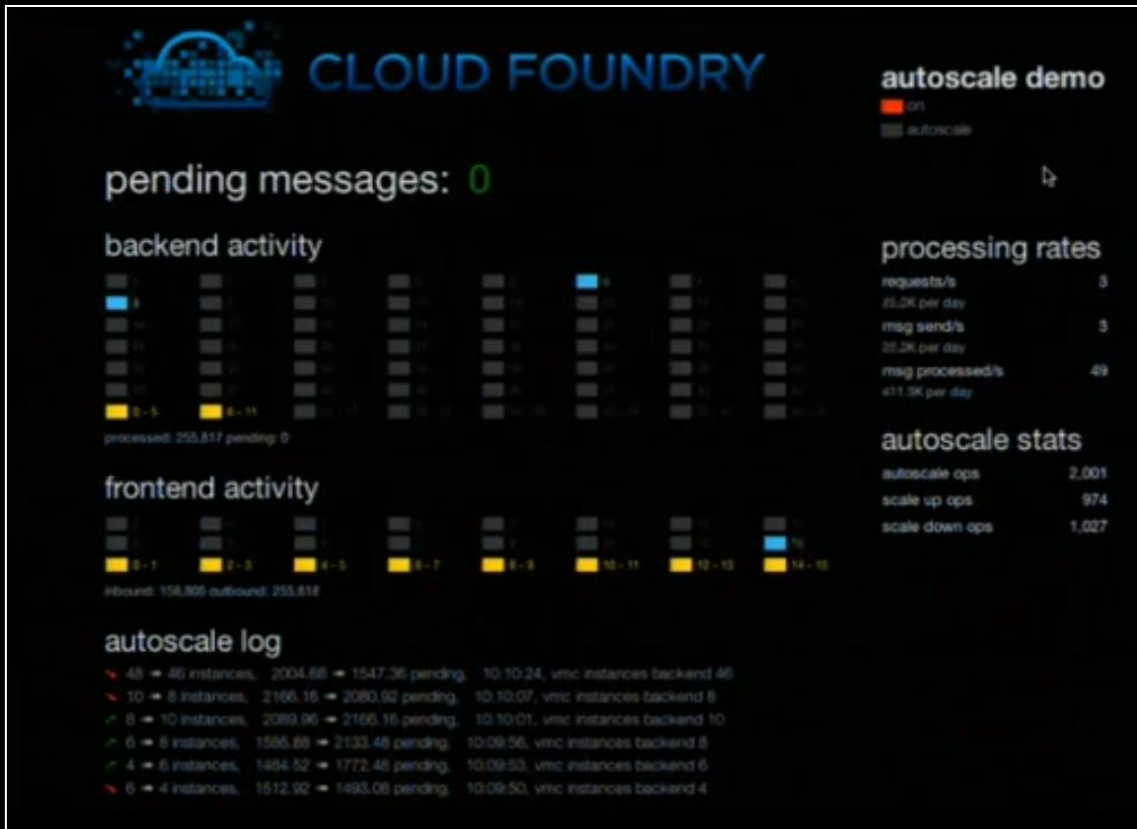
- Modularization & Replication
- Asynchronous Communication
  - Decoupling
  - Queuing
- Orchestration
  - Dynamic Scaling (up and down)

## Persistence Tier

- DBMS Replication & Sharding
- Polygot Persistence
  - RDBMS
  - Object-Relational
  - NoSQL/Unstructured

Microservices

# Dynamic Auto-scaling with Cloud Foundry



- Multi-node Microservices
  - Backend Services
  - Frontend Services
- Message Passing
  - Asynchronous
  - De-coupled
- Relies on Queues
- Relies on Statelessness

Source: Cloud Foundry Launch Event, April 2011 (Part 6 – Scale Out) <https://www.youtube.com/watch?v=Eu72MKWJnYQ> (@ 6:17)



# Mobile

A computer in everyone's hand

# Mobile Device Characteristics

## Features

- Smart Devices (Networking/Data)
- GPS Enabled
- Camera
- Media Rich (Browsing/E-Mail/IM)
- Extensible Apps
- Calendar/PDA
- Music (iPod/MP3-Player)

## Usage

- Rapid Adoption
- Spontaneous & Instantaneous & Continuous generation of data
- Rich Developer Community & Platform



# Developer Focus

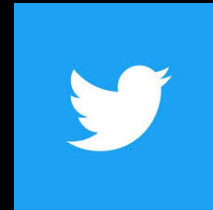
- App Paradigm
  - Platform Specific Implementation
  - Use platform technology
  - Move away from Web
    - (e.g. no FlashPlayer on iOS)
- Proliferation makes mobile 1<sup>st</sup> class citizen
  - e.g. Google Search bias (towards mobile friendly sites)
- Challenge for developers
  - Mobile Screen
  - Tablet Screen
  - Laptop Screen
- AppStore model
  - Delivery & Distribution model for developers
  - Removes Barriers to market entry
- Conducive to 'Innovation' from start-ups
- Scale – potential number of users

# Social

Social Media

# Web 2.0 → Social Media

- Blogs
- Podcasts
- Vlogs
- Wiki
- RSS Feeds → Autofeed
- Instant Messaging
- Facebook
- Twitter
- LinkedIn
- Instagram
- Snapchat



# Social: Characteristics

- Unstructured Data
  - Textual
  - Images
  - Video
- Non-transactional
  - Can afford to be a little out of date
- Scale
  - Exponential data storage trends
  - Cheaper storage
  - Cheaper processing
- Internet Scale
  - Industry developing new solutions & technology
  - Consumption
  - Storage
  - Delivery
  - Distribution
- Dynamics of Computing Changes at Scale
  - Require next generation tech.
  - IT unlike anything we've had to deal with

# Social: Value

- Consumer shift
  - More likely to complain on Twitter
  - Turn to YouTube to learn
  - Use discussion boards to ask questions & get answers
- Companies strive to preserve image & reputation
  - Digital Marketing
  - Subtle Advertising
- Economies of Scale
  - Socializing problems & finding existing solutions
- Statistical value at large-scale
  - Sentiment Analysis
  - Comparative Analysis
- AI & Machine Learning
  - Listening to the conversation

# Big Data

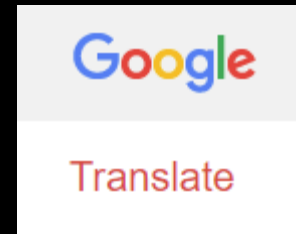
Big Data & Analytics





# Large Data Sets

- When data sets get to be large, very large, they become statistically relevant to the point of great or greater accuracy.
- e.g. Google Translate
  - Search & Compare versus Expert System
  - Volume versus Accuracy → Statistical Relevance





English Spanish French Detect language ▾

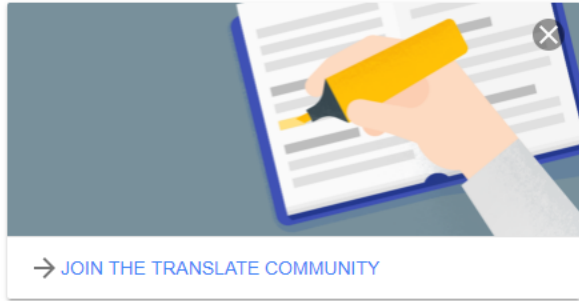


English Spanish Arabic ▾

Translate

0/5000

Type text or a website address or [translate a document](#).



Google Translate for Business: [Translator Toolkit](#) [Website Translator](#)



English Spanish French Spanish - detected ▾



English Spanish Arabic ▾

Translate

sin



3/5000



Suggest an edit

without

## Definitions of sin-

*preposition*

Indica falta o carencia de alguna cosa material o inmaterial.

"un abrigo sin bolsillos; una persona sin escrúpulos; no se quedará sin amigos; lo hizo sin confianza ."

Indica que hay algo que no se ha tenido en cuenta en una suma o en una cantidad; se usa especialmente para hablar de gastos.

"despilfarro sin medida; la compra de la casa me ha resultado muy cara, aun sin los impuestos ."

Prefijo que entra en la formación de palabras con el sentido de 'unión', 'simultaneidad'.

Prefijo de origen latino que entra en la formación de nombres con el significado de la preposición sin.

"sinnúmero, sinrazón, sinfin ."

## See also

sin embargo, sin palabras, sin duda, sin que, sin nombre, sin miedo, sin parar, sin sentido, sin querer, sin piedad

## Translations of sin

*preposition*

- without sin
- out of fuera de, por, sin, falta
- minus menos, sin, falto de

*suffix*

- less sin

*adverb*

- except sin



English Spanish French Catalan - detected



English Spanish Arabic

Translate

sin embar

9/5000

without embar

Suggest an edit



English Spanish French Spanish - detected



English Spanish Arabic

Translate

sin embargo



11/5000



Suggest an edit

Nevertheless

## Synonyms of sin embargo

*adverb*

en vez de

See also

sin, embargo

## Translations of sin embargo

*conjunction*

however	sin embargo, no obstante, por más que
but	pero, sino, sino que, que, sin embargo, sin
nevertheless	sin embargo, no obstante
nonetheless	sin embargo, no obstante
notwithstanding	no obstante, sin embargo, a pesar de que

*adverb*

though	sin embargo
all the same	a pesar de todo, sin embargo

# The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, Google

## The Unreasonable Effectiveness of Data

Peter Norvig



- Boole Lecture at UCC, December 2009
- UBC Computer Science Distinguished Lecture Series  
<https://www.youtube.com/watch?v=yvDCzhbjYWs>

# Big Data & Analytics

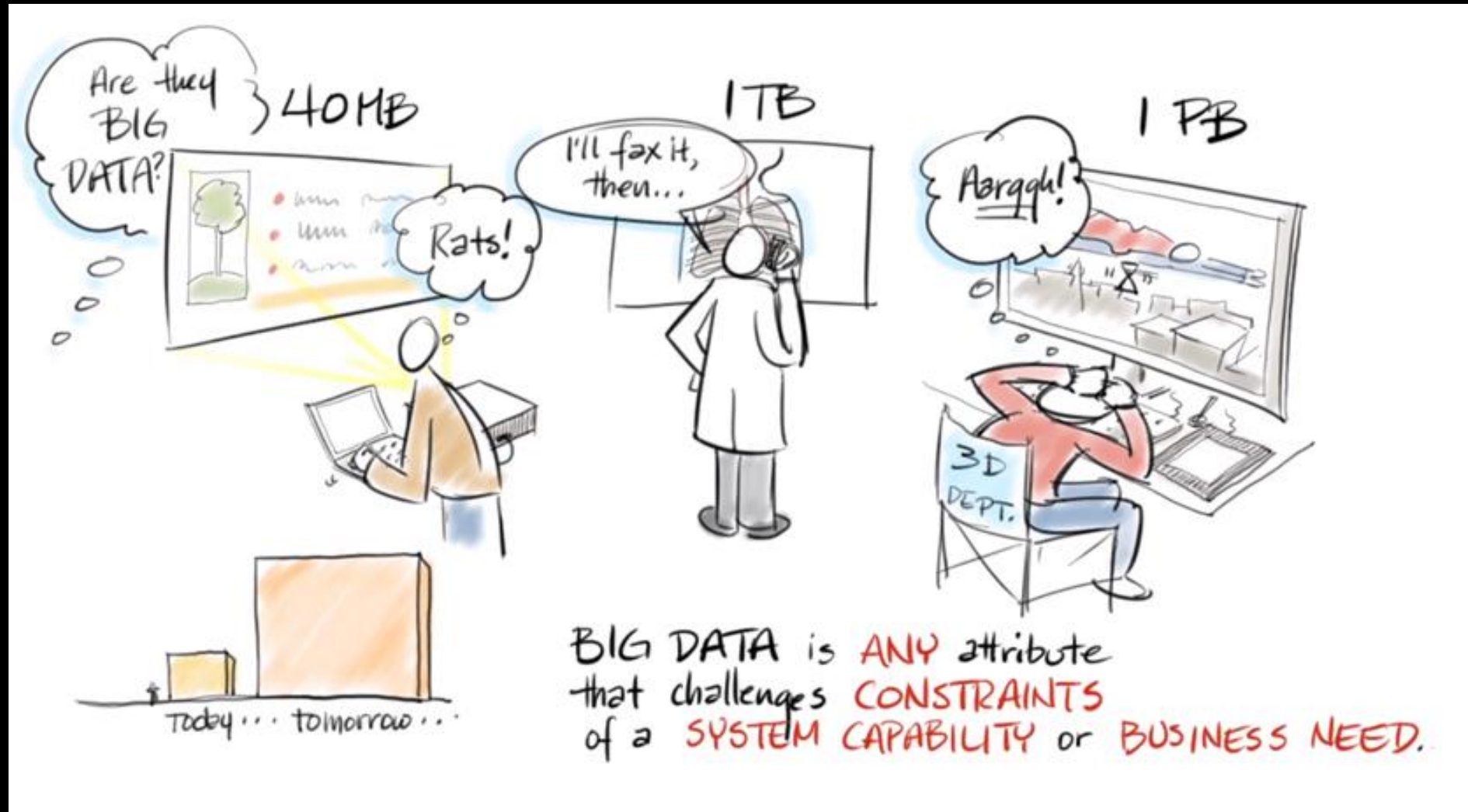
- Scientific Method → Expert Systems (GOFAI) → Statistical Machine Learning



*Good Old-Fashioned Artificial Intelligence*



# Big Data - How big is 'Big Data'?



*“It’s ‘Big Data’ if it pushes the limits of the common technology available to utilize it.”*



Patricia Florissi, EMC

# Big Data Attributes

- Speed
  - Doesn't refer solely to the size of the data
  - Relates to the speed at which data is generated
- Number & variety of sources
  - Sheer volume of data to be analysed within a given timeframe or geographical boundary
  - e.g.
    - Number of Mobile Devices
    - Internet of Things (IoT) – i.e. Smart Meters in homes
- Not all 'Big Data' is the same in structure
  - e.g. unstructured data from social media
  - e.g. structured data from a transactional database
- Generation Diversity & Interconnectedness
  - e.g. Automated Generation or Manually Created
  - Incremental Updates about the data, going from vague to precise over time

Vídeo

How big is Big Data?

# Big Data

- Drives the need to make sense out of the data
- Find meaning & relationships

# Big Data & Analytics

- Comparative Analytics → Predictive Analytics
- Systems Management
  - How is my <whatever> doing in relation to everyone else's?
  - When is my <part of the system> going to go wrong?
- Pre-emptive Management
  - Notice: We just replaced <component> in your system

# How do we architect solutions to support these types of systems?

1. We need capture & consume very large amounts of data.
2. We need to persist/store very large amounts of data.
3. We need to query/analyze very large amounts of data.
4. We may need to exchange or distribute very large amounts of data.

# How do we architect solutions to support these types of systems?

- The Problem – Our existing, mainstream database technology doesn't seem well suited to address the problem at hand.
- The Solution – We need new types of databases
- RDBMS best suited to well-formatted, structured data.
- Not all data requires ACID criteria to integrity
- Even scaled-up DBMS engines on very large hosts are not capable of storing huge datasets.
- Querying these datasets requires massive CPU & memory resources, beyond that of even these large hosts.



# Greenplum



- Highly scalable, shared-nothing database

## MPP Shared Nothing Architecture

Flexible framework for processing large datasets

Master Host and Standby Master Host

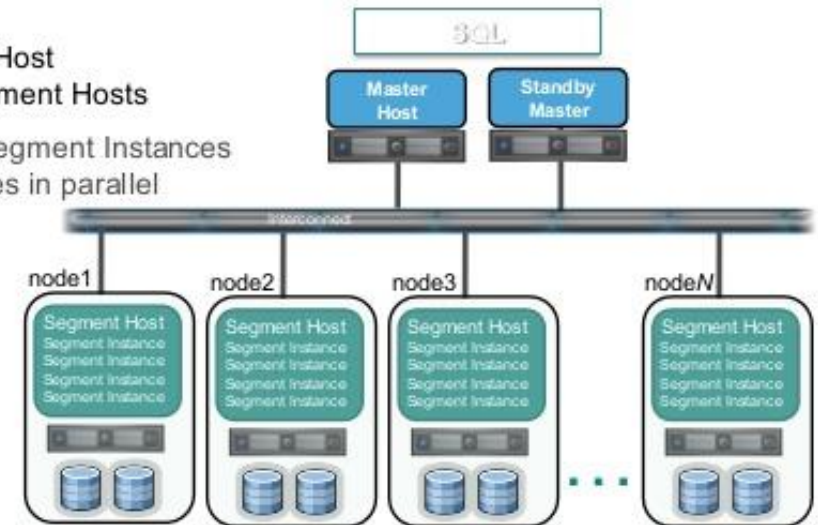
Master coordinates work with Segment Hosts

Segment Host with one or more Segment Instances

Segment Instances process queries in parallel

Segment Hosts have their own CPU, disk and memory (shared nothing)

High speed interconnect for continuous pipelining of data processing



Pivotal.

© Copyright 2013 Pivotal. All rights reserved.

Source: <https://www.slideshare.net/AGrishchenko/greenplum-architecture>

# Greenplum Device



- Greenplum = Software Solution
- Acquired by EMC in 2010
- Deployed on dedicated hardware
  - Fast disks, network, etc.
- Now part of Pivotal

Source: <https://www.storagereview.com/news/emc-greenplum-appliance-based-unified-analytics-platform-now-available>

# NoSQL

## Priorities

- Scalability is more important than consistency
- Solves for 'eventually consistent'
- Close match between model and data
  - e.g. document-based
  - e.g. Name/Value pairs

## Cost

- Less features or less sophisticated than tradition SQL engines
  - Transactional support
  - Query Facilities
  - Data Management

# NoSQL Players

- **Column:** Accumulo, Cassandra, Druid, HBase, Vertica, SAP HANA
- **Document:** Apache CouchDB, ArangoDB, Clusterpoint, Couchbase, Cosmos DB, HyperDex, IBM Domino, MarkLogic, MongoDB, OrientDB, Qizx, RethinkDB
- **Key-value:** Aerospike, ArangoDB, Couchbase, Dynamo, FairCom c-treeACE, FoundationDB, HyperDex, InfinityDB, MemcacheDB, MUMPS, Oracle NoSQL Database, OrientDB, Redis, Riak, Berkeley DB, SDBM/Flat File dbm
- **Graph:** AllegroGraph, ArangoDB, InfiniteGraph, Apache Giraph, MarkLogic, Neo4J, OrientDB, Virtuoso
- **Multi-model:** ArangoDB, Couchbase, FoundationDB, InfinityDB, MarkLogic, OrientDB

Source: <https://en.wikipedia.org/wiki/NoSQL>

# Other Big Data Technologies

- Apache Cassandra
- Apache Hbase
- MongoDB
- Neo4j
- Apache CouchDB
- OrientDB
- Infinispan (Jboss)
- Terrastore
- FlockDB
- Hibari
- Riak
- Hypertable
- Blazegraph
- Hive
- InfoBright
- Infinispan
- Redis
- Greenplum
- GemFire (Apache Geode)

<https://www.datamation.com/data-center/slideshows/16-open-source-big-data-databases.html>

# The emergence of Hadoop



# Google

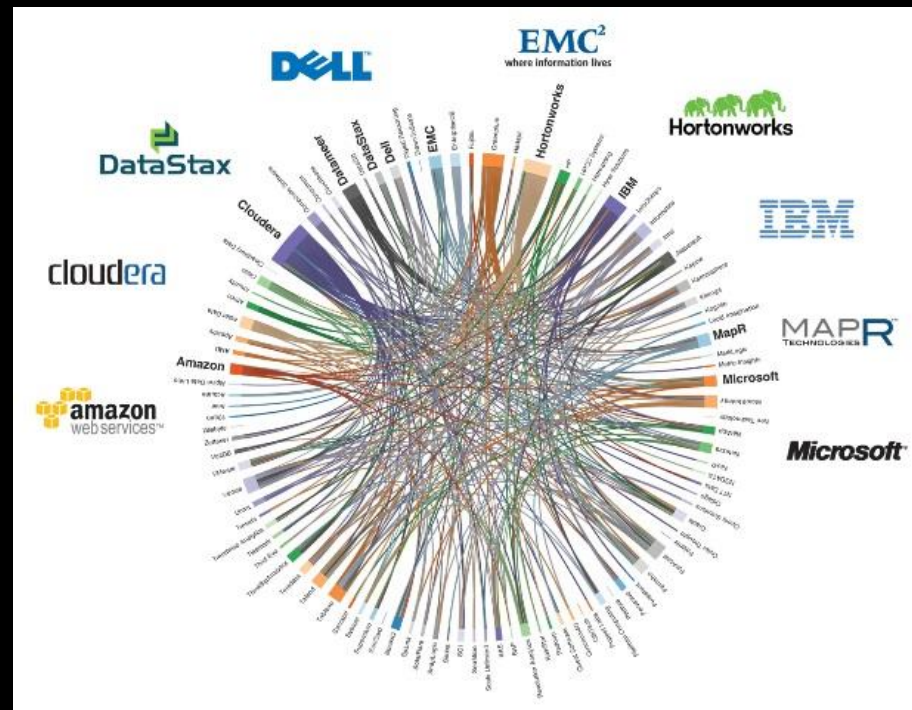
- MapReduce: Simplified Data Processing on Large Clusters (2004)
  - <https://research.google.com/archive/mapreduce.html>
- Bigtable: A Distributed Storage System for Structured Data (2006)
  - <https://research.google.com/archive/bigtable.html>
- GFS: The Google File System (2003)
  - <https://research.google.com/archive/gfs.html>

# Yahoo! *hadoop*

- Hadoop
  - Doug Cutting & Mike Cafarella
- Storage = HDFS (Hadoop Distributed File System)
- Processing = MapReduce



# Other Hadoop Players



# Vídeo

## Demystífyíng Hadoop

# Summary

- The Third Platform
  - New generation of architecture
  - New generation of developer, with different engineering practices
  - Now capable of solving hard/large engineering problems

Q&A

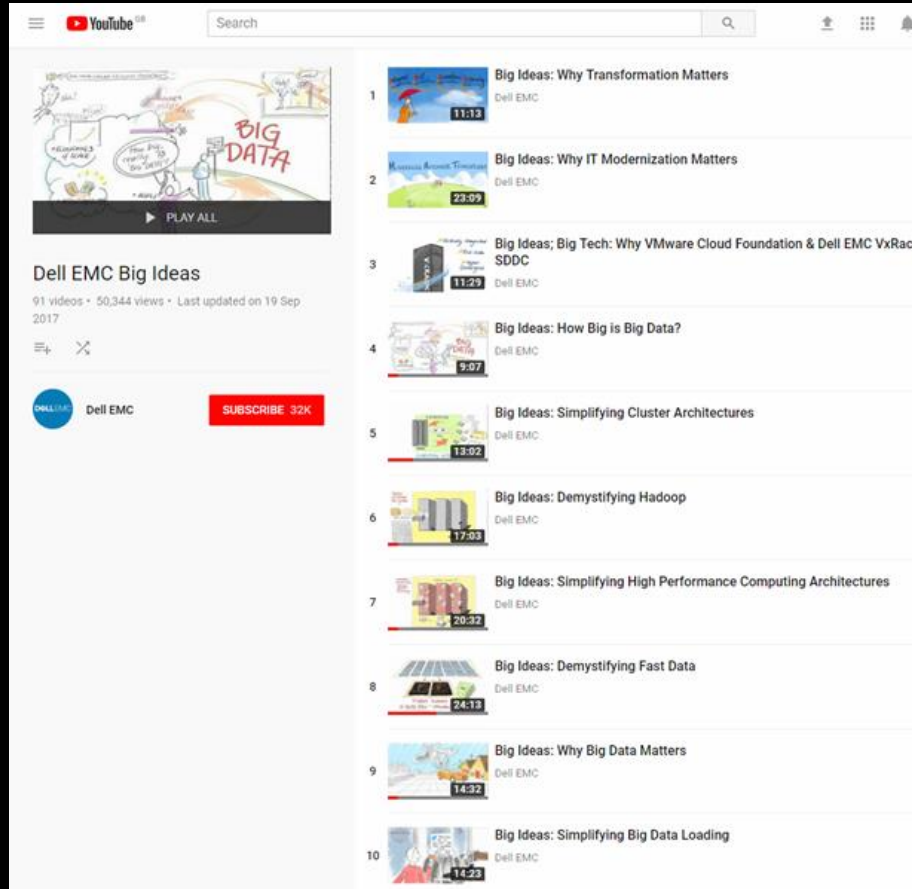
Discussion Time

# Questions

- Can on-prem solutions learn from cloud deployments?

Thank You

# Recommended Reading



## • Dell EMC Big Ideas

- [https://www.youtube.com/playlist?list=PLD298CBF8D0908E4C&feature=view\\_all&list=PLD298CBF8D0908E4C&feature=view\\_all](https://www.youtube.com/playlist?list=PLD298CBF8D0908E4C&feature=view_all&list=PLD298CBF8D0908E4C&feature=view_all)

1. How big is Big Data?
2. Cluster Architectures
3. Hadoop
4. High Performance Computer Architectures
5. Fast Data
6. ...

# Recommended Reading

- Dell EMC Big Ideas
  - [https://www.youtube.com/playlist?list=PLD298CBF8D0908E4C&feature=view\\_all&list=PLD298CBF8D0908E4C&feature=view\\_all](https://www.youtube.com/playlist?list=PLD298CBF8D0908E4C&feature=view_all&list=PLD298CBF8D0908E4C&feature=view_all)
- Peter Norvig – The Unreasonable Effectiveness of Data
  - <https://www.youtube.com/watch?v=yvDCzhbjYWs>



