Cambrian Explosion

NoSQL East

Don't Panic

- You are in the correct place….
- This is a cloud presentation
- And we will talk about NoSQL, big data and the clouds…

When all you have is a hammer everything looks like an IT Infrastructure.

06:48 PM July 13, 2008 from web ☆ ☺

botchagalupe
botchagalupe

- **johnmwillis.com**
  - **Cloud Café Podcasts**

- **redmonk.com/cote**
  - **IT Management Guys Podcast**

- **botchagalupe**
  - **twitter and gtalk**

whoami

Cambrian Explosion

580 Million Years Ago...

Environmental, developmental, and ecological changes accelerated the rate of evolution by an by an order of magnitude.

Cambrian Explosion

"The rate of evolution of any system is dependent upon the organisation of its subsystems"

Herbert Simon, Theory of Hierarchy and componentisation.

Cambrian Explosion

"A Period of Combinatorial Innovation"

Hal Varian a Chief Economist for Google

Historically, you'll find periods in history where there would be the availability of a different component parts that innovators could combine or recombine to create new inventions. In the 1800s, it was interchangeable parts. In 1920, it was electronics. In the 1970s, it was integrated circuits.

Cambrian Explosion

Today

I would argue we are at the base of an explosion of innovation driven primarily by four factors:

- Abstract and fault tolerant components
- Unlimited infrastructure
- World wide collaboration
- Changes in intellectual property

Abstract and Fault Tolerant Components

- Components are abstract
- Bits, software, virtualized
- Clustered
- Fault tolerant
- Parallelized

Unlimited Infrastructure

- You never run out of them
- So there's no shortage
- There are no inventory delays
- You can reproduce them
- You can duplicate them, you can spread them around the world

World Wide Collaboration

- You can have thousands and tens of thousands of innovators combining or recombining the same component parts to create new innovation.

- It's a situation where the components are available for everyone.

World Wide Collaboration

Mechanical Turk Funds	CRITICAL	10-22-2009 11:12:59	0d 21h 53m 53s	2/2	$0.075 left for turking
PING	OK	10-22-2009 11:13:03	0d 16h 4m 49s	1/2	PING OK - Packet loss = 0%, RTA = 0.04 ms

Changes in Intellectual Property

- Revolution in how we view intellectual property
- Open source
- Google – I can be an expert on anything in a matter of minutes… worst case one day.

Big Data (NoSQL)

• UnStructured

• Think "Use Case" first

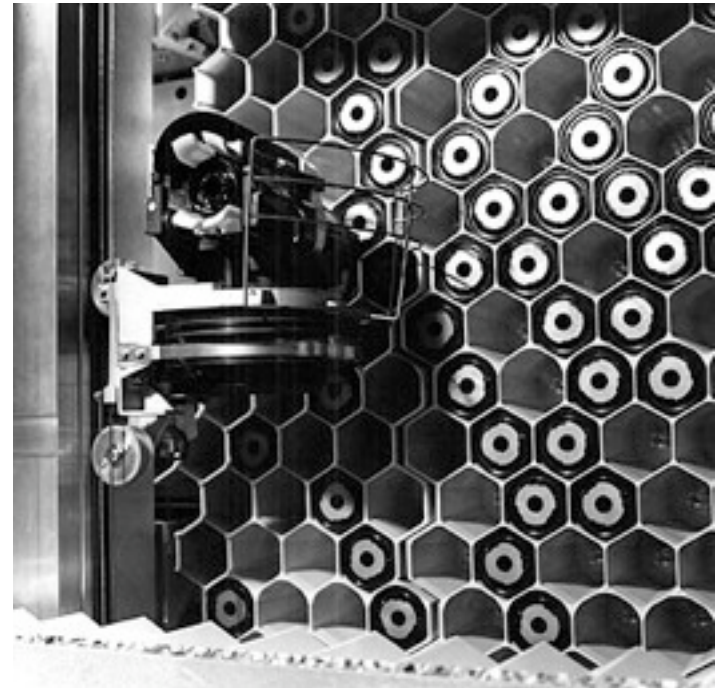• Not everything needs to be in Tables and Columns

Premature optimization is the root of all evil

Donald Knuth

Cambrian Explosion
Big Data 1980

- IBM 3850 Mass Storage
- 500Gig
- $20,000,000.00

Big Data 2010

- Western Digital
- 500Gig
- $50.00

Big Data defined

A History Lesson

• Paxos separate safety properties from liveness properties (Eventual Timely Links)

• Eric Brewer's (CAP) Theorem CAP Theorem

• Microsoft work with storage systems (Boxwood/Paxos

• Google's publishing of papers on GoogleFS, Map/Reduce algorithm, Chubby (lock service) and BigTable

• Amazon's work with Dynamo and then S3 (Eventual Consistency)

• Yahoo's work with Map Reduce and Apache Hadoop

• Cloudera further work with Hadoop

Big Data Folk

- Cassandra
- Redis
- Pig
- Cascading
- MongoDB
- CouchDB
- Dynomite
- Riak
- Hbase
- tin
- Neo4j
- Voldermort
- Dryad
- Sherpa
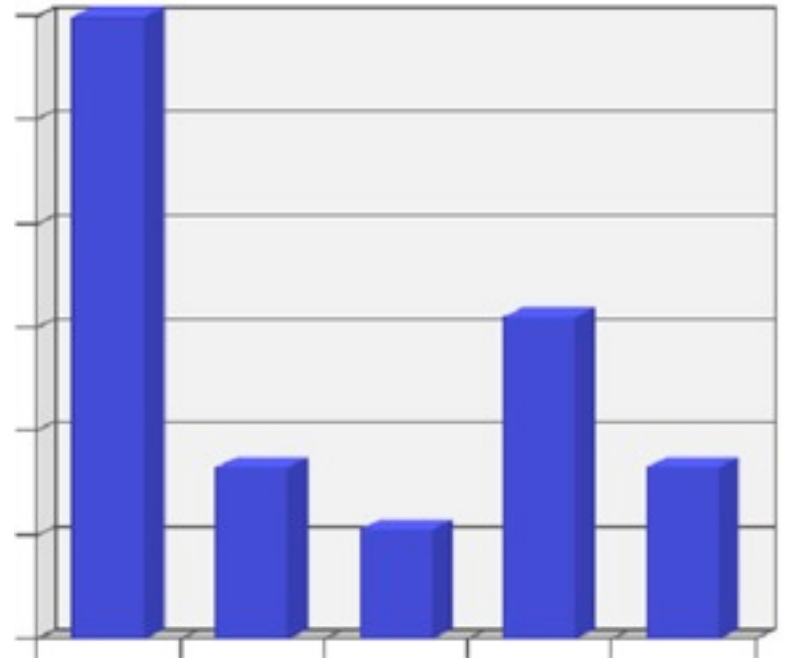
What Tools Do I use?

Big Data Choices

VS

Big Data Choices

VS

Big Data Choices

VS

Big Data Stories

Hadoop Sort Benchmarks (10B 100 byte records)

- 2007 – 1TB 4.95 minutes (400 nodes)
- 2008 - 1TB 3.48 minutes (910 nodes)
- 2009 - 1/2TB 1 minute (1406 nodes)
- 2009 – 100TB 173 minutes (3452 nodes)

Big Data Stories

Examples

- Rackspace Mailtrust
- Silverpop

Flightcaster

- Heroku (UI)
- Cloudera Hadoop
- Cascading
- Clojure (JVM)

The Cambrian Dude!