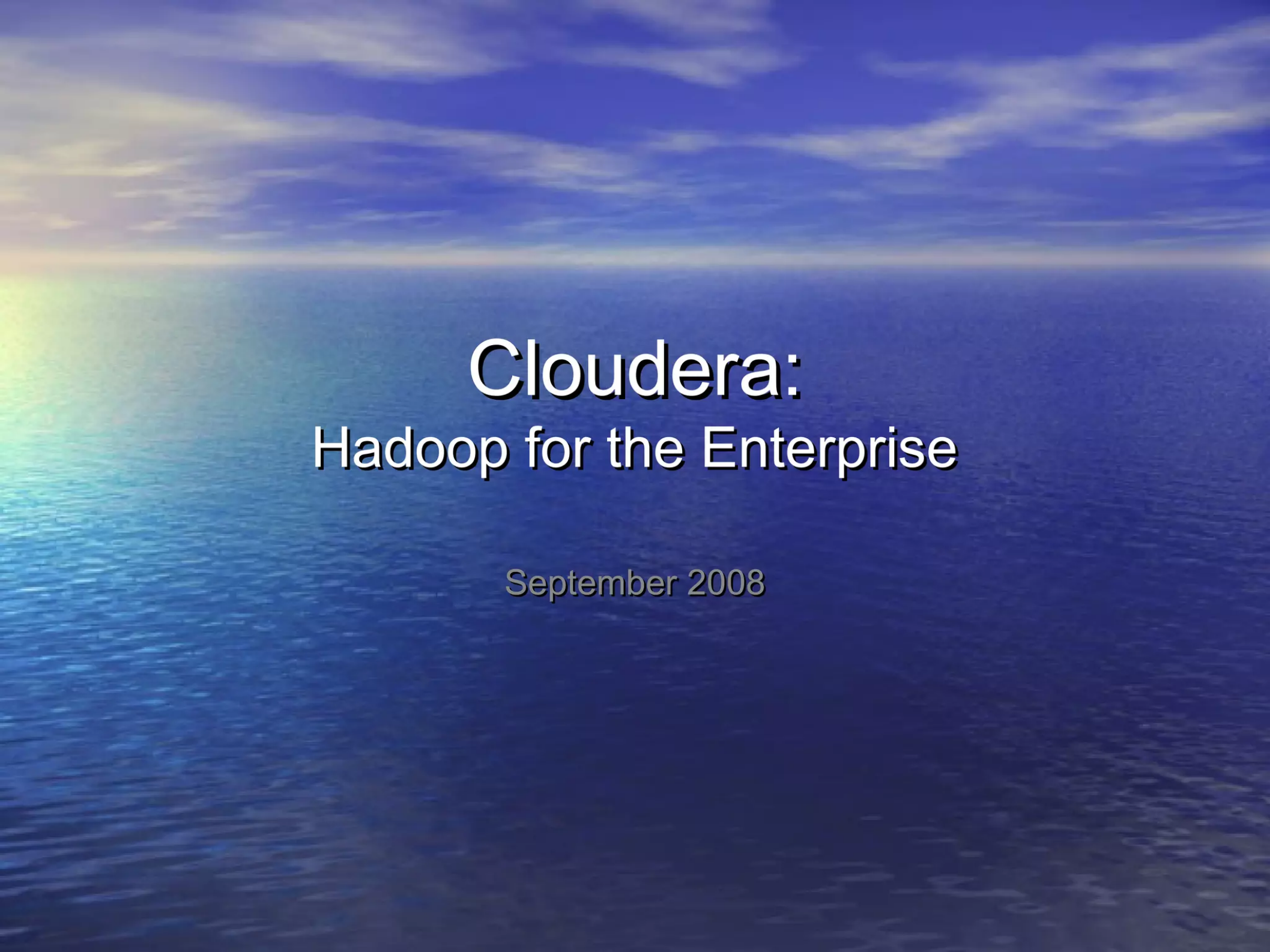


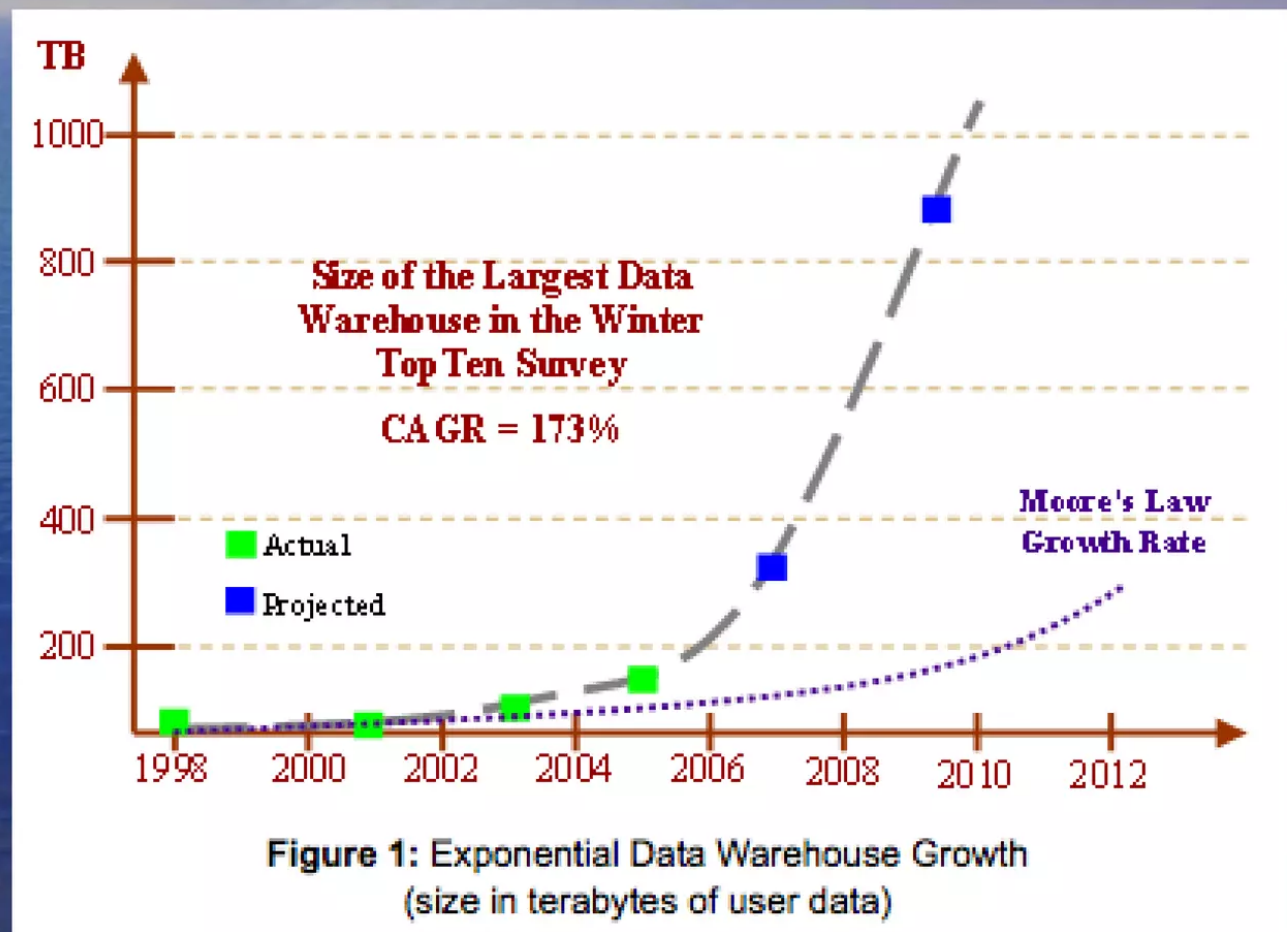


Cloudera: Hadoop for the Enterprise

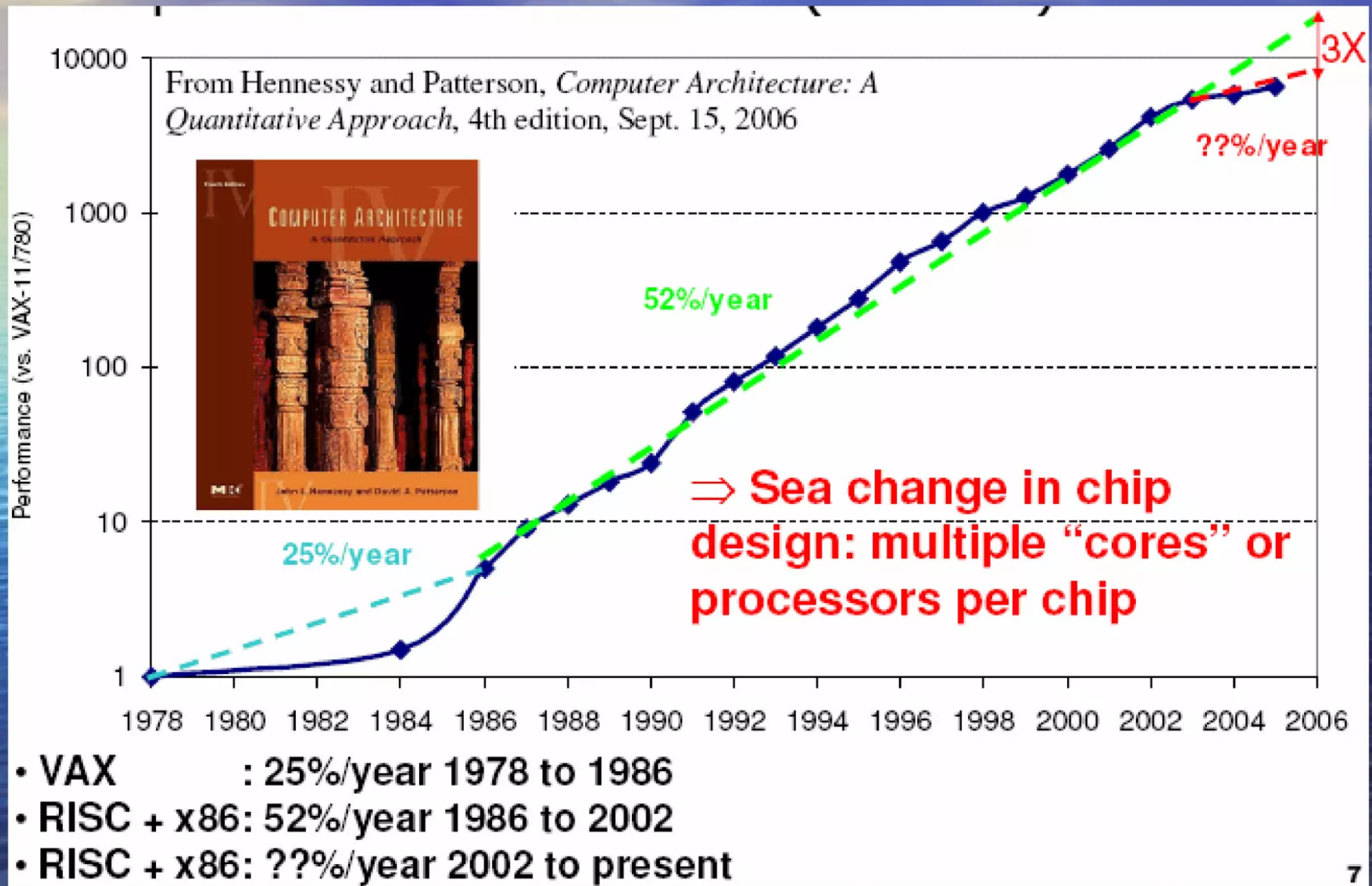
September 2008

Data Growing Much Faster than Moore's Law

Source: Richard Winter,
*Why Are Data
Warehouses Growing
so Fast?*, April 2008



Uniprocessor Performance



Founding Team

- Mike Olson, CEO
 - CEO Sleepycat
 - Britton Lee, Illustra, Informix, Oracle
 - BA, MS CS, Berkeley
- Amr Awadallah, CTO, VP Engineering
 - Founder Aptivia/VivaSmart
 - 8 years at Yahoo! running BI infrastructure, including Hadoop
 - PhD EE, Stanford
- Christophe Bisciglia, VP Technology
 - Created Google/NSF Hadoop cluster and program
 - BA CS, U Washington
- Jeff Hammerbacher, VP Product
 - Ran world's largest operational BI support system on Hadoop, at Facebook
 - BA Mathematics, Harvard

What Is Hadoop?

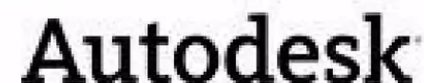


- Core engine:
 - Open source implementation of Google's MapReduce and GFS
 - Hundreds or thousands of servers parallelize a data analysis task
- Interfaces built on top of MapReduce
- Storage layer beneath (HDFS)
- Doug Cutting, Mike Cafarella are advisors

Hadoop is Open Source

- Hadoop is distributed under the Apache License:
 - Reduces concern about lock-in
 - Low-cost, effective distribution strategy
 - Allows innovation by partners, customers
 - Third-party inspection of source code provides assurances on security, product quality
- Business-friendly license encourages commercial development
 - “Open core” licensing
 - Closed-source components, applications

Hadoop Users



Momentum: Google Trends



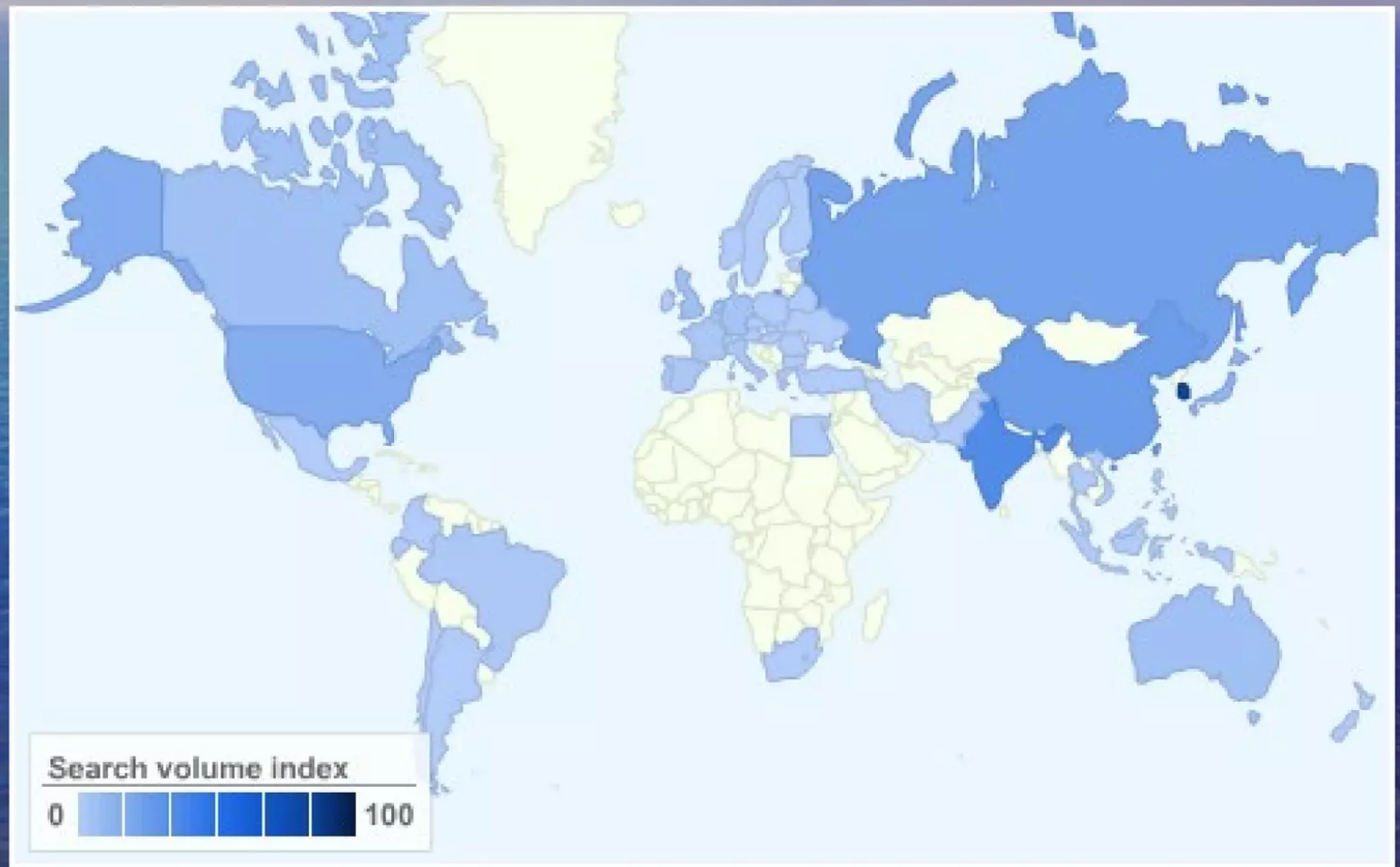
Netezza: \$127M in FY08, \$79M in FY07

Teradata: \$830M in 1H08, \$1.7B in FY07

04/21/17

Worldwide Phenomenon

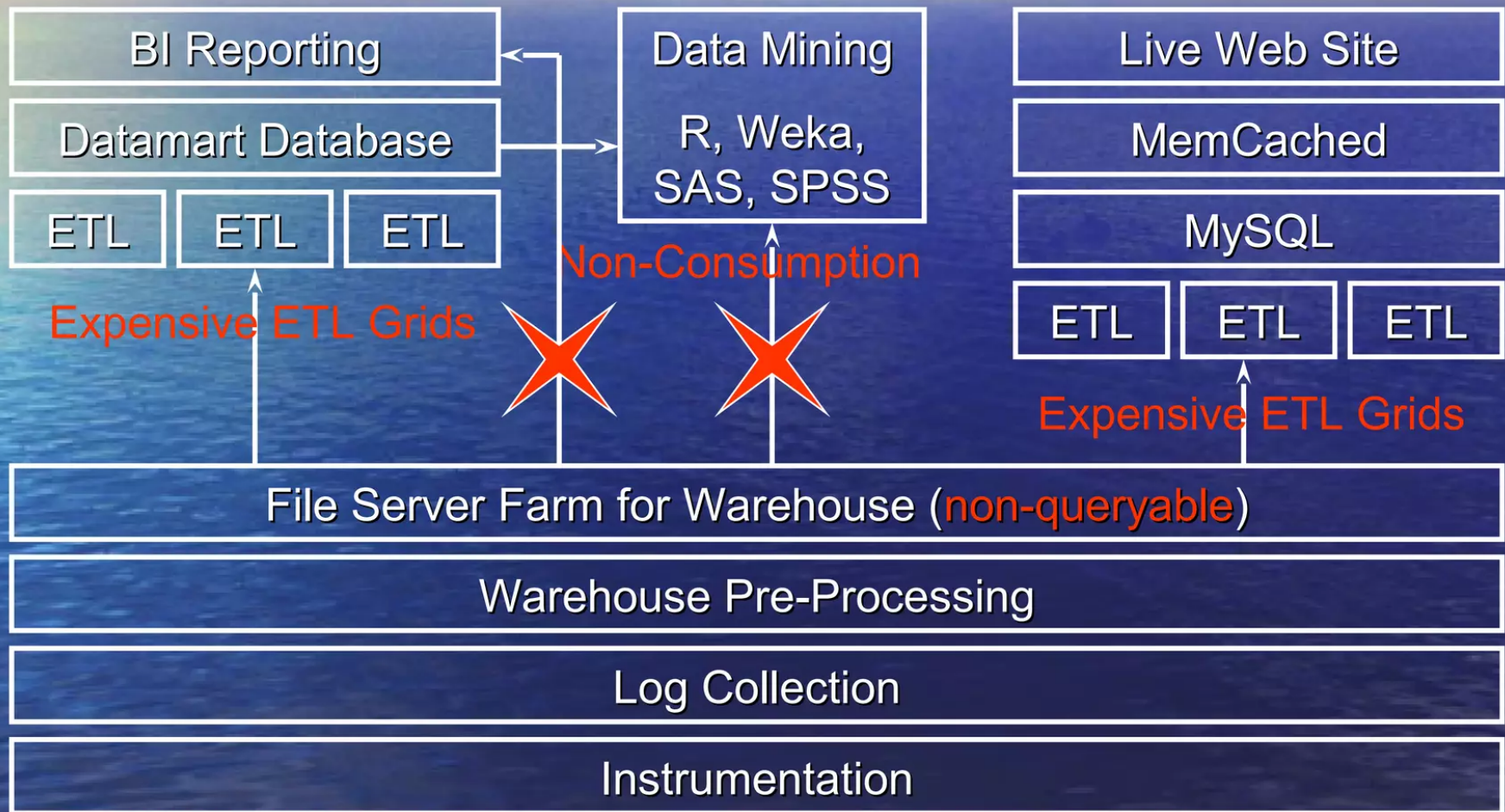
Source:
Google Insights
world map for
searches on
“hadoop”,
Sept 2008.



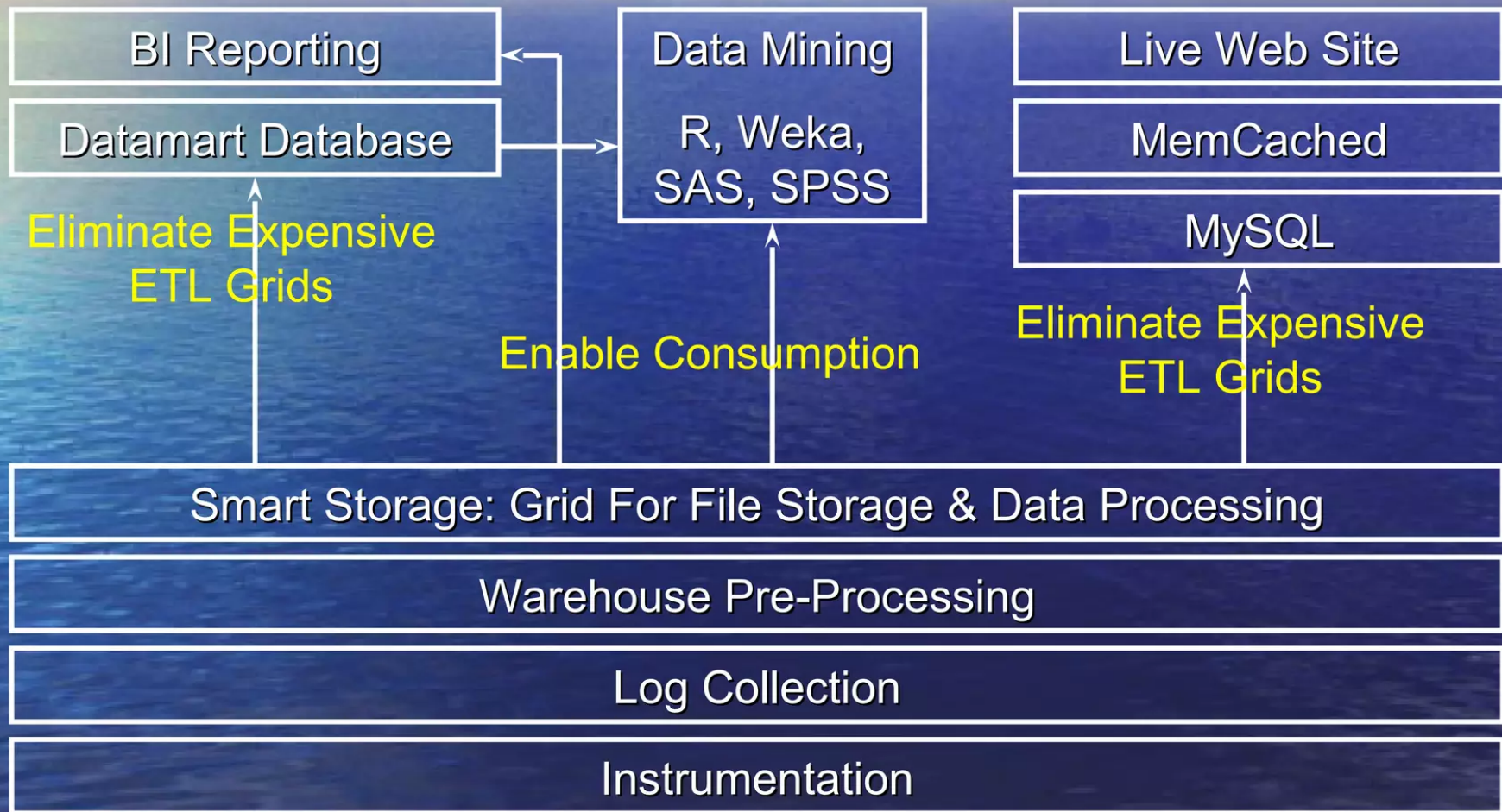
Why is Hadoop Successful?

- Brings computation closer to data allowing both IO and compute scalability.
- Map-Reduce forces developers to think in a parallel way
- Operates on unstructured data, and structured data (HBASE, HIVE)
- Prescriptive development, grows with you without needing to re-architect
- Procedural language offers power

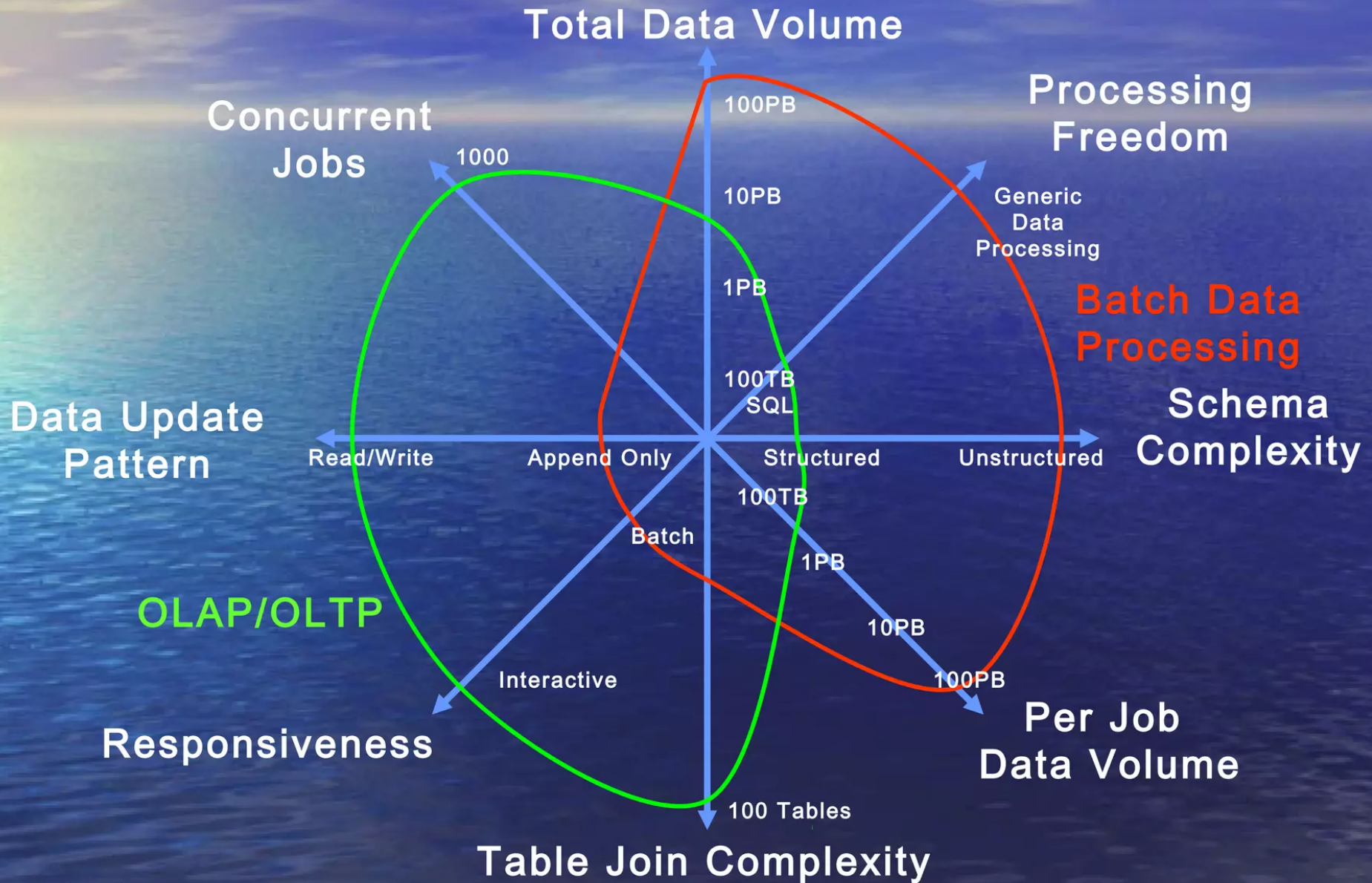
Current Systems Isolate Users from the Event Level Raw Data



Solution: “Smart” Storage Service



BDP versus OLAP/OLTP



The Cloud Wars: \$100+ billion at stake

■ The Cloud - A multi-year shift in the computing paradigm

We are in the midst of a pronounced shift from client-server to Cloud computing, which is more analogous to centralized mainframe computing. Quantum improvements in Internet bandwidth, computing power and memory, coupled with enabling technologies like virtualization, parallel processing and multi-core chips, make it feasible to run large computing tasks on a centralized 'Cloud' infrastructure. The economics are truly compelling, with cost advantages of 3-5x for business apps, and 5-10x or better for personal productivity apps.

Shift creates a \$100+ billion opportunity

Cloud equivalents exist today for most business and personal productivity apps. Starting in the enterprise as OnDemand apps, roughly a \$2 billion software segment. Cloud apps are moving into personal productivity (e.g., email, word processing). Cloud software is not as mature as client-server, but the trajectory is changing. The total \$160bn addressable market opportunity includes \$95 billion in business and productivity apps, and another \$65 billion in online advertising.

Source:
Merrill Lynch
Industry
Overview,
May 7, 2008

Cloudera Differentiators

- Enabling Hadoop as an elastic platform with statistical multiplexing over many customers
- **Multi-Tenant Support:** Concurrency, Priority, Namespace Isolation, Performance Isolation.
- **Monitoring, Reliability, and Availability**
- **Resilience and Fast Recovery:** A non-sexy problem that is critical to enterprises, no time to restart ETL job from scratch, otherwise misses SLA.
- **IDE** to easily debug, deploy, and tune.
- Integration with data mining and analysis functionality (R, Weka, SAS, SPSS)
- **Connector certification:** another non-sexy problem that is ignored by community, make sure system is compatible with other enterprise systems.