

Timo Raitalaakso
+358 40 848 0148
rafu@solita.fi

TUT – NoSQL Seminar (Oracle) Big Data

11.12.2012

SOLITA

Timo Raitalaakso

- › MSc 2000
- › Work: Solita since 2001
 - › Senior Database Specialist
- › Oracle ACE – 2012
- › Blog: <http://rafudb.blogspot.com>

SOLITA

Agenda

- › How to make your data big
- › Big Data needs
- › Big Data technology

SOLITA

Relational model and Normal Form

- > Year 1970
- > “Many of the existing formatted data systems provide users with tree-structured files or slightly more general network models of the data.”

A Relational Model of Data for Large Shared Data Banks

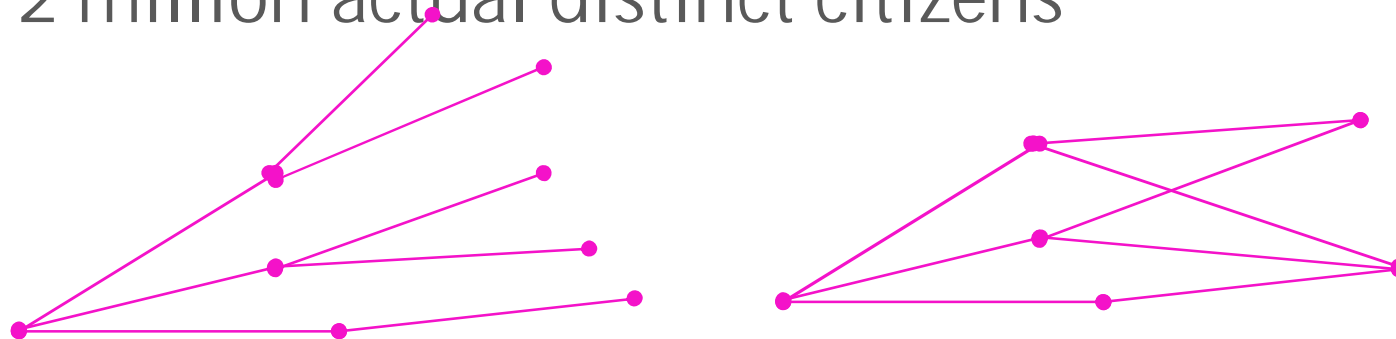
E. F. CODD
IBM Research Laboratory, San Jose, California

<http://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf>

SOLITA

Network vs Relation

- > 10 million finnish citizens in a network model
- > Converted to a relational model
- > 2 million actual distinct citizens



- > Duplicating your - data you get more data

SOLITA

Changes Stored

- › Transaction time
 - › You choose
 - › or does your technology force you to store the whole history
- › Closed World Assumption
- › How about Bitemporal needs?
 - › Valid time dimension of temporal data

SOLITA

Big Operational Data
Big Interaction Data

Big Data technology

SOLITA

Disney's platform is growing at 5TB a day,
and there are still many other types of data
it needs to house

<http://gigaom.com/data/how-disney-built-a-big-data-platform-on-a-startup-budget>

SOLITA

Data Growth



Research firm IDC:

- › CERN generates 1PB/sec during experiments
- › Boeing jets generate 20TB of data per hour.
- › Twitter generates 12TB/day for average traffic
- › Wal-Mart's data stores are 2.5PB
 - › (1m transactions/sec)

http://myexpospace.com/oracle2012/SessionFiles/CON5840_PDF_5840_0001.pdf

SOLITA

Oracle Open World 2012 - Big Data - topics

- > [Implementing Big Data Solutions: From Theory to Practice](#)
- > [Being a Data Scientist with Oracle Big Data](#)
- > [Building an Integrated Data Warehouse with Oracle Database and Hadoop](#)
- > [Big Data Performance Delivered: Oracle NoSQL Database on Cisco UCS](#)
- > [Mining Big Data with RDF Graph Technology: Discovering What You Didn't Know](#)
- > [Overcoming Analysis Paralysis with Big Data in Healthcare](#)
- > [Big Data Analytics: Enabling Innovation and Creating Value](#)
- > [Optimal Oracle Systems for Big Data Performance and Scalability](#)

<http://www.oracle.com/openworld/index.html>

SOLITA

Oracle Open World 2012 - Big Data - topics

- › [Hadoop, Oracle Exadata, and the Industrial Revolution of Data](#)
- › [Hadoop Meets Oracle Exadata](#)
- › [Formula for Growth: Innovation, Analytics, and Big Data](#)
- › [Accelerate Big Data Adoption with Less Risk and Complexity](#)
- › [High-Performance, Integrated Products for Diverse Stages of Big Data Solutions](#)
- › [Oracle NoSQL Database and Big Data](#)
- › [Oracle Big Data Connectors: Using Data from a Hadoop Cluster with Oracle Database](#)

SOLITA

UKOUG 3.-5.12.2012 - Big Data topics

- › [KEYNOTE: The Next Generation of Oracle Business Analytics: Endeca, Big Data, and Beyond](#)
- › [Formulating Your Big Data Strategy: It's Not About Technology - It's About Business Value; It's Not About Data - It's About Insight From Data](#)
- › [SPONSOR: The Next BIG Thing?](#)
- › [Big Data Meets Big Process](#)
- › [NoSQL Databases for Implementing Data Services – Should I Care?](#)

SOLITA

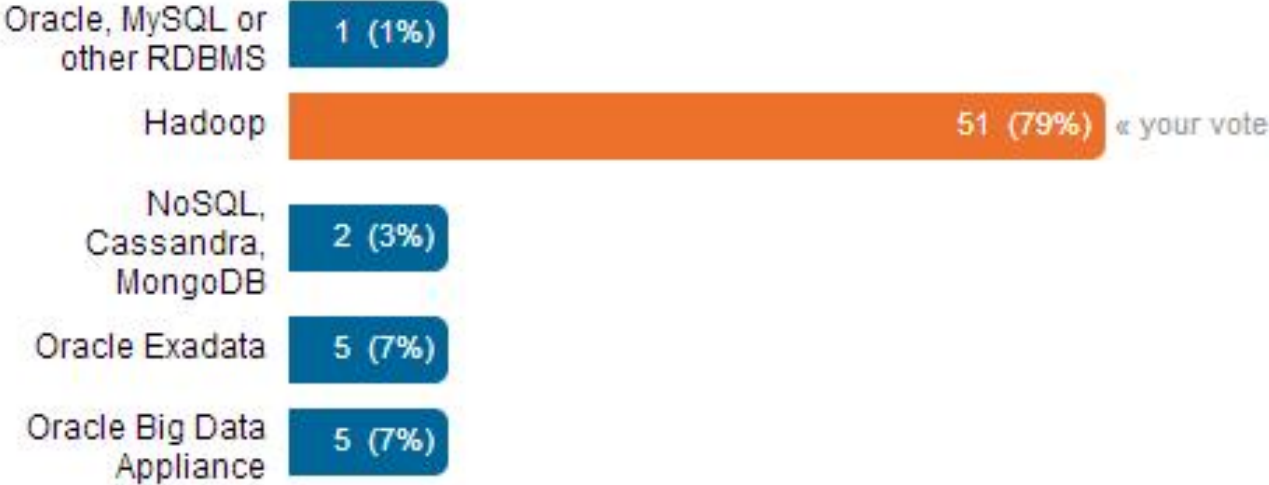
Linked in group – IOUG Big data SIG



Follow Alex

What is the #1 software product comes to mind when you are asked about Big Data?

posted 12 days ago • 64 votes



<http://www.linkedin.com/groups/IOUG-Big-Data-SIG-4612788>

SOLITA

Hadoop



- › Apache project
- › A distributed filesystem
- › Add storage easily
- › Fast inserts
- › Map Reduce Queries

SOLITA

Big Data Appliance - HW

- > 18 Compute and Storage nodes
- > 648 TB of raw storage
- > InfiniBand Connectivity between nodes and across racks



<http://www.oracle.com/us/products/database/big-data-appliance/overview/index.html>

SOLITA

Big Data Appliance – SW - Integrated

- › Oracle Enterprise Linux 5.6
- › Oracle Hotspot Java Virtual Machine
- › Cloudera's Distribution including Apache Hadoop
- › Cloudera Manager
- › Open Source Distribution of R
- › Oracle NoSQL Database Community Edition

<http://www.oracle.com/technetwork/server-storage/engineered-systems/bigdata-appliance/overview/bigdataappliance-datasheet-1453665.pdf>

SOLITA

Oracle NoSQL Database

- › Key/value store
- › Rewritten by Oracle's Berkeley DB team

- › Education available:
 - › Introduction to Oracle NoSQL Database
 - › http://education.oracle.com/pls/web_prod-plq-dad/db_pages.getCourseDesc?dc=D75054GC10

SOLITA

R

- › Language and environment for statistical computing and graphics
- › For analyzing and visualizing data
- › GNU project

<http://www.r-project.org/>

SOLITA

Delivery day



<http://kerryosborne.oracle-guy.com/2012/09/oracle-big-data-appliance-delivery-day/>

SOLITA

Rapid development

- › After first boot
- › Red Hat to latest release
- › Cloudera to latest release
- › ...

SOLITA

Oracle Big Data Appliance First Boot



<http://karlarao.wordpress.com/2012/10/16/oracle-big-data-appliance-first-boot/>

SOLITA

Tera – Peta – Exa

Metric prefixes							
Prefix	Symbol	1000^m	10^n	Decimal	Short scale	Long scale	Since ^[n 1]
yotta	Y	1000^8	10^{24}	1 000 000 000 000 000 000 000 000	septillion	quadrillion	1991
zetta	Z	1000^7	10^{21}	1 000 000 000 000 000 000 000	sextillion	trilliard	1991
exa	E	1000^6	10^{18}	1 000 000 000 000 000 000	quintillion	trillion	1975
peta	P	1000^5	10^{15}	1 000 000 000 000 000	quadrillion	billiard	1975
tera	T	1000^4	10^{12}	1 000 000 000 000	trillion	billion	1960
giga	G	1000^3	10^9	1 000 000 000	billion	milliard	1960

<http://en.wikipedia.org/wiki/Exa->

SOLITA

Oracle 11g rdbms MAX db size?

(Max db_files)*(max bigfile tbs size)
65536*128TB=8388608TB=8EB

SOLITA

Exadata

- › 11.2 Oracle RDBMS
- › Exadata Smart Scan
- › Exadata Smart Flash Cache
- › Exadata Hybrid Columnar Compression
- › InfiniBand Network

<http://www.oracle.com/us/products/database/exadata/>

SOLITA



Exadata X3-8

- › Two 8-socket database servers
- › 14 Oracle Exadata Storage Servers
- › InfiniBand switches
- › more than 22 terabytes of Exadata Smart Flash Cache

SOLITA

Exadata X3-2

- › Eight 2-socket database servers
- › 14 Oracle Exadata Storage Servers
- › InfiniBand switches
- › More than 22 terabytes of Exadata Smart Flash Cache

- › Also Half, Quarter and Eighth Rack available

SOLITA

Exadata – Features

- › 128 CPU cores and 2 TB memory for database Processing
- › 168 CPU cores dedicated to SQL processing in storage
- › 22.4 TB of Exadata Smart Flash Cache
- › 40 Gb/second (QDR) InfiniBand Network
- › Uncompressed and mirrored usable capacity of up to 224 TB per rack

SOLITA

Oracle Exadata Storage Expansion Rack X3-2

- › Over 4.5 Petabytes of user data can be stored in a rack using the included Hybrid Columnar Compression



SOLITA

Mid game - how much could fit inside

- > A HD feature film
10GB
- > Could all feature
films in HD fit
inside Exadata +
Storage Expansion
Rack ?



The screenshot shows the IMDb website header with the logo and navigation menu. Below the header, the page title is "IMDb Database Statistics". The statistics listed are:

Category	Count
Total Titles	2,336,782 (Year Range: 1880 - 2019)
TV episodes	1,408,540
Feature films	279,383
Shorts	249,383
Videos	75,103

<http://www.imdb.com/stats>

SOLITA

Even more

- › Scale the configuration by connecting up to 8 Exadata Database Machines and Exadata Storage Expansion Racks with cables
- › Larger configurations can be built with additional InfiniBand switches



SOLITA

Power
consumption

10,5 kW +
airconditioning



SOLITA



Map Reduce in SQL

- › Implementing MapReduce Programs with SQL and PL/SQL on Oracle Exadata
 - › https://oracleus.activeevents.com/connect/sessionDetail.ww?SESSION_ID=4809
 - › Smart scan / indexes while map filtering
- › “MapReduce for Java developer change your thinking more like functional programming”
 - › https://oracleus.activeevents.com/connect/sessionDetail.ww?SESSION_ID=5840
- › <http://rafudb.blogspot.fi/2009/10/coding-dojo.html>

SOLITA

BDA + Exadata

- › WHITE PAPER: High Performance Connectors for Load and Access of Data from Hadoop to Oracle Database
 - › " , it was possible to load up to 12TB/hour from Oracle Big Data Appliance to Oracle Exadata " = 3,6GB/sec
 - › Measured with X2-8 Exadata
<http://www.oracle.com/technetwork/bdc/hadoop-loader/connectors-hdfs-wp-1674035.pdf>
- › X3 twenty times faster writing
 - › OOW 2012 Larry Ellison keynote
<http://www.youtube.com/watch?v=uyUrVQJERKY>

SOLITA

Oracle's Big Data Platform



SOLITA

Cost- BDA - 3 years support - Compared

BDA:

for US\$ 450,000 you get the following:

- Big Data Appliance Hardware
- Cloudera CDH and Cloudera Manager
- Oracle Linux
- Distribution of Open-source R

+ support
US\$ 626,150

HP + Cloudera Do-it-Yourself System

- 18 servers and 3 IB switches
- Cloudera Enterprise subscription
- ...

DIY system US\$ 762,849
at all list prices

https://blogs.oracle.com/datawarehousing/entry/price_comparison_for_big_data

SOLITA

Cost - Exadata

Eighth Rack US\$ 200,000

+ Oracle RDBMS EE CPU License

<http://www.oracle.com/us/corporate/pricing/exadata-pricelist-070598.pdf>

SOLITA



SOLITA

Summary

- › IDC expects the Big Data technology and services market to grow to \$16.9 billion in 2015 with a compound annual growth rate (CAGR) of 40 percent. IDC Worldwide Big Data Technology and Services 2012-2015 Forecast, doc #233485, March 2012
- › Although Hadoop MapReduce is well suited to problems with key/value data sets, it's not intended for operations that require complex data or transactions.
- › Acquire and store a variety of structured and unstructured information.
- › Must understand whether their use case requires subsecond interactive response or comprises somewhat slower batch operations.

<http://oracle.com.edgesuite.net/ebook/bigdata/>

SOLITA

@timoraitalaakso 

Timo Raitalaakso | rafu@solita.fi | <http://rafudb.blogspot.com>

KIITOS.

SOLITA