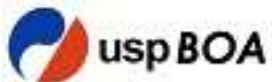




Big Data... *Small Pricetag*

Universal Shell Programming Laboratory, Ltd.



What is Unicage?

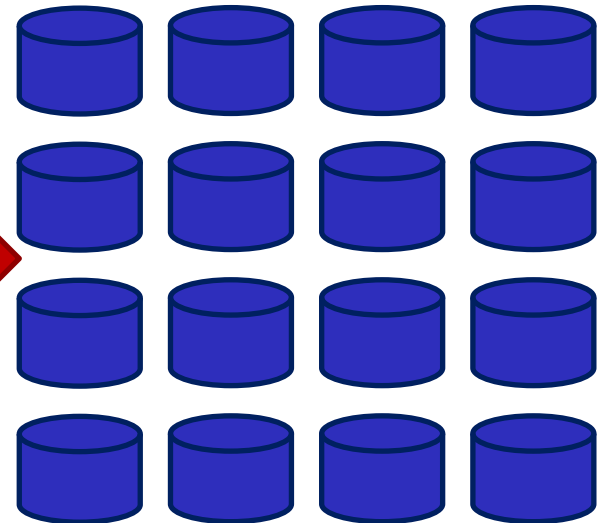


usp lab.

Unicage

is a **Big Data analytics platform**
that combines
the speed and ease of **shell scripting**
with
the power and performance of a **clustered database engine**

```
#!/bin/bash
join0 key=1 MASTER URE | Join data
self 2 3 4 5 | Select field
hsort key=1/2 | Sort
sm2 1 2 3 4 | Sum up
sm4 1 1 2 2 3 4 | Intermediate total
self 1 2 4 3 | Select Field
sm5 1 3 4 4 | Final total
map num=1 | Transpose
sed 's/A/Sales/g' | Text search/replace
sed 's/B/Profit/g' | "
keta 4 6@NF-1 | Align rows
comma 3/NF | Add commas
cat header | Attach header
tocsv > result # Output to CSV
exit 0
```



What can it do for you?



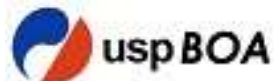
- Return search results on a dataset of **50 billion records** in less than **0.3 seconds**

Allow you to design, develop, test and implement a complex data analytics application in **days instead of months**

- Perform **ad-hoc analysis** on enormous data sets in real-time by writing **simple shell scripts**

Scale performance linearly by adding additional servers to the **cluster** without altering your programs

- Available on **Amazon-AWS** or as binaries for FreeBSD, RedHat/CentOS, Debian and MacOS



Introducing “Unicage”



usp lab.

It's a software development environment using Linux, text files and shell scripting to build “big data” information systems cheaply, quickly and flexibly.

Low Cost

Low Cost / Easy to Program

- Uses inexpensive PC-based servers and OS (Linux)
- Data is plain text, program is shell script, no middleware required
- No forklift upgrades (data and software remain same when hardware and OS change)

Fast

Fast Development Time / Fast Processing

- Programs are very short
- One engineer can design, develop and operate
- We've stripped out unnecessary features to maximize hardware performance

Flexible

Flexible

- Program is simple and easy to customize
- Software functions are not dependent on each other so can be changed easily
- Data necessary for the application is created from organized, structured data

We Follow the Unix way of Thinking



The UNIX Philosophy has been Unchanged for 40 Years

1. Small is Beautiful
2. One program (command) should only do one thing
3. Prototyping should be as fast as possible
4. Portability takes precedence over efficiency
5. Data is stored as plain text
6. Commands are used as “levers” (can be combined & reused)
7. Applications are written in shell script
8. All programs are designed as filters (pipes)

Source: “The UNIX Way of Thinking” by Mike Gancarz

Some Examples of Unicage Commands



usp lab.

Database Commands

JOIN: Table join
COUNT: Count matching records
GETFIRST: Get first matching row
GETLAST: Get last matching row
CCNT: Count columns

I/O Commands

CGI-NAME: Read data from CGI-POST
MIME-READ: Read MIME encoded data

Arithmetic Functions

PLUS: Addition
DIVK: Divide by 1000

CSUM: Sum a field
ROUND: Round a number
RATIO: Find a ratio
RSUM: Sum all fields in a record

Formatting Commands

COMMA: Add commas to number
FMERGE: Merge data into template
VCAT: Vertically concatenate
HCAT: Horizontally concatenate
TRPOSE: Transpose rows/columns
DOW: Get day of week
UP3: Merge files on key field

- 200 custom commands in total
- Independent command set honed to single functions
- Simple to use (manual-less)
- Higher performance than abstracted languages or middleware

Simple Script Example



usp lab.

Creates a summary report from data

```
#!/bin/bash
join0 key=1 MASTER URE | Join data
self 2 3 4 5 | Select field
hsort key=1/2 | Sort
sm2 1 2 3 4 | Sum up
sm4 1 1 2 2 3 4 | Intermediate total
self 1 2 4 3 | Select Field
sm5 1 3 4 4 | Final total
map num=1 | Transpose
sed 's/A/Sales/g' | Text search/replace
sed 's/B/Profit/g' | "
keta 4 6@NF-1 | Align rows
comma 3/NF | Add commas
cat header - | Attach header
tocsv > result | Output to CSV
exit 0
```

→ Execute the shell script to process the data

Scalable Architecture for Big Data

usp BOA (Big Data Oriented Architecture) :

A **Software Appliance** dedicated to Big Data Processing based on Unicage



- OSS Based
- Inexpensive H/W
- Scale Out Architecture

“Big Data” Appliances

- High-Speed Performance
- Maintenance and Support
- Stability



- OSS Based
- Scale Out Architecture
- High-Speed Performance
- Inexpensive Hardware
- Maintenance and Support
- Stability

The Best of Both Worlds

Product Features



usp lab.

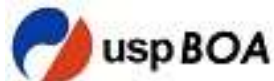
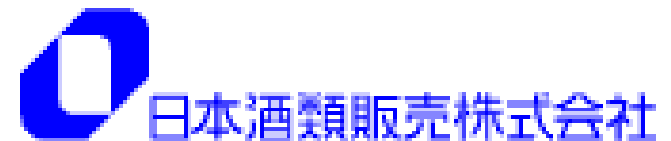
Feature	Description
OSS Base	Runs under any UNIX-based OS (Linux/FreeBSD)
Scale Out	Can be scaled out similar to Hadoop Can increase processing nodes dynamically based on load
High-Speed Processing	High-speed commands (Unicage) and distributed processing Can be used for front-end processing for legacy DB apps and Hadoop
Inexpensive Hardware	Runs on entry-class servers
Maintenance Support	Can be supported by anyone savvy in OS operations and management
Stability	Only uses the features of UNIX that have been stable for 40 years

Customers of Unicage



Many enterprises have adopted Unicage for the following reasons:

1. Fast learning curve for programmers (1 month training + 3 months OTJ)
2. Data Structure is transparent (Structured Text Files)
3. Programs are very short (95% shorter than traditional DB middleware)
4. Runs under latest UNIX/Linux OS to ensure security



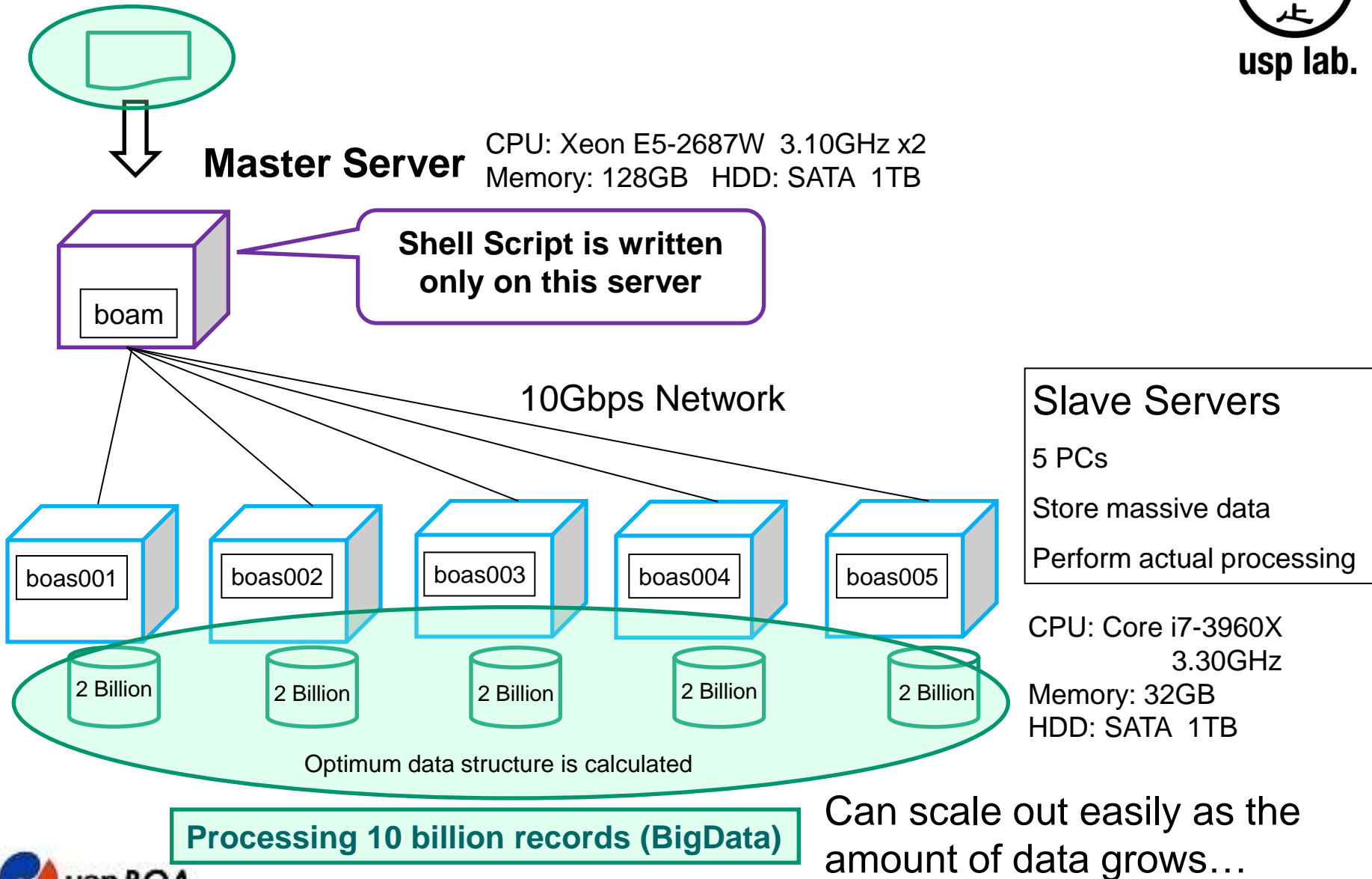
Big Data Processing Examples



usp lab.

<p>1 Batch Processing (Leading Credit Card Company)</p>	<p>Processing of daily transaction details on 60,000,000 credit card accounts OLD: COBOL Program running on Large Server (15hrs. 29mins) ↓ NEW: UNICAGE Program running on 5 PCs (1hr. 56mins)</p>
<p>2 Complex ETL (Leading Investment Bank)</p>	<p>Data Creation for DB Loading of 30,000,000 daily transaction records OLD: JAVA + PostgreSQL (90 minutes) ↓ NEW: Unicage Program running on 1 PC (91.58 seconds)</p>
<p>3 Complex ETL (Large Electric Utility)</p>	<p>Preprocessing of 10GB of Smart Meter data OLD: JAVA on HPUX Itanium 1.6GHz/2Core (15 hours) ↓ NEW: Unicage Program running on 1 PC (FreeBSD 9.1) (4 mins 16 secs)</p>
<p>4 Large Data Search (Biggest Search Engine in Korea)</p>	<p>50.3 Billion Log Records from 5 years (19.2TB) 10 Types of SQL Searches translated to Unicage Search Time: 0.227 sec - 4.763 sec</p>

Environment and Architecture



Benchmark Test: 10 billion records



usp lab.

Process	Description	Speed
1. Big Data Select (apli-select)	Perform a matching select on 10,000 transactions (join and exclude) from among 10 billion records distributed across the slave servers	4.5 secs.
2. Big Data Update (Add & Change, Delete, Sum) apli-update apli-delete apli-sumup	Update (add & change), delete and sum 10,000 transactions from among 10 billion records distributed across the slave servers	5.5 secs.
3. Big Data Search (apli-search)	Search account holder data based on Rank, Gender, Geographical Region, Age Group, Length of Membership and Minimum Average Score from among 10 billion records distributed across the slave servers	1.2 secs.

Thank You!

Contact in USA:

Jim Ryan

jim@farpointventures.com

617.821.1686