Package ‘usl’

January 9, 2020

Type Package
Title Analyze System Scalability with the Universal Scalability Law
Version 2.0.0
Date 2020-01-08

BugReports https://github.com/smoeding/usl/issues
Depends R (&gt;= 3.0), methods
Imports graphics, stats, nlsr
Suggests knitr
VignetteBuilder knitr

Description The Universal Scalability Law (Gunther 2007)
&lt;doi:10.1007/978-3-540-31010-5&gt; is a model to predict hardware and
software scalability. It uses system capacity as a function of load to
forecast the scalability for the system.

License BSD_2_clause + file LICENSE
Encoding UTF-8
RoxygenNote 7.0.2
NeedsCompilation no
Author Neil J. Gunther [aut],
Stefan Moeding [aut, cre]
Maintainer Stefan Moeding &lt;stm@moeding.net&gt;
Repository CRAN
Date/Publication 2020-01-09 19:10:02 UTC

R topics documented:

usl-package ................................................................. 2
confint,USL-method ......................................................... 3
efficiency,USL-method ...................................................... 4
oracledb ................................................................. 5
overhead,USL-method ...................................................... 6
Analyze system scalability with the Universal Scalability Law

Description

The Universal Scalability Law is a model to predict hardware and software scalability. It uses system capacity as a function of load to forecast the scalability for the system.

Details

Use the function `usl` to create a model from a formula and a data frame.

The USL model produces two coefficients as result: `alpha` models the contention and `beta` the coherency delay of the system.

The Universal Scalability Law has been created by Dr. Neil J. Gunther.

References


See Also

`usl`
Description

Estimate confidence intervals for one or more parameters in a USL model. The intervals are calculated from the parameter standard error using the Student t distribution at the given level.

Usage

```r
## S4 method for signature 'USL'
confint(object, parm, level = 0.95)
```

Arguments

- `object`: A USL object.
- `parm`: A specification of which parameters are to be given confidence intervals, either a vector of numbers or a vector of names. If missing, all parameters are considered.
- `level`: The confidence level required.

Details

Bootstrapping is no longer used to estimate confidence intervals.

Value

A matrix (or vector) with columns giving lower and upper confidence limits for each parameter. These will be labelled as $(1-level)/2$ and $1 - (1-level)/2$ in % (by default 2.5% and 97.5%).

See Also

- `usl`

Examples

```r
require(usl)
data(specsdm91)

## Create USL model
usl.model <- usl(throughput ~ load, specsdm91)

## Print confidence intervals
confint(usl.model)
```
efficiency, USL-method  Efficiency of the system

Description

The efficiency of a system expressed in terms of the deviation from linear scalability.

Usage

## S4 method for signature 'USL'
efficiency(object)

Arguments

object A USL object.

Details

The function returns a vector which contains the deviation from linearity for every measurement of the model input. A value of 1 indicates linear scalability while values less than 1 correspond to the fraction of the measurement compared to linear scalability.

Value

A vector of numeric values.

References


See Also

usl

Examples

require(usl)
data(raytracer)

## Show the efficiency
efficiency(usl(throughput ~ processors, raytracer))
Description

A dataset containing performance data for an Oracle OLTP database measured between 8:00am and 8:00pm on January, 19th 2012. The measurements were recorded for two minute intervals during this time and a timestamp indicates the end of the measurement interval. The performance metrics were taken from the v$sysmetric family of system performance views.

Format

A data frame with 360 rows on 8 variables

Details

The Oracle database was running on a 4-way server.

The data frame contains different types of measurements:

- Variables of the "time" type are expressed in seconds per second.
- Variables of the "rate" type are expressed in events per second.
- Variables of the "util" type are expressed as a percentage.

The data frame contains the following variables:

- `timestamp` The end of the two minute interval for which the remaining variables contain the measurements.
- `db_time` The time spent inside the database either working on a CPU or waiting (I/O, locks, buffer waits ...). This time is expressed as seconds per second, so two sessions working for exactly one second each will contribute a total of two seconds per second of `db_time`. In Oracle this value is also known as Average Active Sessions (AAS).
- `cpu_time` The CPU time used during the interval. This is also expressed as seconds per second. A 4-way machine has a theoretical capacity of four CPU seconds per second.
- `call_rate` The number of user calls (logins, parses, or execute calls) per second.
- `exec_rate` The number of statement executions per second.
- `lio_rate` The number of logical I/Os per second. A logical I/O is the Oracle term for a cache hit in the database buffer cache. This metric does not indicate if an additional physical I/O was necessary to load the buffer from disk.
- `txn_rate` The number of database transactions per second.
- `cpu_util` The CPU utilization of the database server in percent. This was also measured from within the database.
Description

overhead calculates the overhead in processing time for a system modeled with the Universal Scalability Law. It evaluates the regression function in the frame `newdata` (which defaults to `model.frame(object)`). The result contains the ideal processing time and the additional overhead caused by contention and coherency delays.

Usage

```r
## S4 method for signature 'USL'
overhead(object, newdata)
```

Arguments

- `object`: A USL model object for which the overhead will be calculated.
- `newdata`: An optional data frame in which to look for variables with which to calculate the overhead. If omitted, the fitted values are used.

Details

The calculated processing times are given as percentages of a non-parallelized workload. So for a non-parallelized workload the ideal processing time will always be given as 100% while the overhead for contention and coherency will always be zero.

Doubling the capacity will cut the ideal processing time in half but increase the overhead percentages. The increase of the overhead depends on the values of the parameters `alpha` and `beta` estimated by `usl`.

The calculation is based on A General Theory of Computational Scalability Based on Rational Functions, equation 26.

Value

overhead produces a matrix of overhead percentages based on a non-parallelized workload. The column `ideal` contains the ideal percentage of execution time. The columns `contention` and `coherency` give the additional overhead percentage caused by the respective effects.

References


See Also

usl.USL-class

Examples

require(usl)
data(raytracer)

## Print overhead in processing time for demo dataset
overhead(usl(throughput ~ processors, raytracer))
References


See Also

usl, scalability, USL-method

Examples

require(usl)
data(raytracer)

peak.scalability(usl(throughput ~ processors, raytracer))
## Peak scalability will be reached just below 450 processors

plot, USL-method

Plot the scalability function from a USL model

Description

Create a line plot for the scalability function of a Universal Scalability Law model.

Usage

## S4 method for signature 'USL'
plot(
x, 
from = NULL, 
to = NULL, 
xlab = NULL, 
ylab = NULL, 
bounds = FALSE, 
alpha, 
beta, 
... )

Arguments

x The USL object to plot.
from The start of the range over which the scalability function will be plotted.
to The end of the range over which the scalability function will be plotted.
xlab A title for the x axis: see title.
A title for the y axis: see title.
Add the bounds of scalability to the plot.
Optional parameter to be used for evaluation instead of the parameter computed for the model.
Optional parameter to be used for evaluation instead of the parameter computed for the model.
Other graphical parameters passed to plot (see par, plot.function).

plot creates a plot of the scalability function for the model represented by the argument x.
If from is not specified then the range starts at the minimum value given to define the model. An unspecified value for to will lead to plot ending at the maximum value from the model. For add = TRUE the defaults are taken from the limits of the previous plot.
xlab and ylab can be used to set the axis titles. The defaults are the names of the regressor and response variables used in the model.
If the parameter bounds is set to TRUE then the plot also shows dotted lines for the theoretical bounds of scalability. These are the linear scalability for small loads and Amdahl’s asymptote for the limit of scalability as load approaches infinity.
The parameters alpha or beta are useful to do a what-if analysis. Setting these parameters over- ride the model parameters and show how the system would behave with a different contention or coherency delay parameter.

See Also
usl.plot.function

Examples
require(usl)
data(raytracer)

## Plot result from USL model for demo dataset
plot(usl(throughput ~ processors, raytracer))
Usage

```r
## S4 method for signature 'USL'
predict(
  object,
  newdata,
  alpha,
  beta,
  interval = c("none", "confidence"),
  level = 0.95
)
```

Arguments

- `object`: A USL model object for which prediction is desired.
- `newdata`: An optional data frame in which to look for variables with which to predict. If omitted, the fitted values are used.
- `alpha`: Optional parameter to be used for evaluation instead of the parameter computed for the model.
- `beta`: Optional parameter to be used for evaluation instead of the parameter computed for the model.
- `interval`: Type of interval calculation. Default is to calculate no confidence interval.
- `level`: Confidence level. Default is 0.95.

Details

The parameters `alpha` or `beta` are useful to do a what-if analysis. Setting these parameters override the model parameters and show how the system would behave with a different contention or coherency delay parameter.

`predict` internally uses the function returned by `scalability,USL-method` to calculate the result.

Value

`predict` produces a vector of predictions or a matrix of predictions and bounds with column names `fit`, `lwr`, and `upr` if `interval` is set to "confidence".

References


See Also

`usl, scalability,USL-method, USL-class`
**Examples**

```r
require(usl)

data(raytracer)

## Print predicted result from USL model for demo dataset
predict(usl(throughput ~ processors, raytracer))

## The same prediction with confidence intervals at the 99% level
predict(usl(throughput ~ processors, raytracer),
        interval = "confidence", level = 0.99)
```

**Description**

`print` prints its argument and returns it invisibly (via `invisible(x)`).

**Usage**

```r
## S4 method for signature 'USL'
print(x, digits = max(3L, getOption("digits") - 3L), ...)
```

**Arguments**

- `x`: An object from class `USL`.
- `digits`: Minimal number of significant digits, see `print.default`.
- `...`: Other arguments passed to other methods.

**Value**

`print` returns the object `x` invisibly.

**See Also**

`usl.USL-class`

**Examples**

```r
require(usl)

data(raytracer)

## Print result from USL model for demo dataset
print(usl(throughput ~ processors, raytracer))
```
**raytracer**

*Performance of a ray-tracing software on different hardware configurations*

**Description**

A dataset containing performance data for a ray-tracing benchmark.

**Format**

A data frame with 11 rows on 2 variables

**Details**

The benchmark measured the number of ray-geometry intersections per second. The data was gathered on an SGI Origin 2000 with 64 R12000 processors running at 300 MHz.

The data frame contains the following variables:

- **processors** The number of CPUs used for the benchmark (1–64).
- **throughput** The number of operations per second.

**Source**


---

**scalability,USL-method**

*Scalability function of a USL model*

**Description**

`scalability` is a higher order function and returns a function to calculate the scalability for the specific USL model.

**Usage**

```r
## S4 method for signature 'USL'
scaleability(object, alpha, beta)
```


Arguments

- **object**
  A USL object.

- **alpha**
  Optional parameter to be used for evaluation instead of the parameter computed for the model.

- **beta**
  Optional parameter to be used for evaluation instead of the parameter computed for the model.

Details

The returned function can be used to calculate specific values once the model for a system has been created.

The parameters alpha or beta are useful to do a what-if analysis. Setting these parameters override the model parameters and show how the system would behave with a different contention or coherency delay parameter.

Value

A function with parameter x that calculates the scalability value of the specific model.

References


See Also

usl, peak.scalability, USL-method

Examples

```r
require(usl)
data(raytracer)

# Compute the scalability function
s <- scalability(usl(throughput ~ processors, raytracer))

# Print scalability for 32 CPUs for the demo dataset
print(s(32))

# Plot scalability for the range from 1 to 64 CPUs
plot(s, from=1, to=64)
```
**show, USL-method**  
*Show objects of class "USL"*

---

**Description**  
Display the object by printing it.

**Usage**  
```
## S4 method for signature 'USL'
show(object)
```

**Arguments**  
- `object`  
The object to be printed.

**Value**  
`show` returns an invisible `NULL`.

**See Also**  
- `usl, USL-class`

**Examples**  
```
require(usl)
data(raytracer)

## Show USL model
show(usl(throughput ~ processors, raytracer))
```

---

**specsdm91**  
*Performance of a Sun SPARCcenter 2000 in the SPEC SDM91 benchmark*

---

**Description**  
A dataset containing performance data for a Sun SPARCcenter 2000 (16 CPUs)

**Format**  
A data frame with 7 rows on 2 variables
Details

A Sun SPARCcenter 2000 with 16 CPUs was used for the SPEC SDM91 benchmark in October 1994. The benchmark simulates a number of users working on the UNIX server and measures the number of script executions per hour.

The data frame contains the following variables:

- load The number of simulated users (1–216).
- throughput The achieved throughput in scripts per hour.

Source


summary.USL-method

USL Object Summary

Description

summary method for class "USL".

Usage

## S4 method for signature 'USL'
summary(object, ...)

Arguments

object A USL object.
...
Other arguments passed to other methods.

See Also

usl.USL-class

Examples

require(usl)
data(raytracer)

## Show summary for demo dataset
summary(usl(throughput ~ processors, raytracer))

## Extract model coefficients
summary(usl(throughput ~ processors, raytracer))$coefficients
Create a model for the Universal Scalability Law

Description

usl is used to create a model for the Universal Scalability Law.

Usage

usl(formula, data, method = "default")

Arguments

- **formula**: An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be analyzed. The details of model specification are given under 'Details'.
- **data**: A data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which usl is called.
- **method**: Character value specifying the method to use. The possible values are described under 'Details'.

Details

The Universal Scalability Law is used to forecast the scalability of either a hardware or a software system.

The USL model works with one independent variable (e.g. virtual users, processes, threads, ...) and one dependent variable (e.g. throughput, ...). Therefore the model formula must be in the simple "response ~ predictor" format.

The model produces two main coefficients as result: alpha models the contention and beta the coherency delay of the system. The third coefficient gamma estimates the value of the dependent variable (e.g. throughput) for the single user/process/thread case. It therefore corresponds to the scale factor calculated in previous versions of the usl package.

The function coef extracts the coefficients from the model object.

The argument method selects which solver is used to solve the model:

- "nls" for a nonlinear regression model. This method estimates all coefficients alpha, beta and gamma. The R base function nls with the "port" algorithm is used internally to solve the model. So all restrictions of the "port" algorithm apply.
- "nlxb" for a nonlinear regression model using the function nlxb from the nlsr package. This method also estimates all three coefficients. It is expected to be more robust than the nls method.
"default" for the default method using a transformation into a 2nd degree polynomial has been removed with the implementation of the model using three coefficients in the usl package 2.0.0. Calling the "default" method will internally dispatch to the "nlxb" solver instead.

The Universal Scalability Law can be expressed with following formula. \( C(N) \) predicts the relative capacity of the system for a given load \( N \):

\[
C(N) = \frac{\gamma N}{1 + \alpha(N - 1) + \beta(N - 1)}
\]

Value

An object of class USL.

References


See Also


Examples

require(usl)

data(raytracer)

## Create USL model for "throughput" by "processors"
usl.model <- usl(throughput ~ processors, raytracer)

## Show summary of model parameters
summary(usl.model)

## Show complete list of efficiency parameters
efficiency(usl.model)

## Extract coefficients for model
coef(usl.model)

## Calculate point of peak scalability
peak.scalability(usl.model)

## Plot original data and scalability function
plot(raytracer)
plot(usl.model, add=TRUE)
USL-class

Class "USL" for Universal Scalability Law models

Description

This class encapsulates the Universal Scalability Law. Use the function `usl` to create new objects from this class.

Slots

frame The model frame.
call The call used to create the model.
regr The name of the regressor variable.
resp The name of the response variable.
coefficients The coefficients alpha, beta and gamma of the model.
coef.std.err The standard errors for the coefficients alpha and beta.
coef.names A vector with the names of the coefficients.
fitted The fitted values of the model. This is a vector.
residuals The residuals of the model. This is a vector.
df.residual The degrees of freedom of the model.
r.squared Coefficient of determination of the model.
adj.r.squared Adjusted coefficient of determination.
efficiency The efficiency, e.g. speedup per processor.
na.action The na.action used by the model.

See Also

usl
Index

*Topic datasets
  oracledb, 5
  raytracer, 12
  specsdm91, 14

coeff, 16, 17
confint, USL-method, 3
df.residual, 17

efficiency (efficiency, USL-method), 4
efficiency, USL-method, 4

fitted, 17
formula, 16

invisible, 11

nls, 16
nlslr, 16
nlxb, 16

oracledb, 5
overhead (overhead, USL-method), 6
overhead, USL-method, 6

par, 9
peak.scalability
  (peak.scalability, USL-method), 7
peak.scalability, USL-method, 7
plot, USL-method, 8
plot.function, 9
predict, USL-method, 9
print, USL-method, 11
print.default, 11

raytracer, 12
residuals, 17

scalability (scalability, USL-method), 12

scalability, USL-method, 12
show, USL-method, 14
specsdm91, 14
summary, USL-method, 15
title, 8, 9

usl, 2–4, 6–11, 13–15, 16, 18
USL-class, 18
usl-package, 2