



# Data Analysis Documentation

The Data Analysis Tool is a html+javascript based utility used by administrators, tuning experts, support engineers and consultants to create professional graphs based various types of input data. The tool is fully automatic and significantly faster than Excel at creating scatterplots, bar charts and line plots together with histogram analysis for almost any column separated input data such as NMON (*CPU\_ALL, MEM, MEM\_NEW, PROC, LPAR, PAGE*), VMSTAT, SDFMON, TAANA, ABAPMETER, MHTML, NIPING and XLSX. DataAnalysis.HTML will run in all modern web browsers like Chrome, Edge, Firefox, Opera, Safari (*Windows / Linux / MacOS / iPadOS*).

SAP internal note [3169320 - Data Analysis](#)

SAP JAM <https://jam4.sapjam.com/blogs/show/bavz22yEdmzrd4PHAGU3Yh>

## Contents

1	SELECT INPUT DATA.....	2
2	REVIEW INPUT DATA.....	3
3	CUSTOMIZE GRAPHIC OUTPUT .....	5
4	ADDITIONAL OPTIONS.....	6
4.1	1 <sup>ST</sup> LINE - X-Axis.....	6
4.2	2 <sup>ND</sup> LINE - Y-Axis.....	7
4.3	3 <sup>RD</sup> LINE - 2 <sup>ND</sup> Y-Axis.....	9
4.4	4 <sup>TH</sup> LINE – HISTOGRAM AND COLORS.....	10
4.5	5 <sup>TH</sup> LINE - CHART OPTIONS.....	11
5	PLOTLY OPTIONS .....	14
6	HISTOGRAM .....	15
7	PERFORMANCE.....	16
8	MEMORY USAGE: .....	16
9	ERRORS / BUGS .....	16



## 1 Select Input Data

DA Data Analysis allows to open and concatenate multiple input files, it will detect automatically the column separator (*tab*, *semicolon*, *comma..*) - any character not included in 0..9, A..Z, a..z or *space* will be detected. To load data into the Data Analysis tool you have three options:

- Use the "Choose Files" button and select one or multiple files (*press Ctrl Key to toggle between file and folder selection - if folder is active DA will select all files in a folder*)
- Drag and drop one or multiple files over the "Choose files" button
- Copy & paste any column separated data into the input box
- In some cases when many large numbers with thousand delimiter are included a wrong column separator, typically , or . might be selected - in that case we need to specify the separator manually (*use S for space, T for tab or specify any other character //\.,;:-)*)
- Specify Decimal format (*point or comma - if numbers are too small the program cannot determine the correct format automatically*)

**DA**  
Data Analysis

Select Data file(s):  No file chosen      Column Separator:  (enter S for space; t for tab - all others are build in: |,;tab/)      Dec.Notation: 1,023,045.06  check to preview data

Copy any column separated data to the clipboard, then paste into the text area and press *Update* button or upload files with the *Choose Files* button

Time	IAS	Instance	Act.	WPs	Dia.WPs	RFC WPs	ICPU	Usr	ICPU Sys	ICPU Idle	Paging in!	Paging out!	Free Mem.	EM alloc.	EM attach.	EM global	Heap Memor.	Fri.	Paging Mem.	Diag.	Upd.	Eng.	Logins	Sessions		
21.10.2019 23:59:30	chaerp05_S4P_15		1	1	100	2	1	1	1	1	125,980,298	7,672	6,884	14,429,931	0	0	181	17	1	0	20					
23:59:30	chaerp06_S4P_16		2	1	75	2	1	95	0	0	0	131,088,048	6,936	6,888	14,381,498	0	0	48	8	0	1	11				
23:59:30	chaerp07_S4P_13		2	1	1	75	28	2	68	0	0	0	126,913,000	7,228	7,044	14,383,546	0	0	1,192	14	0	0	13			
23:59:30	chaerp08_S4P_19		1	1	1	75	3	1	94	0	0	0	0	0	0	0	3,039	11	0	0	11					
23:59:30	chaerp09_S4P_25		2	1	75	11	2	85	0	0	0	121,942,931	7,454	7,248	14,383,546	0	0	2,118	9	0	0	17				
23:59:30	chaerp10_S4P_26		1	1	75	11	1	87	0	0	0	0	0	0	0	0	4,118	9	0	0	19					
23:59:30	chaerp11_S4P_27		1	1	150	1	1	97	0	0	0	0	0	0	0	0	80	8	0	0	18					
23:59:30	chaerp12_S4P_28		4	1	1	75	27	2	69	0	0	0	122,769,540	9,212	9,084	14,384,570	4,575	0	0	1,938	2	0	0	16		
23:59:30	chaerp13_S4P_29		3	1	1	75	27	2	69	0	0	0	0	0	0	0	1,539	10	0	0	0					
23:59:30	chaerp14_S4P_30		1	1	75	2	1	95	0	0	0	0	0	0	0	0	16	9	0	0	14					
23:59:30	chaerp15_S4P_32		2	1	1	75	2	1	96	0	0	0	0	0	0	0	0	0	0	0	11					
23:59:30	chaerp16_S4P_33		2	1	2	75	3	1	94	0	0	0	0	0	0	0	0	0	0	0	11					
23:59:30	chaerp17_S4P_34		2	1	1	75	4	3	91	0	0	0	0	0	0	0	0	0	0	0	13					
23:59:30	chaerp18_S4P_35		1	1	150	2	1	96	0	0	0	0	0	0	0	0	0	0	0	0	41					
23:59:30	chaerp19_S4P_36		1	1	1	75	2	1	96	0	0	0	0	0	0	0	0	0	0	0	8					
23:59:30	chaerp20_S4P_37		4	1	1	75	10	2	87	0	0	0	0	0	0	0	7,944	1	0	0	16					
23:59:30	chaerp21_S4P_38		1	1	1	75	2	1	96	0	0	0	0	0	0	0	0	0	0	0	26					
23:59:30	chaerp23_S4P_40		3	1	1	75	11	3	84	0	0	0	0	0	0	0	2,394	5	0	0	16					
23:59:30	chaerp24_S4P_43		2	1	1	75	4	3	91	0	0	0	0	0	0	0	0	0	0	0	15					
23:59:30	chaerp25_S4P_45		1	1	1	75	2	1	95	0	0	0	0	0	0	0	0	0	0	0	15					
23:59:30	chaerp26_S4P_46		2	1	1	75	28	1	70	0	0	0	0	0	0	0	0	0	0	0	15					
23:59:30	chaerp27_S4P_47		1	1	1	75	2	1	96	0	0	0	0	0	0	0	0	0	0	0	14					
23:59:30	chgreccapp00_S4P_001		4	1	150	4	1	95	0	0	0	0	0	0	0	0	0	0	0	0	26					
23:59:30	chaerp05_S4P_15		1	1	145	2	1	95	0	0	0	0	0	0	0	0	0	0	0	0	21					

Data.Analysis.HTML Vers. 11.22 -  - need help? please contact [c.weyd@ao.com](mailto:c.weyd@ao.com)

\*\*\* Data.Analysis.HTML is based on open-source JavaScript charting library plotly.js v1.58.5 <https://plot.ly/javascript/> - (C)opyright 2012-2019, Plotly, Inc. Licensed under the MIT license ([view the source on GitHub](#)). \*\*\*

Paste content of SDFMON into text area – then use button Update

Author(s): Christoph Weyd  
Changed: 31.03.2022

2 of 17



## 2 Review Input Data

On the second screen one can review the data (*if required specify the date format*) and modify the column headers or go directly to the graphic display.

**DA**  
Data Analysis

**Graphic**

Review data, if required edit header description(s), then press Graphic button to display the histogram & scatterplot - Select Date Format:   - Use:

File	Index	Time	AS Instance	Act. WPs	Dia.WPs	RFC WPs	CPU Usr	CPU Sys	CPU Idle	Paging in	Paging out	Free Mem.	EM alloc.	EM attach.	EM global	Heap Memor.	Pr.	Pagi								
F00000	I00000	23:59:30	chaerp05_S4P_15	1	1	150	2	1	96	0	0	28051768	7148	6884	4416622	0	0									
F00000	I00001	23:59:30	chaerp06_S4P_16	1	1	75	11	2	85	0	0	3.74E-5	6984	6884	4381498	0	0									
F00000	I00002	23:59:30	chaerp07_S4P_13	5	4	75	39	3	57	0	0	2.4E-5	7152	7660	4382522	0	0									
F00000	I00003	23:59:30	chaerp08_S4P_13	3	2	75	2	1	96	0	0	2991388	7336	6884	4382520	0	0									
F00000	I00004	23:59:30	chaerp09_S4P_25	3	3	8	75	2	1	96	0	0	3725556	7246	6892	4385594	0	0								
F00000	I00005	23:59:30	chaerp10_S4P_26	5	5	75	2	1	96	0	0	38055984	7244	6908	4384579	0	0									
F00000	I00006	23:59:30	chaerp11_S4P_27	11	1	150	1	1	96	0	0	37169716	7376	6884	4484296	0	0									
F00000	I00007	23:59:30	chaerp12_S4P_28	6	2	75	59	3	36	0	0	31988092	8812	7548	4384579	0	0									
F00000	I00008	23:59:30	chaerp13_S4P_29	1	1	75	2	1	96	0	0	30824944	7268	6884	4383546	0	0									
F00000	I00009	23:59:30	chaerp14_S4P_30	3	3	75	2	1	96	0	0	33411476	7344	6892	4383546	0	0									
F00000	I00010	23:59:30	chaerp15_S4P_32	5	3	75	46	1	52	0	0	22346968	11328	18716	4387643	5873	0									
F00000	I00011	23:59:30	chaerp16_S4P_33	3	3	75	2	1	96	0	0	33844860	7336	6892	4385596	0	0									
F00000	I00012	23:59:30	chaerp17_S4P_34	3	3	75	2	1	96	0	0	30563212	7368	6892	4386619	0	0									
F00000	I00013	23:59:30	chaerp18_S4P_35	1	1	150	2	1	96	0	0	28822224	7180	6884	4414574	0	0									
F00000	I00014	23:59:30	chaerp19_S4P_36	1	1	75	2	1	96	0	0	39350860	6976	6884	4386619	0	0									
F00000	I00015	23:59:30	chaerp20_S4P_37	4	4	75	2	1	96	0	0	32553724	7388	6896	4381498	0	0									
F00000	I00016	23:59:30	chaerp21_S4P_38	1	1	150	2	1	96	0	0	27555648	7872	6884	4481261	0	0									
F00000	I00017	23:59:30	chaerp23_S4P_40	1	1	75	2	1	95	0	0	31747164	7448	6884	4381498	0	0									
F00000	I00018	23:59:30	chaerp26_S4P_43	3	3	75	2	1	96	0	0	33426192	7284	6892	4382522	0	0									
F00000	I00019	23:59:30	chaerp27_S4P_44	1	1	75	2	1	96	0	0	34225400	7424	6884	4384295	0	0									
F00000	I00020	23:59:30	chaerp28_S4P_46	2	1	75	29	1	69	0	0	43208008	8956	8956	4532026	0	0									
F00000	I00021	23:59:30	chaerp30_S4P_47	2	1	75	11	2	85	0	0	34980960	7858	7000	4387643	0	0									
F00000	I00022	23:59:30	chaerpccapp00_S4P_00	4	4	150	1	0	99	0	0	54299520	20680	19120	4989188	0	0									
F00000	I00023	23:59:00	chaerp05_S4P_15	2	2	150	1	1	96	0	0	28016132	7148	6888	4416622	0	0									
F00000	I00024	23:59:00	chaerp06_S4P_16	1	1	75	11	2	85	0	0	37382896	6984	6884	4381498	0	0									
F00000	I00025	23:59:00	chaerp07_S4P_13	6	4	75	39	3	57	0	0	34263916	7300	7200	4382522	0	0									
+++ type / number of unique values +++																										
- / 1 - / 66533		t / 2891		- / 24		n / 70		n / 71		n / 32		n / 100		n / 19		n / 101		n / 1	n / 1	n / 32908	n / 4471	n / 2408	n / 418	n / 1822	n / 5	n
<input type="button" value="PgD ▼"/> <input type="button" value="PgU ▲"/> <input type="text" value="find:"/> <input type="text" value="replace with:"/> <input type="button" value="Find"/> <input type="button" value="Replace"/> <input type="button" value="Sort"/> <input type="button" value="Delete Values"/> <input type="button" value="Split Column"/> <input type="button" value="Delete Column"/>																										
<input type="button" value="Download Data"/> <input type="button" value="Remove Duplicates"/> <input type="button" value="Delete empty columns"/> <input type="button" value="Delete every 2nd Row"/>																										
Number of Records: 66245 - Records: 66245 (total) / 18400 (selected) / 18400 (display) - Runtime: 1444.00 (131.50) ms																										

The last line in the table is indicating what type of data has been detected and how many distinct values exist in each column - the format is `<type> / <number>`

F00000	undefined	I66163	00:06:00	chaerp15_S4P_32	4	2	75
F00000	undefined	I66164	00:06:00	chaerp16_S4P_33	5	1	75
+++ type / number of unique values +++							
- / 1	- / 1	- / 66453	t / 2897	- / 24	n / 99	n / 96	n / 51

The `<type>` indicates:

- text value
- d date

Author(s):	Christoph Weyd			
Changed:	31.03.2022			3 of 17



t time  
n numeric

The user has some options to manipulate the input data (*the buttons are only available if the user is putting the cursor into the header line of a column*)

- Find and/or Replace values
- Sort Data
- Delete Values
- Split or Delete Columns
- Perform basic calculations like  
= <value> | \$nn [ # [<value> | \$nn]] or  
# <value> | \$nn  
where # can be any of the operators +-\*/  
\$nn refers to the content of column nn

In some occasions the program might not be able to determine the date format automatically, in that case the user can choose from different date formats

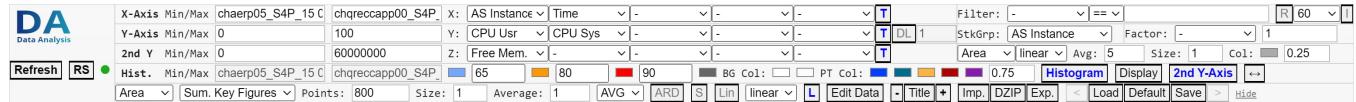
YYYY-MM-DD
DD-MM-YYYY
YY-MM-DD
DD-MM-YY
YYYY-DD-MM
MM-DD-YYYY
YY-DD-MM
MM-DD-YY

Additionally, the user can download the data (tab separated text file with extension .xls). To reduce the amount of data, columns which do not contain significant information (all values identical) can be removed. If files are too big then the user can also remove every 2<sup>nd</sup> data row.



## 3 Customize Graphic Output

Below the various options to customize the graphic output



1. Specify the fields for the X-Axis. Available are non-numeric fields (*date, time, server names...*) and all numerical fields
2. Specify the fields for the Y-Axis. Available are only numerical fields (*integer or float*)

*Important: Certain fields for the X-Axis like Date/Time are automatically proposed - for the Y Axis the DataAnalysis will default CPU User, CPU System, Resp.Time...*

Optional

3. Activate 2nd Y-Axis and specify fields, chart type...
4. Select the chart type - available options are:
  - Scatter Plot
  - Line Chart
  - Bar Chart
  - Area Chart
5. If multiple key figures are select for the Y-Axis specify if those values should be summarized, stacked over each other, or shown independently
  - Summarize Key Figures
  - Stack Key Figures
  - Independent Key Figures
6. Select if any key figures should be used to group/stack the data. Each different value for the selected Stack variable will be shown in a different color. One can specify if only the first n values should be considered. A graph can only be stacked with this option if for each x-value there are an equal number of y-values available. If any x-value will have a different number of y-values this x-value will not be shown (*this might happen if x-values (time) differ by a few seconds - as workaround one can try to activate rounding*)



## 4 Additional Options

In the upper left you see the actual status of the Chart and you can Restart, Refresh or Reset the Graph

- **Restart**  
If you click on the Logo itself you can restart DA Data Analysis to load another data source.
- **Recalculate and Refresh the Graphic**  
**Refresh**
- **Reset Graphic and recalculate Axis scaling**  
**RS**
- **Status**  
Ready  Refresh Pending  Busy 

### 4.1 1st line - X-Axis



- **X-Axis Min/Max – Specify the X-Axis Range**  
**X-Axis Min/Max**
- **Select fields for X-Axis (Show/Hide Data Labels)**  
**X:**
- **Show/Hide X-Axis Labels (blue = Show, gray = hide)**  
 
- **Filter by selected x-value (== x-values equal selected values; <> x-values not equal selected values; multiple conditions can be combined with ||, wildcard \* allowed)**  
**Filter:**



- Round time figures (select R to activate) to multiple of 1, 5, 10, 15 ... 60 ... 3600 seconds

Available Rounding Values are (in seconds)

1
5
10
15
20
30
60
120
300
600
900
1200
1800
3600

- Specify if data points not present in all stack groups should be added or eliminated

- I              Ignore data points which are not present in all stack groups  
 A              Add missing data points with zero value  
 E              Erase Data point from all stack groups if not present in all of them

## 4.2 2<sup>nd</sup> line - Y-Axis

Y-Axis Min/Max	0	100	Y:	CPU Usr	▼	CPU Sys	▼	-	▼	-	▼	T	DL	1	StkGrp:	AS Instance	▼	Factor:	-	▼	1
----------------	---	-----	----	---------	---	---------	---	---	---	---	---	---	----	---	---------	-------------	---	---------	---	---	---

- Y-Axis Min/Max – Specify the Y-Axis Range

Y-Axis Min/Max	0	50
----------------	---	----

- Select fields for Y-Axis (Show/Hide Data Labels)

Y:	CPU User%	▼	CPU Sys%	▼	-	▼	-	▼	-
----	-----------	---	----------	---	---	---	---	---	---

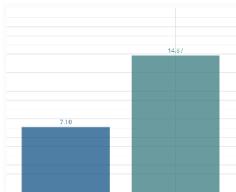
- Show/Hide Y-Axis Labels (blue = Show, gray = hide)



- Data Labels (Bar Charts only)

**DL** 2

Show data Labels above Bars with n decimals



- Stack Group

StkGrp: **Host**

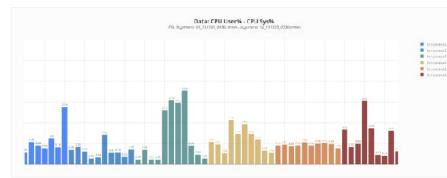
Data Points can be grouped by non-numeric fields as a separate data set. Each Stack Group will show in a separate color which will be interpolated from the four colors defined in the 4<sup>th</sup> line

Example

X-Axis: Host + Time

Y-Axis: CPU Usage; Stack Group: Host

Colors:



- Factor – we can multiply the Y-Values with the value of another column.

Factor: **LPAR logicalC**

A useful example is to multiply the CPU% with the total number of logical CPUs to display the logical CPUs used. We add here an additional factor of 0.01 to adjust for the percentage itself.

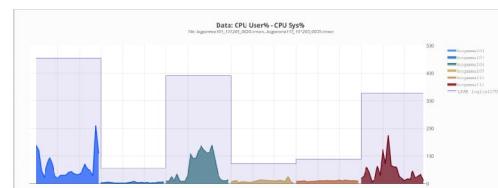
Example

X-Axis: Host + Time

Y-Axis: CPU Usage; Stack Group: Host

Factor: logical CPUs \* 0.01

2<sup>nd</sup> Y-Axis: logical CPUs





### 4.3 3<sup>rd</sup> line - 2<sup>nd</sup> Y-Axis

2nd Y Min/Max	0	60000000	z:	Free Mem.	-	-	-	-	T	Area	linear	Avg:	5	Size:	1	Col:	0.25
---------------	---	----------	----	-----------	---	---	---	---	---	------	--------	------	---	-------	---	------	------

- 2<sup>nd</sup> Y-Axis Min/Max – Specify the Y-Axis Range

2nd Y Min/Max	0	500
---------------	---	-----

- Select fields for 2<sup>nd</sup> Y-Axis (Show/Hide Data Labels)

z:	LPAR logicalC	-	-	-	-	-
----	---------------	---	---	---	---	---

- Show/Hide 2<sup>nd</sup> Y-Axis Labels (blue = Show, gray = hide)

T	T
---	---

- Details for 2<sup>nd</sup> Y-Axis

Area	steps	Avg:	1	Size:	1	Col:	0.1
------	-------	------	---	-------	---	------	-----

Here we specify how the 2<sup>nd</sup> Y-Axis is drawn. From left to right

- Chart Type (Area, Scatter, Line)
- How to connect Data Points (line, spline, steps)
- Average n data points
- Point/Line Size
- Point/Line Color
- Transparency (from 0.05 to 1.00)



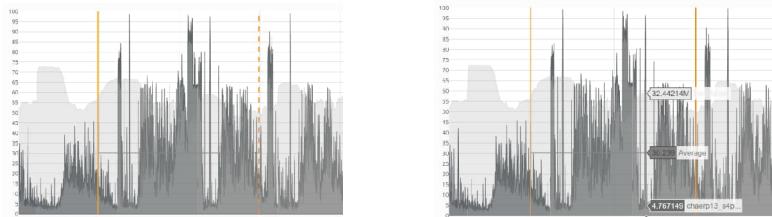
## 4.4 4<sup>th</sup> line – Histogram and Colors

Hist. Min/Max chaerp05\_S4P\_15 0 chqreccapp00\_S4P\_ 65 80 90 BG Col: PT Col: 0.75 Histogram Display 2nd Y-Axis ↔ 2D

- Histogram Range (Display only)

**Hist.** Min/Max chaerp05\_S4P\_15 0 chqreccapp00\_S4P\_

Single click on the Chart to set first the left border (the right border will show in a dashed orange line), then click again to set the left border.



- Specify data range and color for the histogram (3rd line)

65 80 90

- Select background and chart color

BG Col: □ □

- Specify color for the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> select Y-Value and Transparency. If a stack/group variable is used this for color will specify the color range used.

0.75

The transparency ranges from 0.05 (almost invisible) to 1.00 (opaque)

- Show/Hide the histogram

**Histogram**

- Display Data used in Chart

**Display**

- Enable/Disable 2<sup>nd</sup> Y-Axis

**2nd Y-Axis**

- Swap primary with 2nd Y-Axis

↔



- Switch to 3D Display

**2D**

## 4.5 5<sup>th</sup> line - Chart Options

Area Sum. Key Figures Points: 800 Size: 1 Average: 1 AVG ARD Lin linear L Edit Data - Title + Imp. DZIP Exp. < Load Default Save > Hide

- Select Chart Type - available Chart Types are:

Bar  
Scatter  
Line  
**Area**

- Handling of Data Points. If multiple key figures are select for the Y-Axis specify if those values should be summarized, stacked over each other or shown independently

**Sum. Key Figures**  
Stack Key Figures  
Indep. Key Figures

- Maximum number of points to be drawn - if the data contains after aggregation m points and only x point should be drawn then only every  $n^{th} = m/x$  points will be drawn

Points: 800

- Specify point size (*also line width in points, default = 3*)

Size: 1

- Select how many data points should be aggregated (*calculate average of n points*) Average: 1

- Specify calculation type if multiple data points are cumulated - default is AVG (average), available options are:

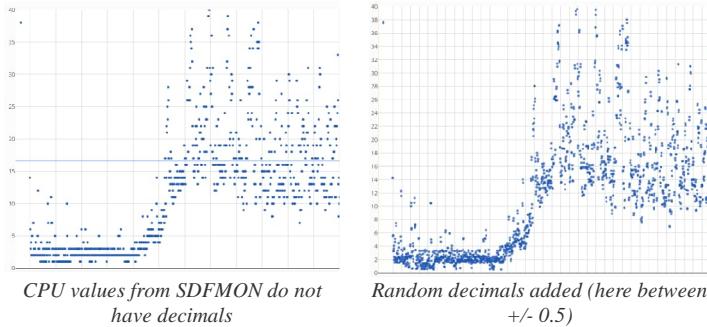
SUM  
**Avg**  
MAX  
MIN



- Add random decimals to prevent banding.

**ARD**

If the measurements like CPU usage is only available with limited decimals values the graph will look unnatural because all values are on discrete lines. With button S one can add to each key figure a random value  $r$  between  $-d/2 < r < d/2$  where  $d$  is the smallest detected difference between the y values. See below Example



- Reduce Number of Data Points (Algorithm)

**S** **T**

**S** Skip every n'th data point

**T** Largest-Triangle-Three-Buckets

(see sveinn-steinarsson/plot-downsample: Downsample plugin for Plot charts. (github.com))

- Switch between linear or logarithmic Y-Axis

**Log** **Lin**

- Specify how data points get connected in a line chart (linear, spline, or steps)

**linear**  
spline  
steps

- Display or Hide Chart Legend

**L**

- Edit/Preview Data

**Edit Data**

- Set Chart & Axis Title (*supported are the modifiers: <b> Bold </b>, <i> italic </i>, <sub> subscript </sub>, <sup> superscript </sup>, and line breaks <br>*) and increase or decrease the font size

**-** **Title** **+**



- Import/export Layout Profile or Export Chart

**[Imp.] [DZIP] [Exp.]**

The two buttons IMP. and Exp. allow to import or export a JSON which contains all chart settings. Those settings can be used when another data source is loaded.

The button DZIP is generating a compressed JSON file which contains the chart settings and the data itself. Typically the DZIP file is around 20% of the original file size.

- Load/Save Settings set defaults. Undo/Redo last changes.

**[<] [Load] [Default] [Save] [>]**

The buttons Load and Save import/export the current chart settings into memory. They can be applied to the next file loaded. The two buttons < > allow to undo or redo the last changes.

- Autohide the chart options (that will make the charts bigger if the cursor is moved over them, the menu is shown again if the cursor is moved to the top)

**Hide   Show**



## 5 Plotly Options

The below options are visible when you hover the mouse above the graph



- Zoom - use the mouse to select a rectangular area to be shown
- Move or Re-Scale axis (*move mouse to the end of the axis scales until an arrow is shown, then lets-click and move*)
- Download Graph as PNG file
- Autoscale / Reset Scale
- Export and open chart in Plotly Chart Studio
- Change scaling of X-Axis (double click on graph) – A popup allow to enter the tick distance (in hours, minutes, seconds...)

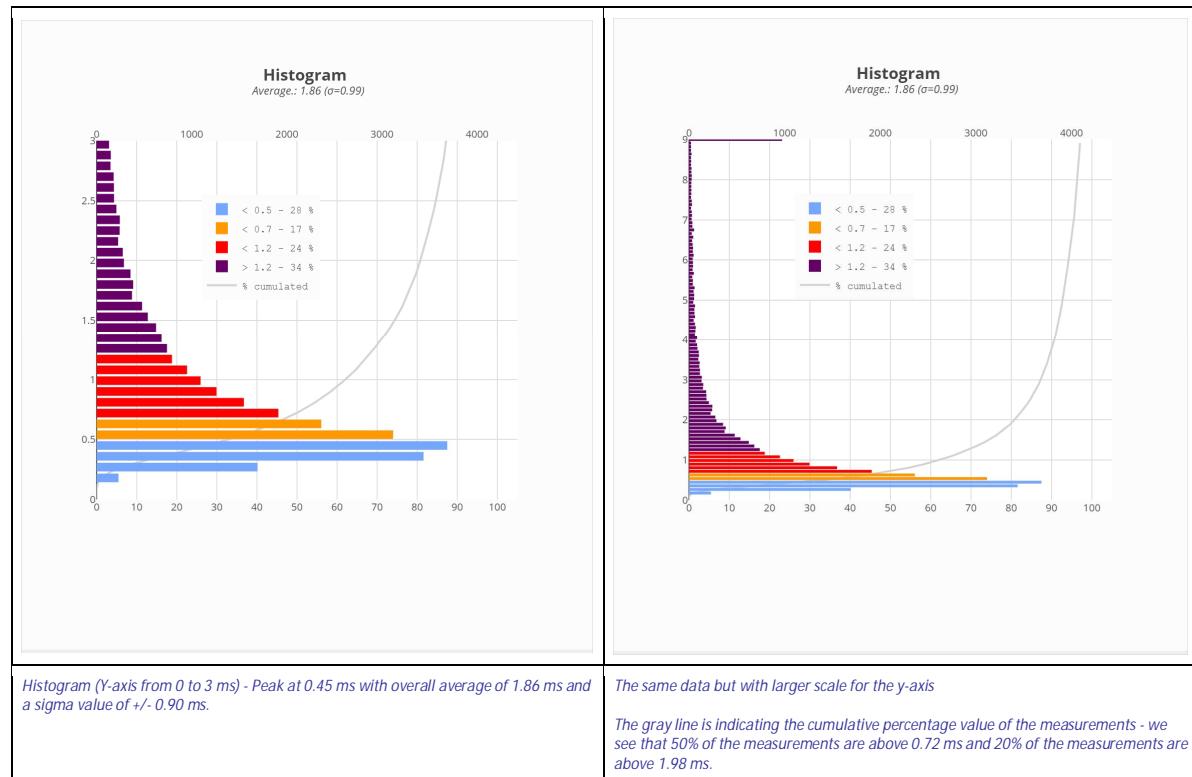


- If multiple values are shown, click on the legend to show/hide individual key figure or stack values



## 6 Histogram

The histogram is calculated for the specified y-axis range divided in 100 equal intervals. For each of those intervals we count how many measurements are recorded and display the result as a horizontal bar chart. The color coding can be specified in the top of the screen:



Especially for highly fluctuating data (*NIPING results, CPU usage, number of active work processes*) a line/bar or scatterplot alone are difficult to interpret and often lead to wrong conclusions. A quantifiable metric to describe the difference in fluctuations is not available on those charts.

*Only a histogram analysis allows to exactly describe the random data distribution. Especially for sizing purposes a statement like "95% (2 sigma) of all measurements are below 75 of CPU usage" is much more accurate than the standard deviation a line drawn above the highest observed values. Scatterplots combined with a histogram show the reality, how the individual measurements are fluctuating and at the same time to quantify the fluctuations, distribution and upper lower limits in percentage to the number of recorded measurements. Using maximum/minimum or averages do NOT show the reality and should NOT be used in the context of capacity analysis. Assuming the histogram follows a gauss normal distribution then we can conclude that:*

- In the interval of the deviation  $\pm\sigma$  of the expected value, 68.27 % of all measured values can be found,
- In the interval of the deviation  $\pm 2\sigma$  of the expected value 95.45 % of all measured values are found,



## 7 Performance

Depending on the file type, record length, number of columns the Data Analysis tool can load and parse approximately **140.000** records per second. The throughput of the graphic engine depends on the number of operations and fields but typically Data Analysis can calculate and draw around **15.000** measurement points per second. We can estimate the average total processing time  $T$  of the Data Analysis tool by  $T \text{ (ms)} = R/1000 * 7 + P/1000 * 60$  where  $R$  is the total number of records loaded,  $P$  the number of data points in the graphic.

Example(s):

- We process 5 NIPING files (*with 1 sec frequency*) which contain a total of **432.000 records** - from those records we plot **20.000 data points**. The estimated total processing time will be at  $432 * 5 + 20 * 60 = 4224 \text{ ms (4.2 seconds)}$ .
- A single NIPING file of 86.400 records with 10k data point will be ready in  $0.60 + 0.60 = 1.2$  seconds.
- We open 30 nmon files (*24h capture, snapshot every 5min, 57 CPU cores, each file with 150.000 records*) with the nmon analyzer to extract CPU\_ALL, PROC, LPAR, MEM, MEM\_NEW then each file takes about 1 minute of processing time. If we merge all files together with the nmon analyzer the process takes around 30 minutes. The Data Analysis tool can load and display the same data in less than 10 seconds.

## 8 Memory Usage:

The memory usage of Data Analysis tool depends on the file size and number of data points to be displayed. When we open a SDFMON file with 1.586.920 records the initial memory consumption reached 5GB of memory. During the creation of the graphic the total memory consumption of the browser (opera) increased to 11.5 GB of memory drawing all 1.586.920 data points (runtime 75 seconds). When we limit the number of data points to 200.000 the memory consumption reached only 2.1GB. The memory consumption scales with the number of data points drawn in the graphic and can be estimated at about 5-10 kB per data point.

## 9 Errors / Bugs

In case of any errors, please send an email to [c.weyd@sap.com](mailto:c.weyd@sap.com) and include a copy (*or link*) to the used data source and a screenshot of the browser window (*including the used settings in the top*). Please also describe which browser and operating system was used.

*Data Analysis is based on open-source JavaScript charting library [plotly.js](https://plot.ly/javascript/) <https://plot.ly/javascript/>  
(C)opyright 2012-2019, Plotly, Inc. Licensed under the MIT license ([view the source on GitHub](#).)*

