Subject: Problems ... again ...
Posted by rats on Sun, 18 Jul 2004 23:03:05 GMT
View Forum Message <> Reply to Message

Hi,
I am having some problems with a huge file that I have to manipulate ...
This file has more than 5 million rows by 4 columns ...
One of this columns is the ATTRIBUTE, and the other ones are: X, Y and Z (all values in meters).
First I have separated this file in 4 new ones ... each one for one attribute.
But the files are still huge.
I have to make some statistics with the points in these files, but the USER have to specify different SPACINGS ... so the statistic will exist to make some MEAN, MEDIAN, etc... with the points inside this spacing.

What I have done till now is:
For example: Attribute 3
1) I have created a NEW INDEX COLUMN in this file ... Beginning with value 1 and the last value will be the last line number.
2) Created 2 ARRAYS . One with the INDEX column, X column and Z column.
The other one with the INDEX column, Y column and Z column.
I have sorted this two arrays by the X and Y columns, so that this two columns begin with the lowerst value. I did that to find the same points inside the spacing (using some intersection with the two arrays by the INDEX COLUMN)
3) I am "moving" the "grid" by the spacing value that the user choose ... so if he chooses 2 meters for the spacing: I first FIX the X axis by two meters and then I "walk" thru the Y Axis (2 by 2 meters) and then I make the statistic for the "intersection points" inside the grid(2x2m) till it reaches the end. After that I move another 2 meters in the X axis and do the same thing for the Y axis ... I do that till the end of the X and Y axis.

I know that this method is too slow ... it works ... but it takes a looong time to terminate.

Any ideas ?

Thank you very much ...
Rafael