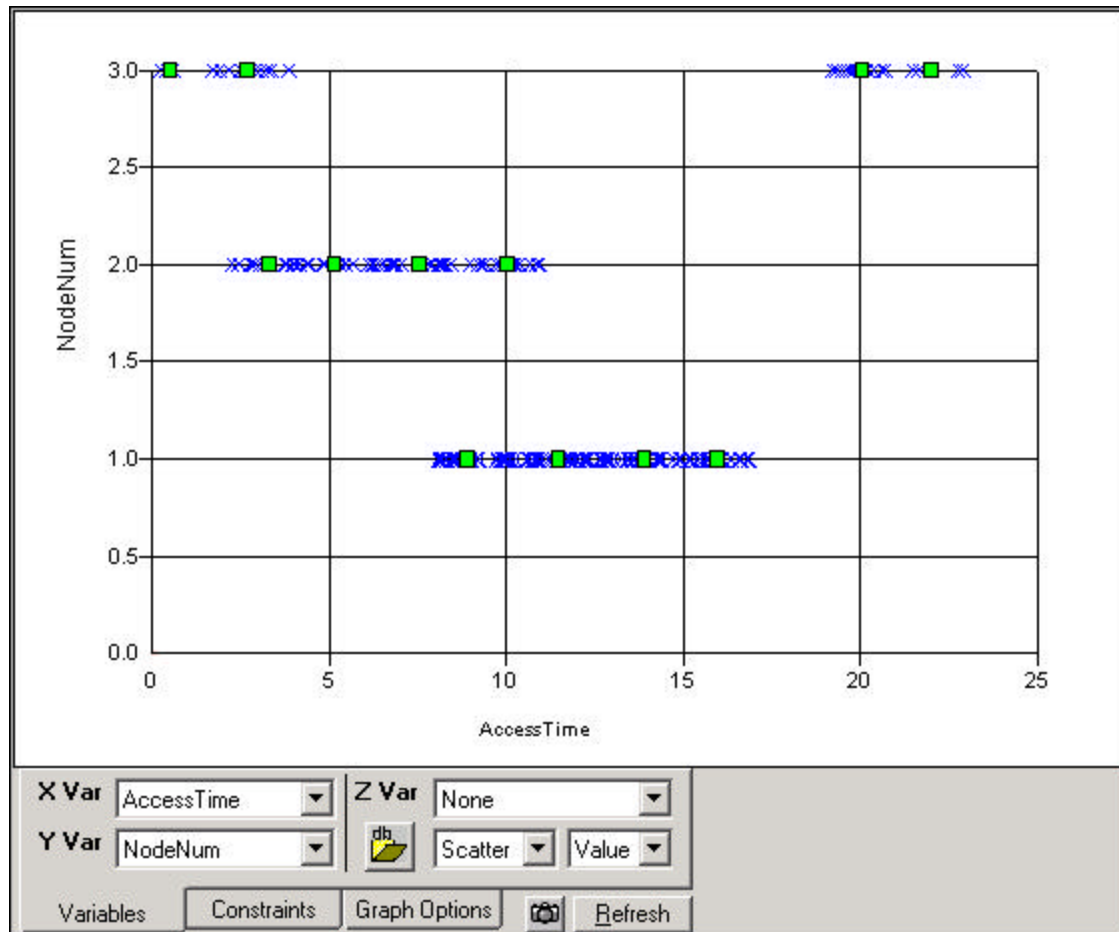


File Access Model Example

Variables:

- NodeNum represents different offices (i.e., 1, 2, 3 corresponds to New York, London, Singapore)
- AccessTime represents logged access time Eastern Std Time. 0 - 23 hrs



The green dots are clusters that were derived from "normal" operational data. In effect, this model has learned the standard operating hours (& thus "normal" file access times) of each office. In an implementation, if a new file access case does not map closely to one of the existing clusters, then it would be flagged as an exception.

Because there are only 2 variables, it would be relatively easy to write a set of SQL queries with hard constraints to characterize the relationship between NodeNum & AccessTime & catch those cases that fall outside predetermined bounds. The point of this example is that no one had to write those rules...they were automatically derived directly from the data.

When you add even a few more variables (i.e., user department, day of week, program dependency, type of access, etc.), you wind up having to write a lot more rules & they are subject to errors in specification and omission. However, a data driven model can automatically discover the relationships between a larger number of variables with relatively little additional effort.