IBM Research

# Using Model Trees to Characterize Computer Resource Usage

Steve Heisig
T. J. Watson Research Center
Steve Moyle
Oxford Comlab

Oct 31, 2004

# What are we (not) addressing?

- Monitoring – Using existing data sources
- **Interpretation – Human and Programmatic Models**
  - Adaptive thresholding for resource consumption
    - Processor, memory, IO bandwidth, paper, plastic
- Resolution - What to change
- Adaptation - How to change it

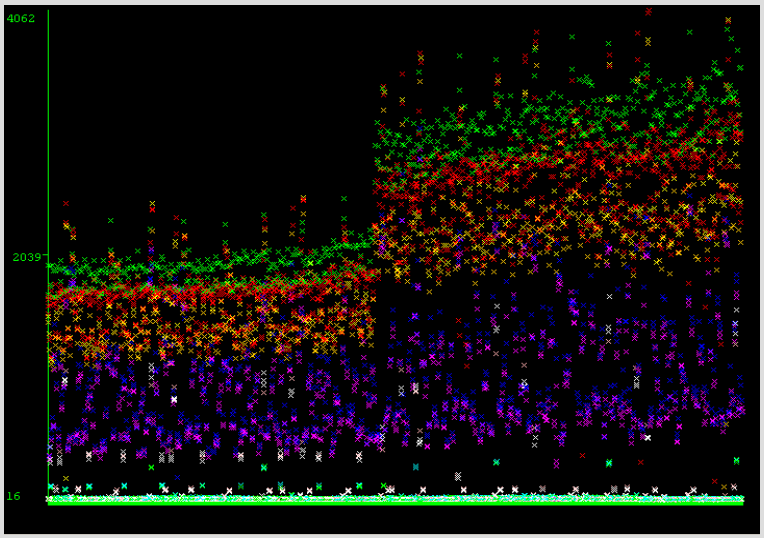

© 2004 IBM Corporation

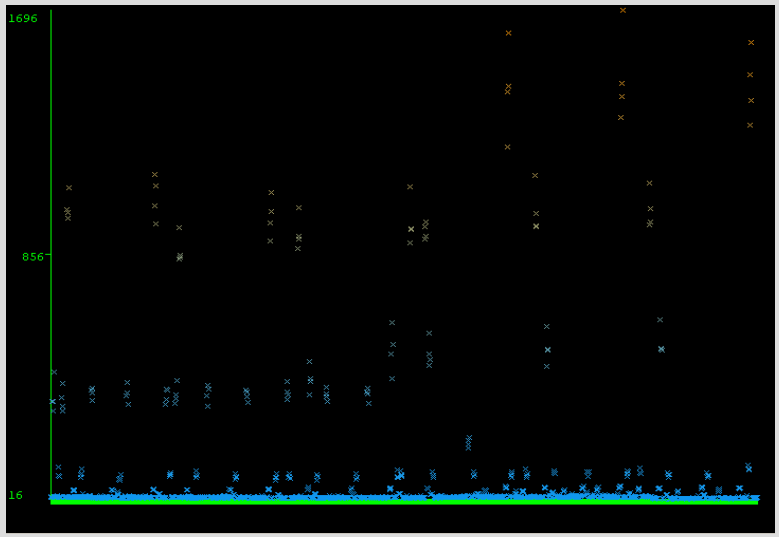
# What are the domains, properties, applications being targeted?

- Systems and workloads with good resource consumption instrumentation
  - Models are improved by ability and willingness of domain experts to provide additional business process information
  - This could potentially be automated if this information becomes available.


© 2004 IBM Corporation

# Processor Characterization




04072003 LOAN INTEREST (MONTHLY) (CALCULATION)
04072003 MONTHLY CHARGED COMMISSION (EXTRACTION)
04072003 MONTHLY NOT PRE-NOTIFIED DEBIT INTEREST (APPLICATION)
04072003 MONTHLY PRE-NOTIFIED DEBIT INTEREST (EXTRACTION)

07072003 LOAN INTEREST (MONTHLY) (APPLICATION)
07072003 MONTHLY CHARGED COMMISSION (APPLICATION)
07072003 MONTHLY PRE-NOTIFIED DEBIT INTEREST (APPLICATION)

11072003 MONTHLY CHARGED COMMISSION (CALCULATION)
11072003 MONTHLY PRE-NOTIFIED DEBIT INTEREST (CALCULATION)


© 2004 IBM Corporation

## Decision tree:

```
Month <= 0.5 :
|  Day <= 12.5 :
|  |  HY <= -7.45 :
|  |  |  HY <= -9.15 :
|  |  |  |  Day <= 1.5 : LM1
|  |  |  |  Day >  1.5 :
|  |  |  |  |  EuroCalc=Y <= 0.5 : LM2
|  |  |  |  |  EuroCalc=Y >  0.5 : LM3
|  |  |  HY >  -9.15 :
|  |  |  |  EuroCalc=Y <= 0.5 :
|  |  |  |  |  Day <= 7.5 :
|  |  |  |  |  |  Day <= 3.5 : LM4
|  |  |  |  |  |  Day >  3.5 : LM5
|  |  |  |  |  Day >  7.5 : LM6
|  |  |  |  EuroCalc=Y >  0.5 : LM7
|  |  HY >  -7.45 :
|  |  |  Day <= 9.5 :
|  |  |  |  Day <= 4.5 :
|  |  |  |  |  Day <= 3.5 : LM8
|  |  |  |  |  Day >  3.5 : LM9
|  |  |  |  Day >  4.5 : LM10
|  |  |  Day >  9.5 :
|  |  |  |  Day <= 10.5 : LM11
|  |  |  |  Day >  10.5 : LM12
|  Day >  12.5 :
|  |  Day <= 27.5 : LM13
|  |  Day >  27.5 :
|  |  |  Day <= 29.5 : LM14
|  |  |  Day >  29.5 :
|  |  |  |  Day <= 30.5 : LM15
|  |  |  |  Day >  30.5 : LM16
Day >  19.5 :
|  |  HX <= 2.45 :
|  |  |  Day <= 27.5 :
|  |  |  |  Day <= 21.5 :
```

## Linear Regression Leaf Nodes

Linear Regression Model 1
CPUSECS = 530.3333

Linear Regression Model 2
CPUSECS = 59.0417 * Day -333.5

Linear Regression Model 3
CPUSECS = 51.5

Linear Regression Model 4
CPUSECS = 21.5

Linear Regression Model 5
CPUSECS = 99.75

Linear Regression Model 6
CPUSECS = 374.6667

Linear Regression Model 7
CPUSECS = -4.4495 * Day + 69.8258


© 2004 IBM Corporation

# New ideas/approaches

- Model Trees are useful for distilling large amounts of performance/diagnostic data to compact, computationally efficient, human understandable rules.
- Model Trees allow human domain experts to augment raw performance data sources. The resulting models can be more useful than those built using unannotated data.
- Insights can be used for automation, human intervention, or OS optimization.


© 2004 IBM Corporation