vTimeSeries

Vladimir Adam, Karan Toor, Cesar Polanco, Ryan McGinley, Nick Cross
Data Center
VMware Infrastructure
VMware Infrastructure

- vCenter
  - Cluster 1
    - ESX Host 1
      - VM 1
      - VM 2
    - ESX Host 2
      - VM 3
      - VM 4
  - Cluster 2
    - ESX Host 3
      - VM 5
      - VM 6
Varying Workload
VMware's Current Solution

- DRS (Distributed Resource Scheduler)
- Overload detection based on immediate trend
VMware's Current Solution

- VM Migration to less stressed host
Problem

- VM migration is expensive!
- Does not take long term trends into account.

So what's our solution?
Our Solution

System Administrators Can

- Forecast hardware demands
- Discover correlations between statistics
Our Solution
Collector/Retriever

- Use vCenter API to get Data
- Data packed up every 115 seconds
Collector/Retriever

- Use vCenter API to get Data
- Data packed up every 115 seconds

Where does the data go?
Database

MySQL DB

- Created objects for both input and output
Database

MySQL DB

● Created objects for both input and output
● Using database created some conflicts
  ○ Naming conventions
  ○ Creating tables/Entities
Database

MySQL DB
- Created objects for both input and output
- Using database created some conflicts
  - Naming conventions
  - Creating tables/Entities

Where does the analysis take place?
Single Exponential Smoothing

\[ s_t = (1 - \alpha)^{t-1} x_0 + \alpha \sum_{i=1}^{t-1} (1 - \alpha)^{i-1} x_{t-i} \]
Double Exponential Smoothing

\[ s_t = \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i (x_{t-i} + b_{t-i-1}) \]

\[ b_t = \beta \sum_{i=0}^{t-2} (1 - \beta)^i (s_{t-i} - s_{t-i-1}) \]

\[ F_{t+m} = s_t + mb_t \]
Triple Exponential Smoothing

\[ s_t = \alpha \frac{x_t}{c_t - L} + \sum_{i=0}^{t-1} (1 - \alpha)^i (x_{t-i} + b_{t-i}) \]

\[ c_i = \frac{1}{N} \sum_{j=1}^{N} \frac{x_{L(j-1)+i}}{A_j} \]

\[ b_t = \beta \sum_{i=0}^{t-2} (1 - \beta)^i (s_{t-i} - s_{t-i-1}) \]

\[ A_j = \frac{1}{L} \sum_{i=1}^{L} x_{L(j-1)+i} \]

\[ c_t = \gamma \frac{x_t}{s_t} + (1 - \gamma)c_{t-L} \]

\[ F_{t+m} = (s_t + mb_t)c_{t-L} + ((m-1) \mod L) \]
Correlation Coefficient

\[ \rho_{xy}(\tau) = \frac{1}{n} \sum (X_t - \mu_x)(Y_{t+\tau} - \mu_y) \]

\[ \sigma_x \sigma_y \]
Correlation Among Stats
Front-end

- Remote Procedure Calls
- Google Web Toolkit
- Highcharts
Demo
Conclusion

- Forecast a typical day (e.g. Mon, Tue, etc..)

- Do automated scheduling based on information