Cloudatlas: Ways to Make Live Migration Easy and Expectable

Zhang Chao, Xie Feng

Alibaba Cloud
Migration Use Case:

**Product Ability**
- VM PLACEMENT
- LOAD BALANCE
- AGGREGATION
- POWER

**Benefit of Service Provider**
- HARDWARE MAINTENANCE
- ROOM MAINTENANCE
- ROLLING UPGRADE
- MAINTENANCE AND UPGRADE

**Improving Customer Experience**
- STABILITY
- HOTPLUGGING
- PRODUCT IMPROVING

**Gap between Basic Migration Ability and Its Use Scenery**

**Migration Optimization**
- VM
Migration Performance & Cost

- CPU Usage
- Network Usage

- Migration Totaltime
- Migration Downtime
- Success Rate

What we should do as our customers are demanding more Vcpu and Memory every day?

Just assign more cpus to our migration thread?
Example of Migration Cost

VM configuration:
32 vcpu 128G MEM

The dirty page rate difference between peak load period and low peak period may reach 5 times
Challenges

- WHY should a VM migration be performed?
- WHICH VM is required to migrate?
- WHEN to migrate it?
- HOW many resources are required?
DATA Boosted Migration Architecture

Why?

EVENT CENTRE (Maintenance Event Scheduling Event)

DATA ANALYZER

PREDICTER

MONITOR

MIGRATION CANDIDATE

MIGRATION TIME

MIGRATION PLAN

Which?

When?

How?

Migration Data Flow

WARNING

START MIGRATION

CANCEL/RESTORE

POST CHECK
Work Flow of the Migration Analyzer

Worklad Prediction

- Period
  - 60%~80%
- Non Period
  - 20%~40%

Workload Classification

- Migration-insensitive
  - Migrate immediately
- Migration-sensitive
  - Time Sensitive
  - Soft_Scheduler (Less Resource )
  - Resource Sensitive
  - Hard_Scheduler (CPU, bandwidth assignment)
Basics of Signal Processing

Transforming the original time series into the frequency domain can help us to detect the intrinsic pattern of the series.
Power Spectrum Density Estimation

Input Series

\[ x(n) \rightarrow R_N(n) \rightarrow x_N(n) \]

Truncate

\[ x_N(n) \rightarrow \text{N point FFT} \rightarrow X_N(k) \]

\[ \frac{1}{N} |X_N(k)|^2 \rightarrow \hat{S}_x(k) \]

Criteria:
- Top frequencies are selected.
- Lowest frequency is used as the base frequency.
- Original series are repeated in base frequencies as a prediction.
Time Series Period Estimation

Input time series

Input time series with noises

Period Estimation
Estimation of Predict Accuracy

Assume:
noises has a distribution of $w(n) \sim N(\mu, \delta)$

For simplicity:
$x[-T:] - x[-2T:-T]$ has a distribution of $N(0,2\delta)$

Thus:
RMSE of $(x[-T:] - x[-2T:-T])$ can be used to indicate the predict accuracy of the future workload.

Input time series with noises. SNR: 5db

Input time series with noises. SNR: 1db

In sample RMSE: 1.98

In sample RMSE: 3.21
Estimation of Real VM Usage data

In sample RMSE: 21.8, out of sample RMSE 19.8

VM1

Time Cost to Estimate a Whole day workload: 10–20ms

In sample RMSE: 0.3, out of sample RMSE 0.2

VM2
Workload Time Series Model

Use Time Series Model to predict next 24 hours workload of VM

ARIMA Model: AutoRegressive Integrated Moving Average Model
• A Traditional Machine Learning method
• Advantage:
  Low calculation cost
  Good performance on Small Data Set
• Disadvantage:
  Lower accuracy than Deep Learning

LSTM: Long Short Term Memory networks(a special kind of RNN)
• A Deep Learning method
• Advantage:
  High accuracy
• Disadvantage:
  High calculation cost
  Need a Large Data Set
Regression algorithm Flow Chart

- ARIMA Model Fit (one-hour sampled data)
- LSTM Model Training (one-minute sampled data)
- Model evaluation
- Good
- Predict the low workload one-hour time window over the next 24 hours
- Predict a optimal migration time window
- Predict the low workload ten-minute time window in the one-hour time window
Regression algorithm Result: ARIMA

ARIMA Predict Next 24 Hours Result

75% of non-periodic VM RMSE < 10
93% of non-periodic VM RMSE < 15

Low workload time window
Regression algorithm Result: LSTM

LSTM Real Time Predict Result

80% of non-periodic VM  
RMSE < 5

91% of non-periodic VM  
RMSE < 10
## Algorithm Performance

<table>
<thead>
<tr>
<th>Workload Type</th>
<th>Algorithm</th>
<th>Time cost</th>
<th>Perc.</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong Period</td>
<td>FFT</td>
<td>10~30ms</td>
<td>60%~80%</td>
<td>&lt;10</td>
</tr>
<tr>
<td>Weak Period</td>
<td>ARIMA</td>
<td>&lt; 10min</td>
<td>10%~15%</td>
<td>&lt;15</td>
</tr>
<tr>
<td></td>
<td>LSTM</td>
<td>&lt; 2hour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non Period</td>
<td>ARIMA</td>
<td>&lt; 10min</td>
<td>9%~15%</td>
<td>&lt;15</td>
</tr>
<tr>
<td></td>
<td>LSTM</td>
<td>&lt; 2hour</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Workload Classification algorithm (Work in process)

Use a Machine Learning binary classification algorithm to predict whether a VM is migration-sensitive

Feature
- Average vCPU utilization (1 hour before migration)
- Amplitude of fluctuation with vCPU utilization (one day before migration)
- VM Instance Type (How many vCPU/Memory?)

Result
- Migration-insensitive VM (downtime <= 100 ms)
- Migration-sensitive VM (downtime > 100 ms)
SUMMARY: Workload Prediction Flow Chart

Classification algorithm ➔

Whether is migration-sensitive ➔

insensitive ➔ Migrate immediately

sensitive ➔

FFT ➔

Whether is periodical ➔

YES ➔ Period repeat ➔ Predict a low workload time window

NO ➔ Regression algorithm
Questions?