

Методы прогнозирования процессорной нагрузки

М.В. Кудинова, А.Л. Мелехова, А.А. Веринов

Московский физико-технический институт (государственный университет)

Прогнозирование процессорной нагрузки является важной составляющей задачи распределения ресурсов и планирования заданий. Оно используется во всевозможных распределённых системах, повышая аккуратность балансировки. Более аккуратная балансировка в свою очередь позволяет увеличить плотность виртуальных серверов на одном физическом, то есть повышает эффективность использования ресурсов. В этом случае точный прогноз процессорной нагрузки позволяет своевременно мигрировать одну или несколько виртуальных машин, предотвращая "перегрузку" физических машин. Однако, из-за разнообразия и многочисленности методов прогнозирования, сложно выбрать метод подходящий для конкретной ситуации.

В данном обзоре рассматриваются наиболее популярные подходы к прогнозированию процессорной нагрузки, достоинства и недостатки применяемых методов, а также задачи, в которых они хорошо себя зарекомендовали. Основное внимание уделено регрессионным (ARFIMA и её частным случаям) и Байесовским моделям, а также фильтру Калмана и кластеризации. Приведены математические основы и алгоритмы реализации соответствующих методов, приведены принимаемые в моделях предпосылки.

В рассмотренных работах было показано, что при предсказании помимо нагрузки на процессор могут быть успешно использованы корреляции между потреблением различных компонент одного приложения, и различная информация об активности операционной системы. Был сделан вывод, что при выборе метода предсказания следует учитывать такие характеристики нагрузки, как соотношение длины задания и времени прогнозирования, характерное время стационарности нагрузки, важность случайных флуктуаций. По итогам рассмотрения работ, приводится сравнение различных подходов к прогнозированию процессорной нагрузки.

Литература

1. *Dhiman, G. [и др.] A system for online power prediction in virtualized environments using gaussian mixture models. // Design Automation Conference (DAC), 2010 47th ACM/IEEE. 3 (2010), 807–812. 2010.*
2. *Di, S. [и др.] Characterization and Comparison of Google Cloud Load versus Grids. // International Conference on Cluster Computing (IEEE CLUSTER). (2012), 230–238. 2012.*

3. *Di, S.* [u др.] Host load prediction in a Google compute cloud with a Bayesian model. // International Conference for High Performance Computing, Networking, Storage and Analysis, SC. 2012.
4. *Dinda, P.A., O'Hallaron, D.R.* Host load prediction using linear models. // Cluster Computing. 3, (2000), 265–280.
5. *Gray, R.M.* Gauss mixture vector quantization. // 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221). 3, (2001), 1–4. 2001.
6. *Kalman, R.E.* A New Approach to Linear Filtering and Prediction Problems. // Transactions of the ASME-Journal of Basic Engineering. 82, Series D (1960), 35–45. 1960.
7. *Kalyvianaki, E.* [u др.] Adaptive Resource Provisioning for Virtualized Servers. // ACM Transactions on Autonomous and Adaptive Systems. 9, 2. 2014.
8. *Padala, P.* [u др.] Adaptive control of virtualized resources in utility computing environments. // ACM SIGOPS Operating Systems Review. 41, 3, 289. 2007.
9. *Zhang, Y.* [u др.] CPU load predictions on the computational grid. // IEICE Transactions on Information and Systems. E90-D, 1 (2007), 40–47. 2007.
10. *Ивченко, Г.И., Медведев, Ю.И.* Введение в математическую статистику. Издательство ЛКИ. 2010.
11. *Орлов, Ю.Н.* Кинетические методы исследования нестационарных временных рядов. МФТИ. 2014.