

Задача прогнозирования нагрузки для повышения энергетической эффективности вычислительного кластера¹

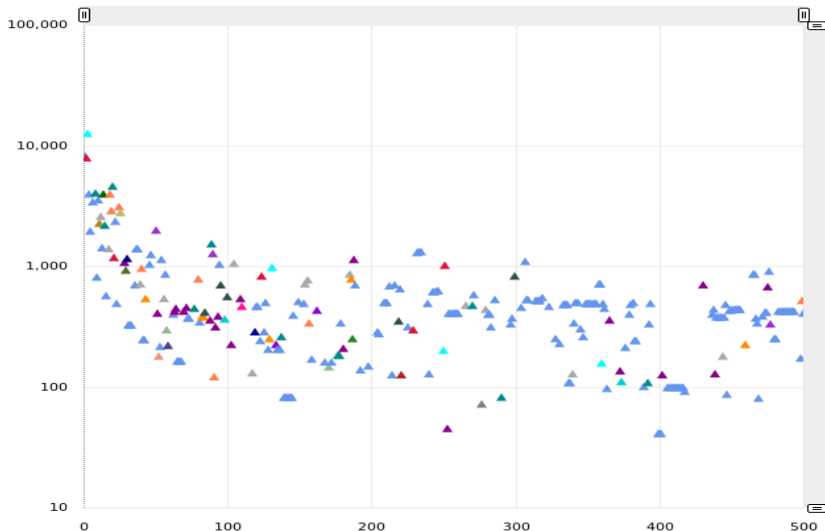
Е. Е. Ивашко, А. С. Румянцев, А. Л. Чухарев

Институт прикладных математических исследований
Карельского научного центра РАН

04.04.2013 г.

¹ Исследования выполнены при поддержке РФФИ (проект 12-07-31147) и Фонда содействия малым формам предпринимательства в научно-технической сфере (г/к №10491р/16862 от 08.06.2012).

Энергопотребление вычислительных кластеров



1 час работы лидеров top-500 = 1 тонна алюминия!

Снижение энергопотребления — комплексная задача:

- производство энергоэффективного оборудования;
- проектирование ЦОДов с низкими накладными расходами;
- *экономия за счет нестационарности нагрузки.*

Стандартные методы (поддерживаются оборудованием):

- S-состояния АСРІ;
- C-состояния процессора;
- снижение частоты процессора.

Конфликт интересов

Владелец:

- 1 оптимальная загрузка
(работа без простоев);
- 2 экономия ресурсов
(энергосбережение).

Пользователь:

- 1 максимально быстрое
получение результата;
- 2 минимальное время
отклика.

Задача: повышение энергоэффективности вычислительного кластера без ущерба для качества обслуживания.

Интенсивность потока задач зависит от качества обслуживания (удовлетворенность пользователя).

Некоторые обозначения

- Множество пользователей $U = \{u_1, \dots, u_n\}$
- Сессии доступа в момент $\tau_i(u)$ продолжительностью $\sigma_i(u)$
- Поток задач в моменты $t_j(u) \in [\tau_i(u), \tau_i(u) + \sigma_i(u)]$ (для некоторого i)
- Задаче требуется $N_j(u)$ процессоров на время $S_j(u)$
- Задача ожидает в очереди $D_j(u)$
- Время до готовности всех ресурсов к принятию заявки V_j

Типичные метрики:

- время отклика системы $D_i + S_i$;
- замедление задачи $\frac{D_i + S_i}{S_i}$.

Качество восприятия на заявке i (если ресурсы не готовы):

$$R_i(V_i) = \alpha(D_i + V_i + S_i) + \beta \frac{D_i + V_i + S_i}{S_i}.$$

Качество обслуживания это $F(R_1(V_1), \dots, R_n(V_n))$, в простейшем случае $\frac{1}{n} \sum_{i=1}^n R_i(V_i)$.

Без ущерба для качества обслуживания:

$$P \left(\frac{F(R_1(V_1), \dots, R_n(V_n))}{F(R_1(0), \dots, R_n(0))} \in [1, 1 + \varepsilon] \right) = 1 - \delta.$$

Снижение энергопотребления

- Пусть имеется свободный ресурс из K устройств (процессоров, узлов).
- Уменьшение K — приход новой задачи (очередь пуста), завершение обслуживания (очередь не пуста).
- Момент времени уменьшения \hat{t} , величина уменьшения \hat{N} .
- Устройства имеют режимы C_0, \dots, C_k (большой индекс — меньшее потребление).

Необходимо построить $\{\kappa_0, \dots, \kappa_k\}, \{\tilde{t}_0, \dots, \tilde{t}_k\}$, где κ_i — количество устройств в режиме C_i , а \tilde{t}_i — время выхода из него. В простейшем случае

$$\kappa_i = K - \hat{N}, \kappa_{i-1} = \hat{N}, \tilde{t}_i = \hat{t}, \tilde{t}_{i-1} = \hat{t} - T_{i-1},$$

где $T_{i-1} < \hat{t}$ — наименьшее время пребывания в состоянии C_{i-1} .

Прогноз уменьшения ресурса

Пусть в момент t очередь отсутствует. Тогда

- Для всех активных пользователей ($\tau_i(u) < t < \tau_i(u) + \sigma_i(u)$) реализуем величины $\hat{t}_j(u)$, $1 \leq j \leq m$ ближайшей постановки задачи в очередь, а также $\hat{N}_j(u)$, $\hat{S}_j(u)$.
- Для не активных реализуем время $\hat{t}_{au_j}(u)$ начала сессии и время $\hat{t}_j(u)$, $1 \leq j \leq m$ ближайшей постановки задачи в очередь, а также $\hat{N}_j(u)$, $\hat{S}_j(u)$.

Тогда

$$\hat{t} = \varphi(\hat{t}_1(U), \dots, \hat{t}_m(U)), \quad \hat{N} = \psi(\hat{N}_1(U), \dots, \hat{N}_m(U)).$$

В простейшем случае $\varphi(\cdot) = \psi(\cdot) = \min(\cdot)$.

Если очередь в системе не пуста \Rightarrow учитывать уходы заявок.

$$E(V) \rightarrow \min$$
$$P \left(\frac{F(R_1(V_1), \dots, R_n(V_n))}{F(R_1(0), \dots, R_n(0))} \in [1, 1 + \varepsilon) \right) = 1 - \delta.$$

Планируется проведение численных экспериментов на базе вычислительного кластера ЦКП КарНЦ РАН «Центр высокопроизводительной обработки данных».

Спасибо за внимание!

cluster@krc.karelia.ru