# Вводная

В данном документе рассматривается составление типичного профиля нагрузочного тестирования: для определения максимальной производительности системы.

**Профиль нагрузочного тестирования** – это набор операций и их интенсивности, подобранные таким образом, чтобы в совокупности они составляли нагрузку, сопоставимую с типичной продуктивной.

Подготовка профиля нагрузки – один из важнейших этапов нагрузочного тестирования.  
Составив неверный профиль, вы получите результаты тестирования, не относящиеся к реальному применению тестируемой системы.

Начать подготовку профиля стоит с изучения **цели тестирования**, например:

* определение максимальной производительности – требуется подготовить профиль, максимально нагружающий систему;
* оценка влияния конкретных процессов на работу системы – требуется подготовить профиль за период, когда на продуктивном контуре работают эти процессы и оценивать их влияние не на всю систему, а только на те операции, которые работают совместно с изучаемыми процессами;
* выявление наиболее тяжёлых операций для дальнейшей оптимизации – требуется сперва согласовать список операций-кандидатов, а потом подготовить один или несколько профилей для их изучения

и так далее.

# Сбор статистики

## Период

Период анализируемой статистики должен включать несколько циклов продуктивной нагрузки, по которым можно выделить типичную нагрузку. Рекомендуется брать статистику за период не меньше месяца.  
Период не должен быть слишком большим – если взять статистику за полгода, итоговый профиль будет отображать усреднённую нагрузку за **весь** период, что неверно, если уровень нагрузки постоянно растёт. Кроме того, большое количество данных усложнит анализ.

При этом следует учесть, что месяцы тоже отличаются: в декабре нагрузка стандартно выше, зато в январе или мае много праздничных дней. Статистика за эти месяцы может быть не показательной. В таких случаях рекомендуется сравнить её со статистикой за соседние месяцы для оценки отклонения. Как вариант, можно увеличить анализируемый период.

## Агрегация

В рядовых случаях интенсивность операций агрегируют по часам (в случае, если интенсивность составляет более 100 операций в час).

# Анализ статистики

## Дни для составления профиля

У большинства систем есть более и менее нагруженные периоды:

* день и ночь
* будние и выходные дни
* дни недели (в понедельники нагрузка часто выше, чем в пятницы).
* нагрузка может быть привязана к числам месяца
* или порядковым дням недели (например, первые четверги месяца)

Необходимо выделить дни с типичной для требуемого профиля нагрузкой.   
Для поиска максимальной производительности это набор дней, нагрузка в которые похожа и является наивысшей. Количество отобранных дней должно быть достаточным, что бы формируемую ими нагрузку можно было назвать **типичной**.

Например, один день в месяц с нагрузкой, заметно отличающейся от нагрузки в другие дни выглядит не типично и на его основе составлять профиль нельзя – даже если это нагрузка в этот день наивысшая.   
А вот если таких дней несколько и работа системы подразумевает соответствующее количество дней с наивысшей нагрузкой в месяце – профиль надо строить именно на основе этих дней.

|  |
| --- |
| На графике ниже каждая линия отражает нагрузку в час в разбивке **по дням**:   * Зелёным – дни с типичной максимальной нагрузкой. * Красным – дни с нетипичной нагрузкой * Серым – дни с низкой нагрузкой   Профиль строится только по статистике «зелёных» дней. |

## Пик-час и покрываемый им диапазон времени

Для выбранных дней производится разбивка по операциям с усреднением интенсивности операций по часам.

Основой для профиля служит час пиковой нагрузки – час, когда суммарная нагрузка по операциям наивысшая.

Необходимо оценить диапазон времени, в котором отношение операций друг к другу незначительно отличается от пик-часа. Этот период будет покрыт профилем тестирования. Для остального времени требуется составление отдельного профиля.

|  |
| --- |
| На графике ниже каждая линия отражает нагрузку в час в разбивке **по операциям**.  Вертикальными линиями обозначены:   * Зелёным – пиковый час. * Желтым – границы периода, покрываемого профилем, составленному по пиковому часу.   Для профиля отбираются операции, с наивысшей нагрузкой в пиковый час. |

Если профиль строится для анализа конкретных операций, следует убедиться, что выбраны дни и часы с **максимальной** интенсивностью именно по этим операциям.   
И наоборот, в случае наличия малозначимых но высокоинтенсивных операций следует выбирать пик-час без их учёта.

## Выбор операций для включения в профиль

На основе интенсивности операций за выбранный пик-час необходимо выбрать набор операций, включаемых в профиль.

* Для начала следует определиться с требуемой **точностью** профиля. Обычно она составляет 90% и более.   
  Это значит, что необходимо упорядочить операции по убыванию интенсивности и включить в профиль те, что суммарно подают не менее 90% нагрузки.   
  Важно понимать, что первоочередное значение имеет не процент, который каждая операция составляет от общей нагрузки, а **суммарный** процент выбранных операций (в профиль могут попасть операции, составляющие менее 1% от суммарной нагрузки).

Так же в профиль могут быть включены операции, находящиеся за пределами требуемой точности профиля:

* Критичные операции – операции, которые необходимо включить в профиль вне зависимости от интенсивности из-за их важности для бизнеса или влияния на работу системы.

Если про эти операции известно, что основная нагрузка по ним проходит в конкретные дни/часы, следует убедиться, что эти часы попали в анализ, а, возможно весь профиль стоит строить на этих часах.

* Операции, включение которых в профиль не требует трудозатрат (уже существуют СНТ).

В некоторых моделях подачи нагрузки может возникнуть требование к интенсивности операций профиля:

* Интенсивность каждой операции должна быть достаточной, чтобы на каждой ступени теста операция выполнялась достаточное количество раз для проверки на соответствие требованиям.

|  |
| --- |
| Предположим, что:   * требуется включить в профиль не менее 98% нагрузки * по модели нагрузки требуется, чтобы интенсивность операции была не менее 200 в оп./ч.   По таблице ниже видно, что:   * Операции 1-12 включены в профиль, поскольку они в сумме составляют 98% нагрузки (см. столбец «Промежуточный итог») * Операции 20-22 включены в профиль как критичные * Дополнительно в профиль включены операции 13, 15, 17-19, поскольку для них уже есть СНТ и их интенсивность достаточно для включения в данную модель тестирования   C:\Users\User\AppData\Local\Temp\SNAGHTML265e1771.PNG |

## Определение интенсивности операций в профиле

По результатам предыдущих шагов составлен список операций и диапазон времени работы системы, который будет покрыт профилем. В итоговый профиль включается **наивысшая** интенсивность каждой операции за покрываемый профилем диапазон времени. После этого оценивается, насколько получившийся профиль перекрывает по интенсивности пиковый час. Допустимое значение – порядка 10%. В случае превышения допустимого значения необходимо скорректировать диапазон времени, покрываемый профилем.   
Таким образом, итоговый профиль эмулирует наивысшую возможную нагрузку.

|  |
| --- |
| По таблице ниже видно, что по различным операциям интенсивность максимальная в разные часы с 10 до 14 часов (зелёные ячейки). |
| Собрав эти пики воедино, мы увидим, насколько отклоняется интенсивность этих операций за остальные часы от часов, выбранных для составления профиля:    Как видно из таблицы выше, интенсивность больше профильной только у операции LN\_DELREQF и это превышение находится за пределами диапазона, покрытого пик-часом (10-16 часов) |
| Если бы мы взяли для составления профиля интенсивность по пик-часу, отклонения выглядели бы так: |

## Пакетные и атомарные операции

Некоторые операции могут выполняться как атомарно, так и последовательно. Например, в банковской сфере после создания клиента для него обычно следует создание счетов и активация ДКО. Пакетные операции могут формировать нагрузку, отличную от выполнения тех же операций атомарно. В таких случаях из собранного профиля сперва выделяется необходимая интенсивность для пакетных операций, а оставшаяся интенсивность подаётся атомарно.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Например:  По логам веб-сервера известно, что статистика следующая:   |  |  | | --- | --- | | **Операция** | **Операций в час** | | Вход в систему | 100 | | Работа с данными системы | 80 |   Из этой статистики видно, что из 100 случаев входа в систему, только 80 продолжаются работой с данными, а 20 ограничиваются входом. Соответственно, на основании данной статистики будут сформированы следующие Use-Case:   |  |  | | --- | --- | | **Use-Case** | **Операций в час** | | 1) Вход в систему 2) Работа с данными системы | 80 | | 1) Вход в систему | 20 | |

# Сравнение профиля с предыдущим

После составления профиля рекомендуется сравнить его с предыдущим профилем, в случае его наличия.   
В сравнение входит:

* Сравнение **суммарного** количества операций
* Сравнение интенсивности каждой отдельной операции в профиле   
  (выполняется после нормирования старого профиля до уровня нового: интенсивность каждой операции старого профиля умножают на такой коэффициент, что бы суммарное число операций старого профиля совпадала с суммарным числом операций нового профиля).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Из таблицы ниже видно, что:   * Интенсивность нового профиля увеличилась на 67% * Доля UC1 в новом профиле увеличилась на 14% * Доля UC2 в новом профиле уменьшилась на 33%  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **№** | **Use-Case** | **Интенсивность (по новому профилю), оп./ч.** | **Интенсивность (по старому профилю) оп./ч.** | **Нормированная интенсивность (по старому профилю) оп./ч.** | **Отличие нового профиля от норм. старого, %** | | 1 | 1) Вход в систему 2) Работа с данными системы | 80 | 42 | 70 | 14% | | 2 | 1) Вход в систему | 20 | 18 | 30 | -33% | | **ИТОГО:** | | **100** | **60** | 100 |  | | **Рост интенсивности профиля составил:** | | **67%** |  |  |  | |

## Как часто обновлять профиль

Необходимость обновления профиля зависит от изменения нагрузки на тестируемую систему. Рекомендуется актуализировать профиль раз в квартал.