Цифровой блок НИУ ВШЭ

Отдел суперкомпьютерного моделирования

Москва, 2024
Суперкомпьютерные дни в России

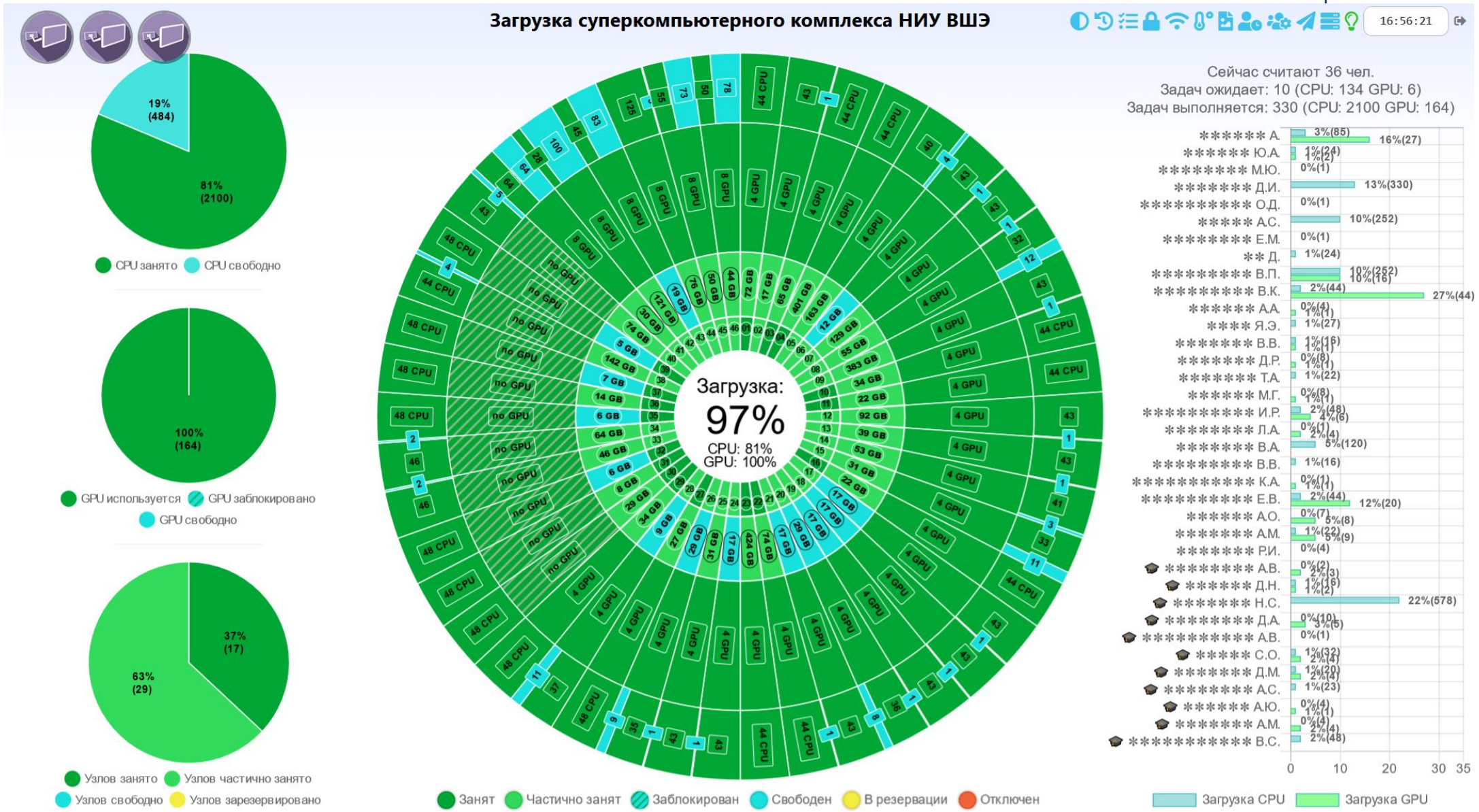
Разработка новых индикаторов для системы HPC TaskMaster

А.А. Раимова, В.И. Козырев, Р.А. Чулкевич, П.С. Костенецкий

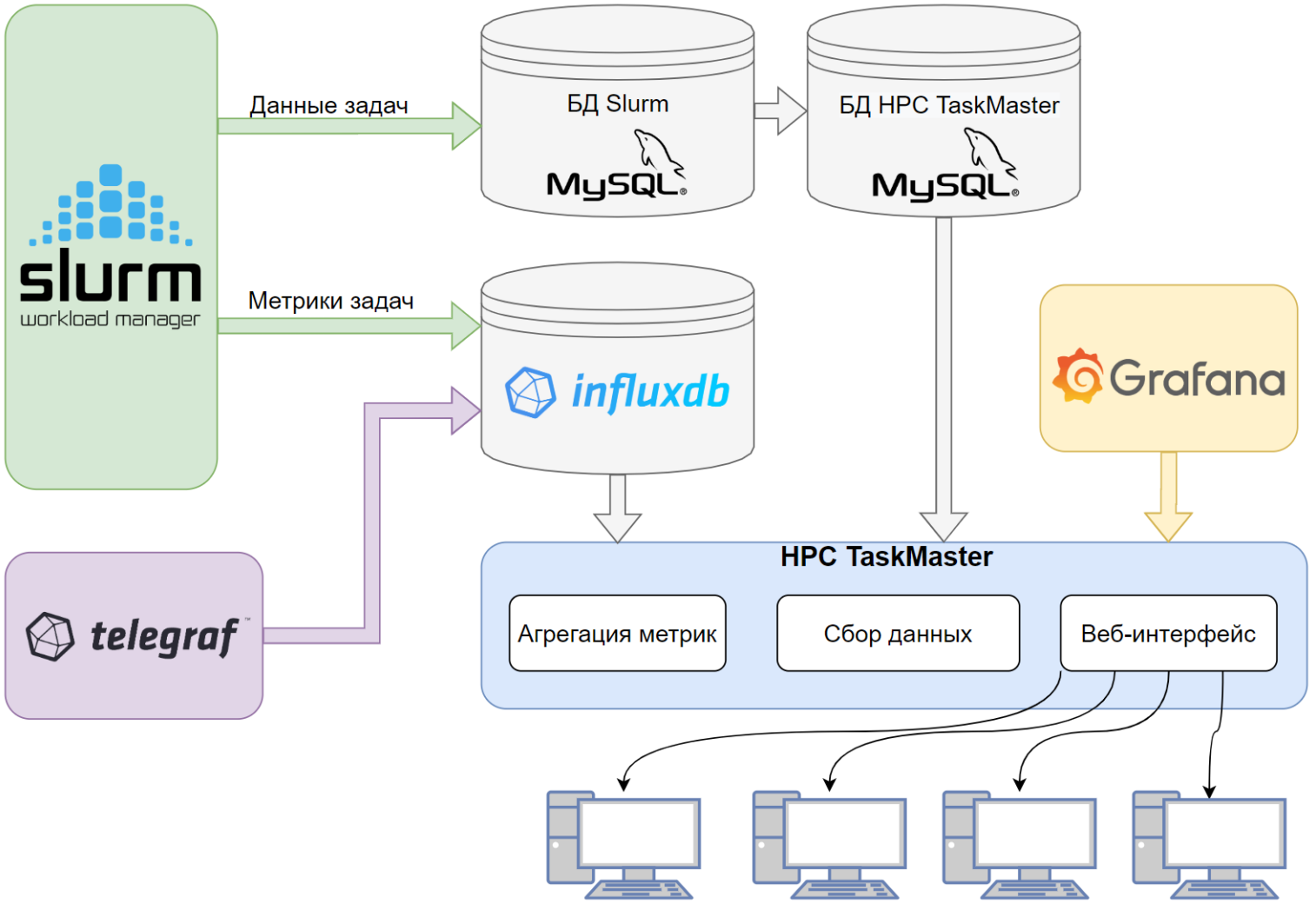
Характеристики суперкомпьютера cHARISMa (Computer of HSE for Artificial Intelligence and Supercomputer Modelling)

- **10 место в ТОП 50**
- Пиковая производительность: **2 Петафлопс** (2 квадриллиона операций в секунду над числами с двойной точностью)
- LINPACK-производительность: **927.4 Терафлопс**
- **46** вычислительных узлов
  - **6** узлов **с 1 ТБ** ОЗУ, **8 GPU A100 80 ГБ SXM**
  - **10** узлов с **1,5 ТБ** ОЗУ, 4 GPU V100 32 ГБ
  - **19** узлов с **768 ГБ** ОЗУ, 4 GPU V100 32 ГБ
  - **11** узлов с 384 ГБ ОЗУ для расчётов на CPU
- **2** управляющих узла
- **148 GPU NVIDIA Tesla A100 80 ГБ**
- **164 GPU NVIDIA Tesla V100 32 ГБ**
- **2584** ядер центральных процессоров
- Оперативная память: **40,3 ТБ RAM** + **7.5 ТБ GPU Memory**
- Дисковая память: **1,15 ПБ**
- параллельная СХД на базе Lustre **840 ТБ**
- локальные диски **128 ТБ**
- сервер резервного копирования **182 ТБ**
- Коммуникационная сеть: **2 x InfiniBand EDR**
- (**2x100 Гбит/с**, топология **FatTree**)

Загрузка суперкомпьютера



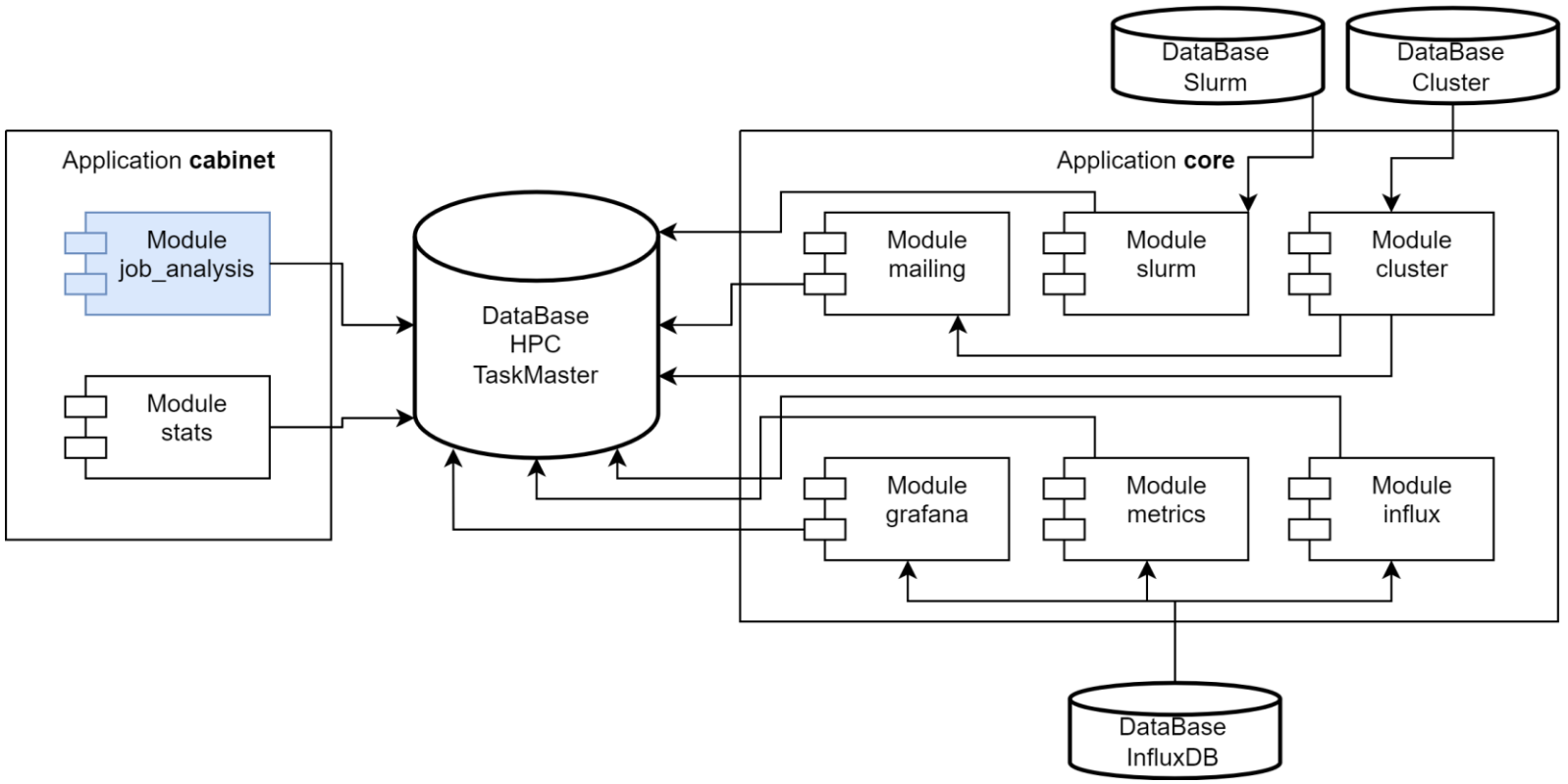
Описание системы HPC TaskMaster



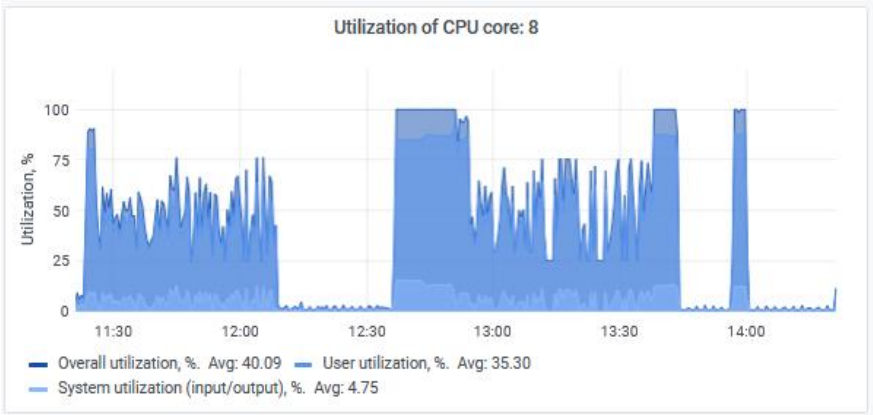
- Открытый исходный код
- Информация о задачах, а не об узлах
- Автоматический анализ метрик
- Интерактивные графики

Описание системы HPC TaskMaster

- Модуль **mailing**: отправка сообщений пользователям о неэффективных задачах;
- Модуль **slurm** собирает данные о задачах из БД SLURM;
- Модуль **cluster**: статистика использования суперкомпьютера пользователями и проектами;
- Модуль **grafana**: дашборды с графиками по использованию ресурсов;
- Модуль **job_analysis**: обработка агрегированных метрик и параметров задач для назначения индикаторов, тегов и выводов;
- Модуль **stats**: создание статистики о пользователях и их задачах;
- Модуль **metrics**: агрегирование метрик и обработка данных;
- Модуль **influx** поддерживает подключение к InfluxDB и собирает метрики задач.



Проблема



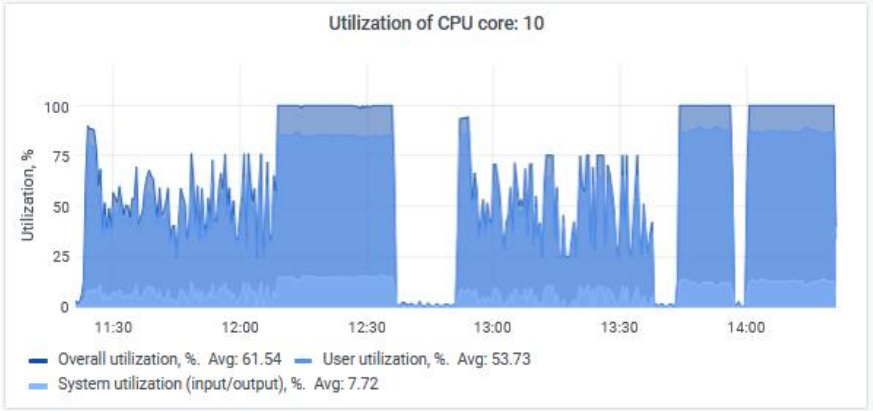
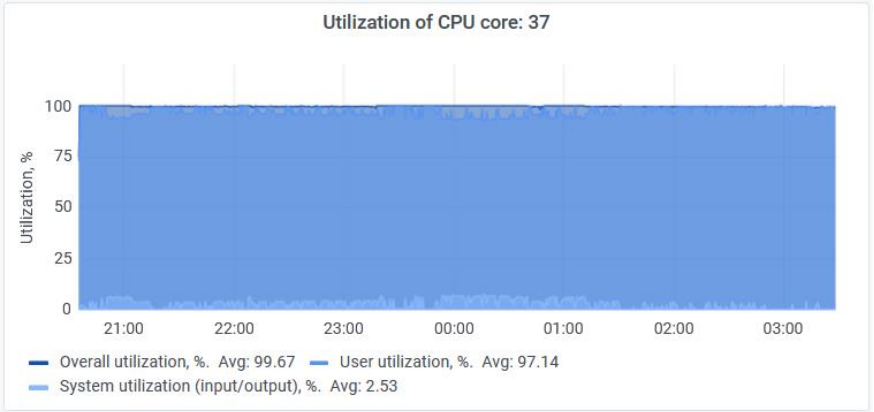


График непараллельной задачи



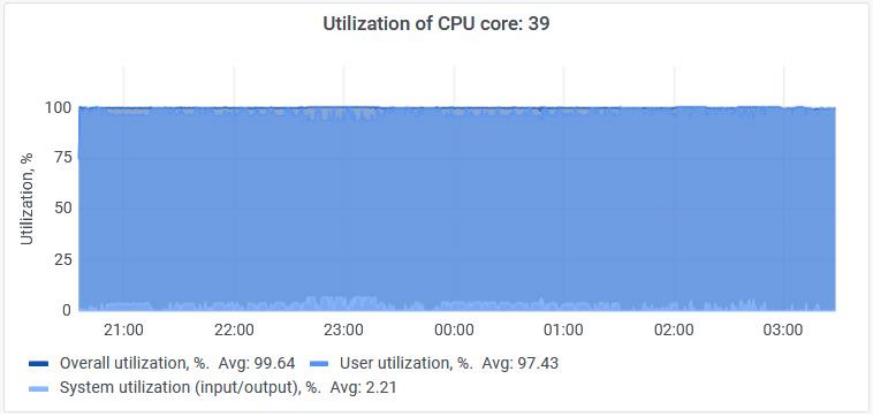


График параллельной задачи

Нужен **индикатор**, который будет обнаруживать задачи, которые **не** поддерживают **распараллеливание**, но при этом запущены на **нескольких** ядрах

Исследование

Алгоритм 1

Корреляция Спирмена

$$r_s = 1 - \frac{6\sum_{i=1}^{m} d^2}{n(n^2-1)}$$

Алгоритм 2

Мера САП-трансформ

Локальный тренд

$$a_i = \frac{6\sum_{j=0}^{k-1}(2j-k+1)y_{i+j}}{hk(k^2-1)}, i \in (1,\dots,m)$$

Мера локальных трендовых ассоциаций

$$coss_k(y,x) = \frac{\sum_{i=1}^{m} a_{yi} \cdot a_{xi}}{\sqrt{\sum_{i=1}^{m} {a_{yi}}^2 \cdot \sum_{j=1}^{m} {a_{xj}}^2}}$$

Результат

Критерий	Корреляция Спирмена	САП-трансформ	Ускоренный САП-трансформ
F1-мера	0.407186	0.918699	0.91498
Минимальное время	0.00119872	0	0
Максимальное время	0.0196644	369.336	2.95074
Среднее время	0.00234576	1.42191	0.0529625

- Алгоритм САП-трансформ был **оптимизирован** путем подсчета меры только за **последние *p*** минут;
- **Сравнение** двух алгоритмов производилось по двум критериям: **время** выполнения и значение метрики **F1-мера**;
- Для сравнения алгоритмов была собрана выборка из **1502** объектов ;
- Среднее время подсчета метрик у ускоренного САП-трансформ **уменьшилось** в **3,5 раза**.