

Организация потоковой обработки данных в объединенной системе работы с данными НИЦ «Планета»

Бурцев М.А., Прошин А.А., Матвеев А.А., Мазуров А.А., Кобец Д.А.

Институт космических исследований РАН



XVIII Всероссийская Открытая конференция
СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА
20 ноября 2020 г.



О системе



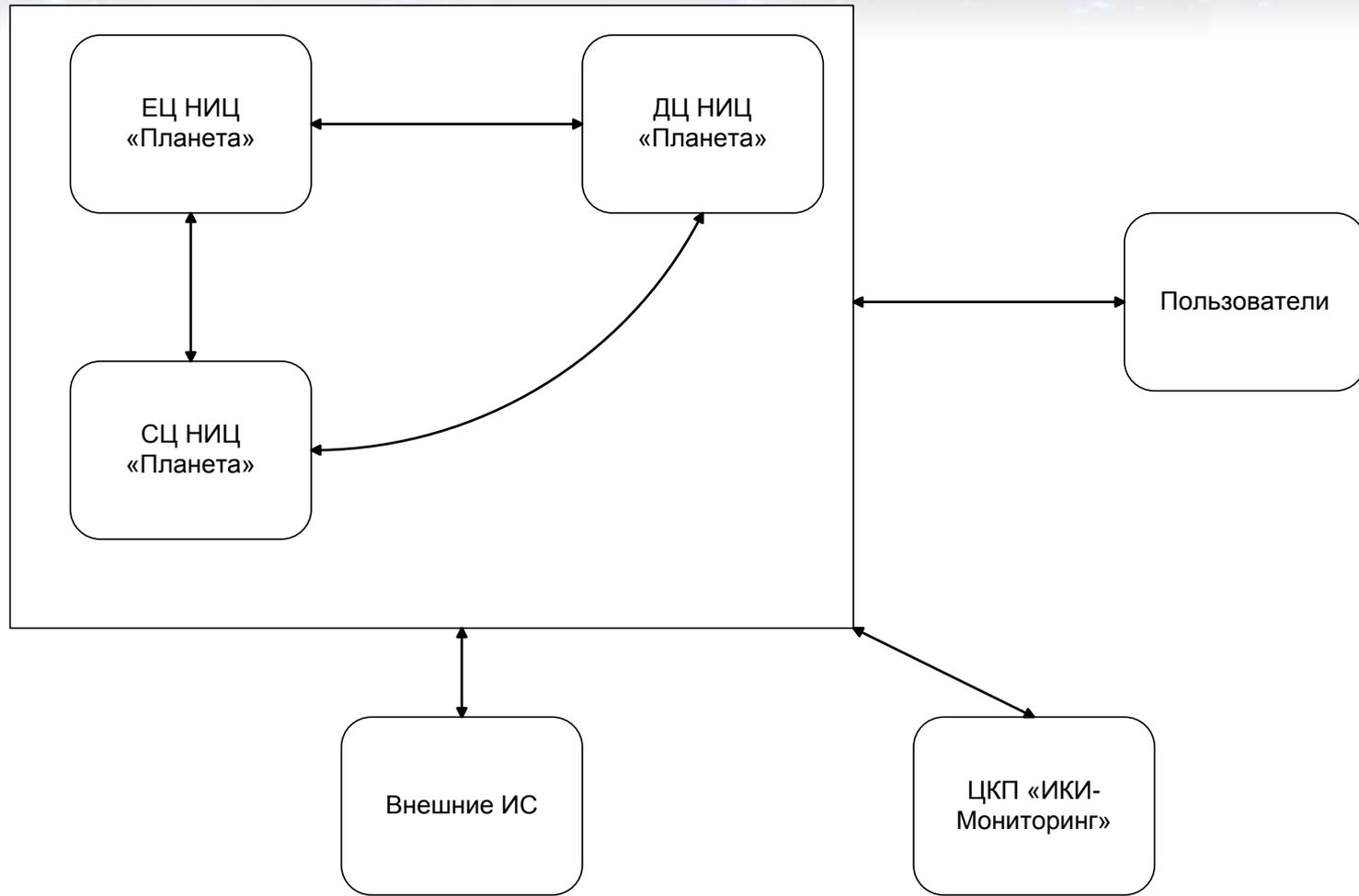
Цель создания системы

Объединённая система работы с данными центров ФГБУ «НИЦ «Планета» введена в эксплуатацию в 2013 году.

Она представляет собой единое, универсальное средство работы с данными центров приёма, обработки и распространения, охватывающее все этапы их жизненного цикла от получения со станции до предоставления их специалисту.



Структура системы



Основные задачи системы

- Предоставление возможности однотипной, прозрачной работы с данными всех центров НИЦ «Планета»;
- Унификация и автоматизация систем обработки данных в различных центрах НИЦ «Планета»;
- Обеспечение возможности распределенной работы с архивами спутниковых данных различных центров;
- Предоставление инструментов для анализа спутниковых данных и информационных продуктов, получаемых на их основе в различных центрах;
- Поддержка унифицированной схемы онлайн-взаимодействия с различными системами дистанционного мониторинга.



Система потоковой обработки

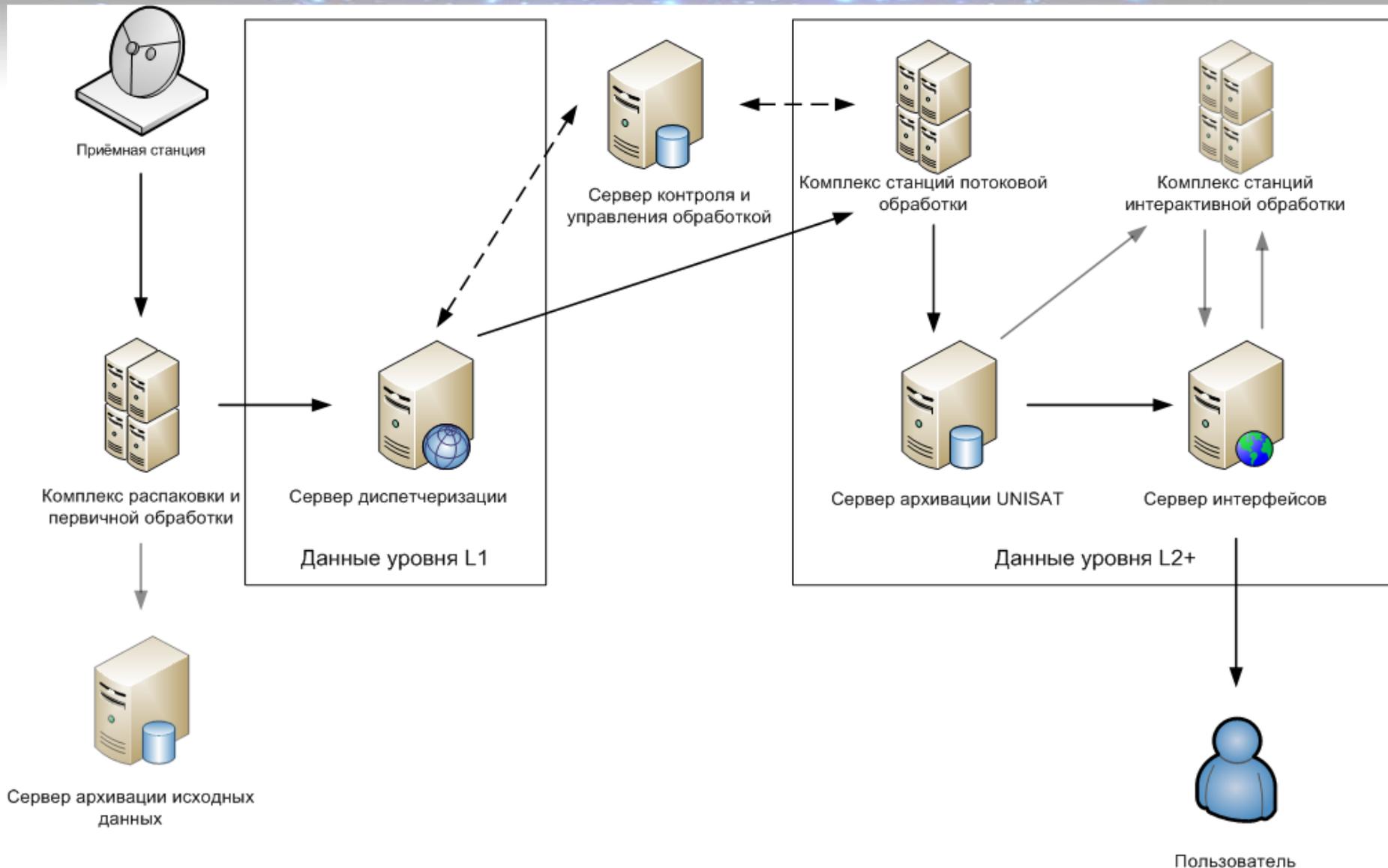


Основные задачи системы

- Доведение всех поступающих разнородных данных до уровня L2+ в едином формате UNISAT с поддержкой механизмов построения «виртуальных продуктов» и интерактивного анализа и обработки;
- Организация обработки разнородных данных в едином контуре управления и контроля;



Общая схема организации потока



Система управления и контроля

Централизованный сервер управления и контроля обеспечивает:

- Ведение БД с настройками видов обработки;
- Распределение типов обработки по станциям обработки;
- Контроль и балансировку выполнения заданий на обработку между станциями обработки;
- Сбор информации о состоянии заданий на диспетчерах;
- Сбор детальной информации о выполняемых процессах обработки на станциях и последующую её агрегацию средствами BI-аналитики.



Типовое задание на обработку

Описание типового задания на обработку включает:

- Адреса серверов диспетчеризации и сбора результатов;
- Точки расположения физических данных для обработки;
- Адрес расположения инструкций для обработки;
- Настройки допустимого количества ресурсов, которые можно выделить на данный тип обработки.



Распределение и контроль выполнения

Распределение и балансировка заданий по станциям позволяет:

- Задавать, какие именно станции могут выполнять определённые типы обработок;
- Управлять приоритетами выполнения обработки;
- Контролировать состояние ресурсов обработки во избежание узких мест в обработке.



Сбор информации

В ходе работы постоянно идёт сбор информации о:

- Состоянии всех очередей обработки на диспетчерах;
- Состоянии станций обработки;
- Всех выполняемых заданиях на станциях обработки с точностью до каждого шага обработки;

Это позволяет получить полную статистику и профили работы системы обработки для определения качества её работы.



BI-аналитика

В ходе работы постоянно идёт сбор информации о:

- Состоянии всех очередей обработки на диспетчерах;
- Состоянии станций обработки;
- Всех выполняемых заданиях на станциях обработки с точностью до каждого шага обработки;

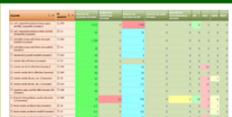
Это позволяет получить полную статистику и профили работы системы обработки для определения качества её работы.



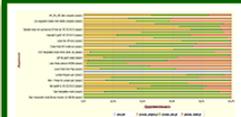
BI-аналитика

PROCI Планета

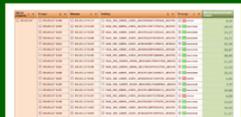
Аналитика обработки задания



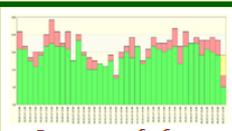
Статистика



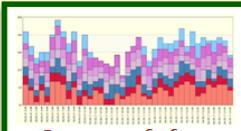
Продолжительность периодов подготовка, обработки и выгрузки данных



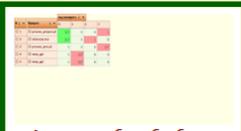
Детализация обработки задания



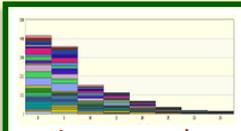
Детализация обработки задания (статусы, часы)



Детализация обработки задания (обработчики, часы)



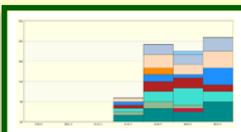
Анализ ошибок обработки задания



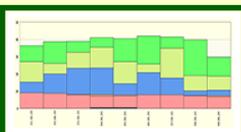
Анализ простоя .rdy



Детализация обработки задания (статусы, дни)



Детализация обработки задания (обработчики, дни)



Аналитика кластера (задания; посуточно)

Аналитика работы обработчика

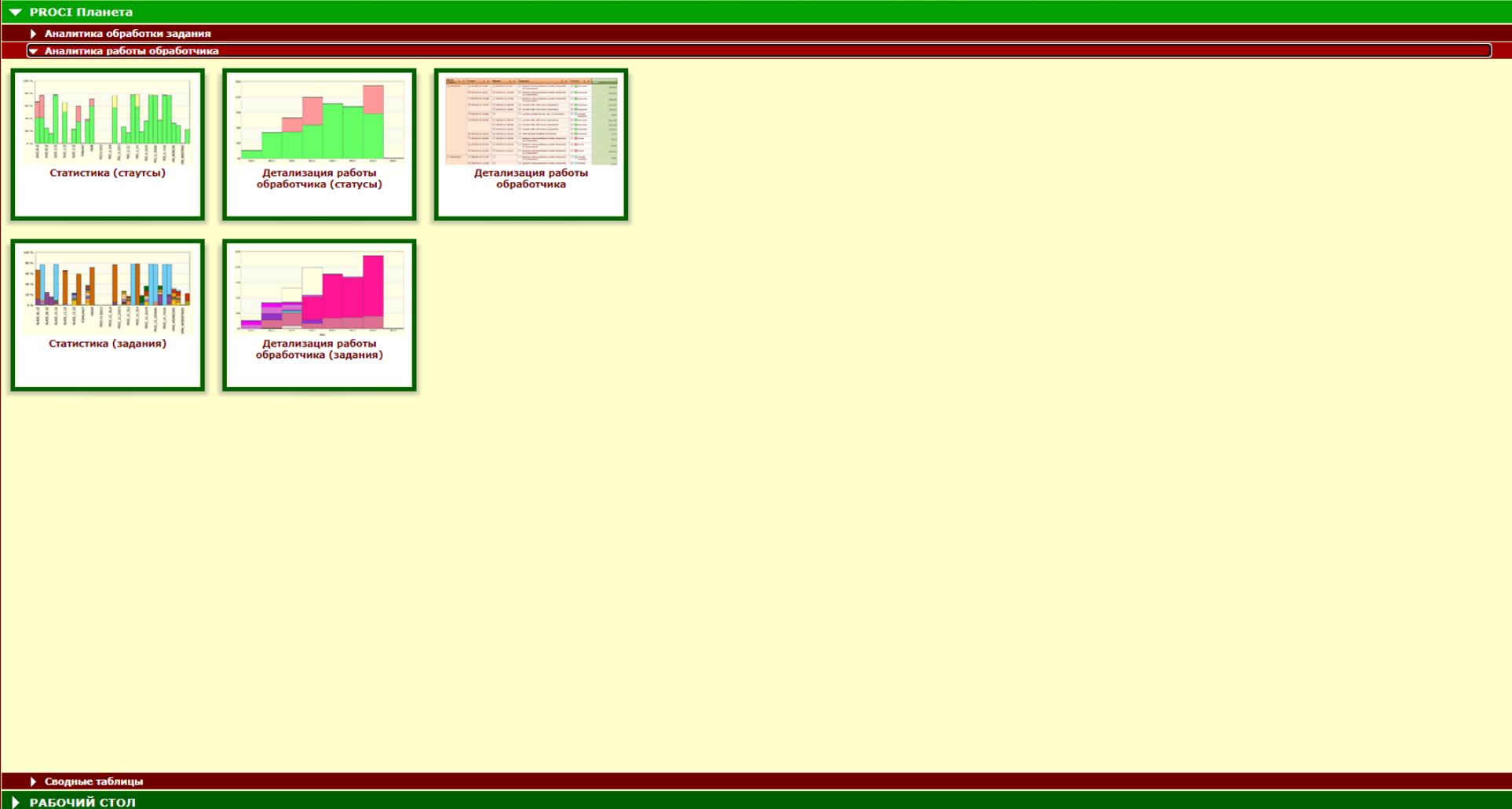
Сводные таблицы

РАБОЧИЙ СТОЛ

Доступная статистика по типам обработки



BI-аналитика



Доступная статистика по станциям обработки



BI-аналитика

PROCI Планета

РАБОЧИЙ СТОЛ

Статистика

Параметры

Навигатор полей

Аналитика задания - Статистика

Статистика (с 18 ноября 2020 до 19 ноября 2020)

Статистика (с 18 ноября 2020 до 19 ноября 2020)

Группы заданий	Задания	.rdy	Удачных	Продолжительность сессий [мин]	Среднее количество потоков в сессии	Среднесессионная продолжительность потока [мин]	Процесс	Ошибки	Crash	Пустых
регулярные-приоритетные	canopus mss products (session)		1	7	1	7,5				
	canopus pss products (session)						1			
	kmss rgb (session)		33	1 834	1	55,6				
	kmss rgb m22 (session)	18	25	971	1	38,8				
	m22 mtvza process (session)		8	124	1	15,5				
	modis qlook (session)		17	13	1	0,8				3
	msu mr rgb united (session)		43	1 245	1	29,0				2
	msu mr rgb united (session) m22	1	3	11	1	3,6				
	ndvi smis modis proc (session)		17	162	1	9,5				
	night channels modis proc (session)		16	44	1	2,8		1		71
	npp channels (session)		11	244	1	22,2				
	sst smis proc (session)		17	64	1	3,7				2
Итого			191	4 720	1	24,7	1	1		78

Пример статистики по выполненным заданиям обработки



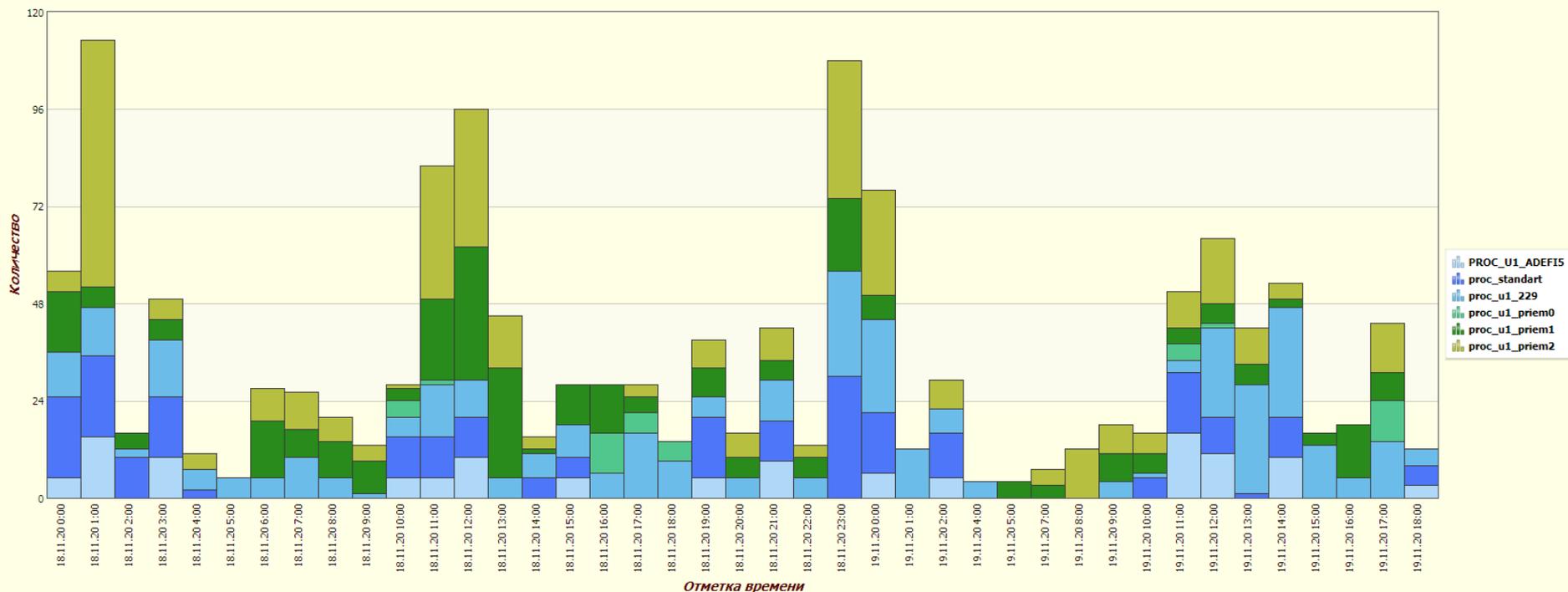
BI-аналитика

Аналитика задание - детализация граф. (обработчики, по часам)

Количество обработанных наборов - (всего 1 395 потоков с 18 ноября 2020 по 19 ноября 2020 на 6 обработчиках)

Обработчики Отметка времени ID сессии ID потока Имя набора Статус

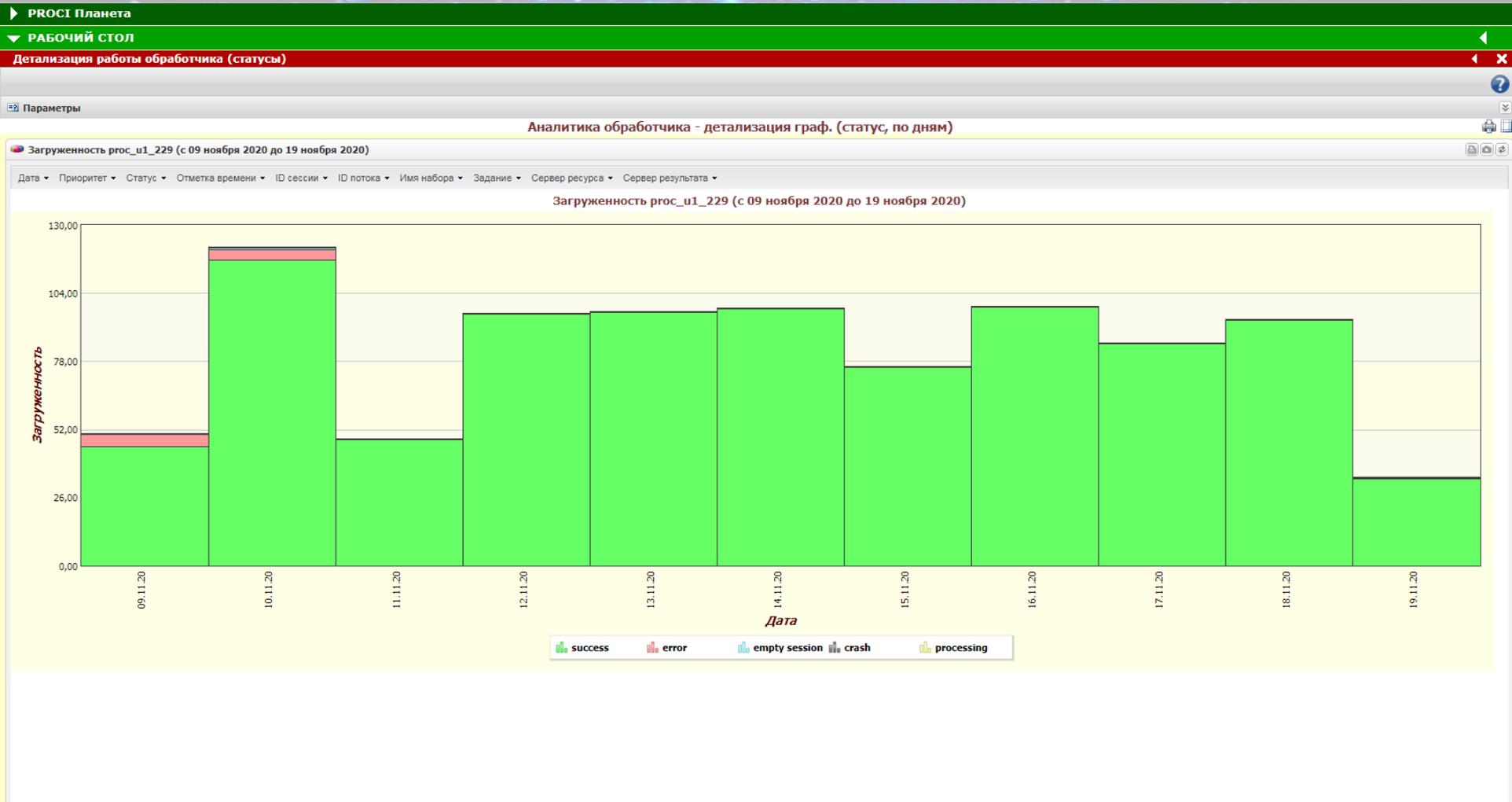
Количество обработанных наборов - (всего 1 395 потоков с 18 ноября 2020 по 19 ноября 2020 на 6 обработчиках)



Пример статистики по суммарной загрузке станций обработки



BI-аналитика



Пример статистики по распределению загрузки станции обработки

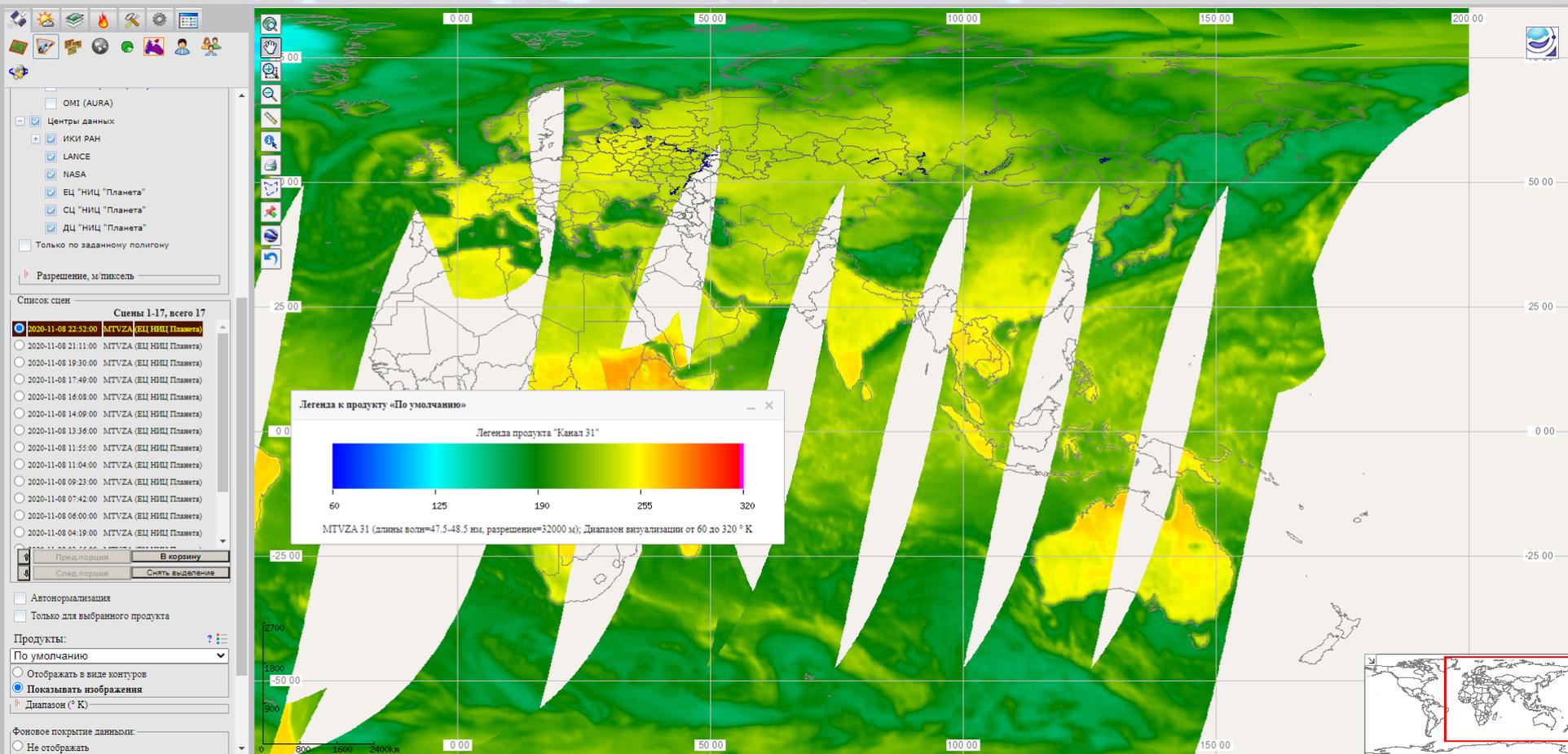


Обрабатываемые данные

- NOAA / AVHRR;
- TERRA, AQUA / MODIS, AIRS;
- Suomi NPP, NOAA 20 / VIIRS, CRiS;
- «Метеор-М» / МСУ-МР;
- «Метеор-М» / КМСС;
- «Метеор-М» / МТВЗА;
- Геостационарные КА («Электро-Л», Meteosat, GOES, Himawari);
- «Канопус-В» №1, 3, 4, 5, 6, Канопус-В-ИК;
- «Ресурс-П» №1, 2, 3;



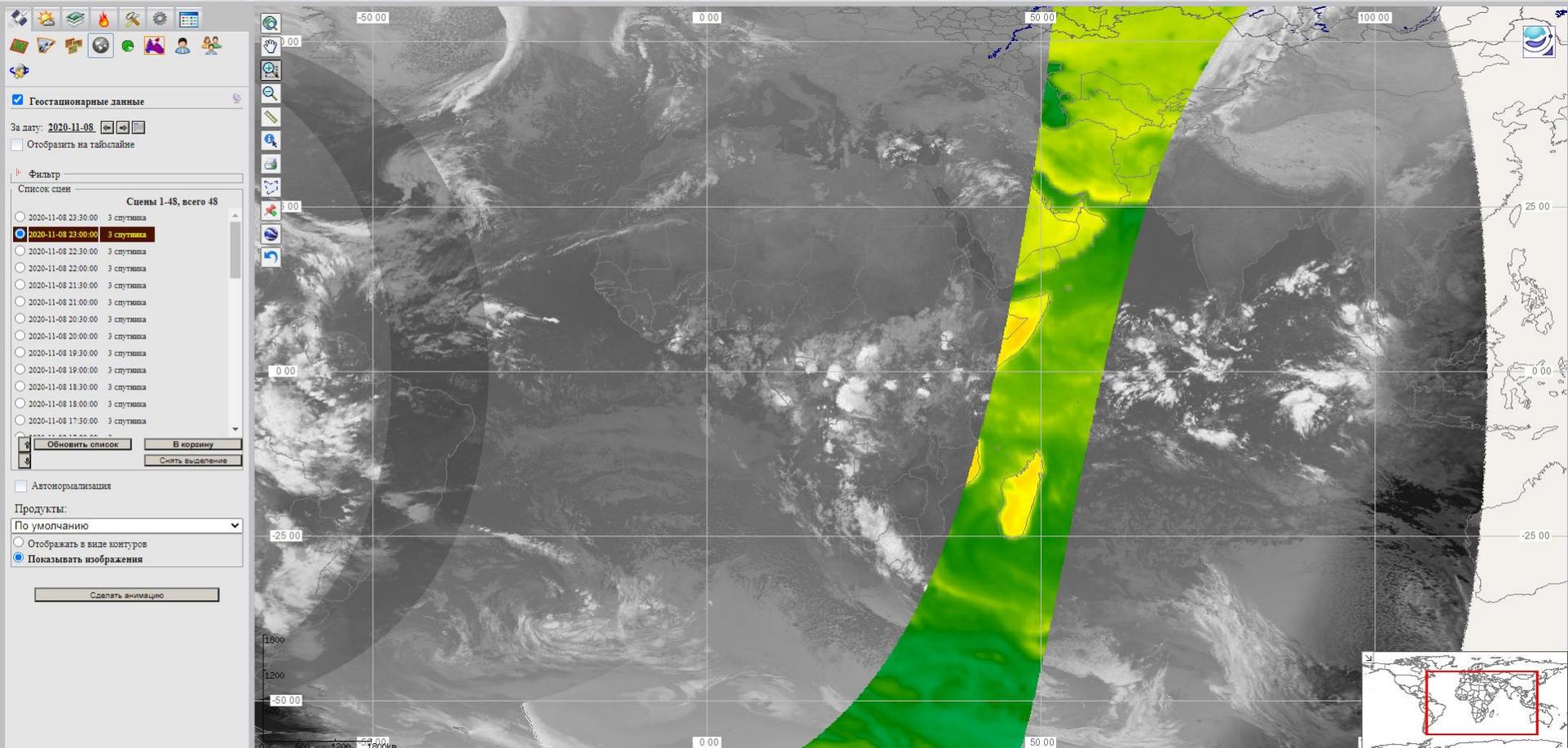
Примеры результатов



Динамическое покрытие данными MTVZA, «Метеор-М» №2-2 за 08.11.2020



Примеры результатов



Пример совмещения данных МТВЗА, «Метеор-М» №2-2 и Meteosat 11/SEVIRI, 23:00 UTC 08.11.2020

Результаты

- Приведение всех данных во всех центрах к единому стандарту и виду;
- Обеспечение возможности динамической генерации тематических продуктов на лету;
- Обеспечение возможности онлайн-обработки и анализа данных;
- Гибкая настройка всех процедур обработки и её инфраструктуры;
- Детальный контроль и анализ состояния системы обработки для парирования ошибок и оптимизации.



***Спасибо за
внимание!***

