

# Универсальная система управления комплексными задачами обработки спутниковых данных TASK JOBS

**Прошин А.А., Марченко В.В., Руткевич Б.П.**

Институт космических исследований Российской академии наук, Москва

Современные проблемы дистанционного зондирования Земли из космоса, 14-18 ноября 2022г

Эффективность любых систем обработки данных во многом зависит от того, насколько оптимально реализовано управление запусками процедур обработки и контроль за их выполнением. Естественно, это справедливо и для систем обработки данных ДЗЗ, которые, как правило, оперируют большими объемами информации, что требует весьма значительных вычислительных ресурсов. Конкретная реализация системы управления заданиями по обработке спутниковых данных во многом зависит от специфики выполняемых задач, используемого программного обеспечения и операционных систем, установленных на серверах обработки.

В рамках функционирования ЦКП «ИКИ-Мониторинг» в настоящее время используется ряд систем управления обработкой. Основной из них является разработанная в отделе «Технологии спутникового мониторинга» система управления многопоточной обработкой спутниковых данных на кластере windows серверов, который в настоящее время включает в себя около 100 компьютеров и виртуальных машин. Отдельная система управления заданиями реализована для поддержки работы интерактивных инструментов обработки и анализа спутниковых информации, реализованных в рамках веб интерфейсов для доступа к данным ЦКП «ИКИ-Мониторинг». Кроме этого в настоящее время используется ряд частных решений для управления специфическими задачами по обработке данных, в частности, для получения информации по объектам мониторинга, выгрузке данных из больших архивов и др. Наличие ряда различных программных компонент, предназначенных для решения во многом схожих задач, безусловно, усложняет их поддержку и развитие. Поэтому актуальной для отдела «Технологии спутникового мониторинга» ИКИ РАН задачей является разработка более универсальных решений, позволяющих управлять широким спектром различных типов обработки спутниковой информации.

## Основные возможности системы TASK JOBS

В настоящем докладе представлена новая универсальная система управления комплексными задачами обработки данных TASK JOBS. Ключевым ее преимуществом является возможность управления составными задачами по обработке, как спутниковых данных, так и другой информации. Типичным примером такого рода задач является построение композитных спутниковых изображений по большим территориям, фрагменты которых формируются в параллельном режиме на кластере серверов обработки. Также к таким задачам можно отнести и любые другие обработки, которые могут быть выполнены в параллельном режиме и для которых важно получить итоговый результат в заданное время. Таким образом, основной особенностью представленной системы управления является возможность отслеживания не только отдельных заданий по обработке данных, но и их комплектов.

Также при разработке системы TASK JOBS большое внимание было уделено гибкости ее настроек, что во многом обеспечивается использованием JSON структур для описания как отдельных заданий по обработке, так и задач, в которые они входят. Это позволяет использовать эту систему для широкого спектра различных задач. В частности, наряду с выполнением комплексных (составных) задач обработки данных, поддерживается также режим с потоковым поступлением новых заданий по обработке спутниковых данных. А в тех случаях, когда результатом обработки отдельного задания является относительно компактная текстовая информация, он может быть сохранен в самой системе управления обработкой. Таким образом, при создании рассматриваемой системы ставилась задача предоставления разработчикам программной среды для организации самых разных вариантов управления задачами обработки.

Для взаимодействия с системой в автоматизированном режиме реализован специальный программный интерфейс, а также интерфейс REST API, позволяющий минимизировать требования на программное обеспечение, устанавливаемое на сервера обработки. Также разработан служебный веб-интерфейс, позволяющий специалистам конфигурировать задачи и контролировать их выполнение в наглядной форме.

## Архитектура построения системы TASK JOBS

В рамках реализации системы TASK JOBS используются три основных типа объектов:

- **Тип задач**

Разработчик может сам решить, что будет являться типом задачи в каждом конкретном случае, но по логике системы тип задач соответствует типу обработки или конкретному программному пакету. Например, тип задач может соответствовать конкретной процедуре для построения временного ряда композитных изображений (изображений состоящих из заданного набора фрагментов).

- **Задача**

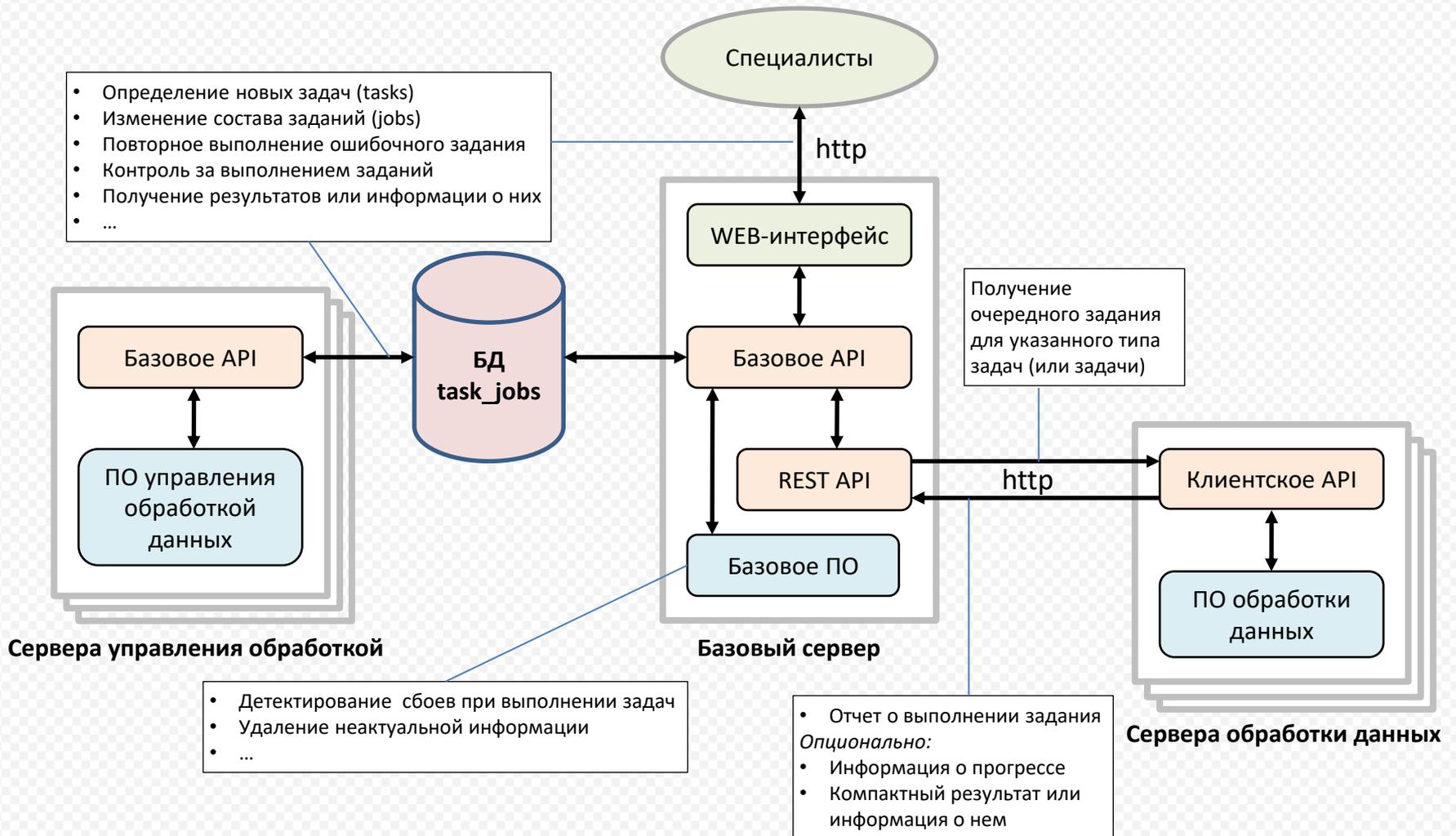
В базовом варианте под задачами понимаются описания комплексных процедур обработки данных, соответствующих заданному типу задач. В случае построения композитных изображений, задача может соответствовать построению композитного изображения за заданный временной интервал, указываемый в ее параметрах. Хотя задачи также могут определяться специалистом при помощи веб-интерфейса, основным вариантом их конфигурации

- **Задание**

Задания соответствуют выполнению «элементарных» процедур обработки на отдельных серверах обработки и являются элементом задачи. Так при построении композитного изображения, задания соответствуют процедурам получения отдельных его фрагментов, описания которых определяются в его параметрах.

Конфигурация новых типов задач, самих задач и соответствующих им заданий может быть произведена разработчиками при помощи специализированного веб-интерфейса. Но основным вариантом является автоматическая их генерация при помощи реализованного программного интерфейса, устанавливаемого на серверах управления обработкой. На серверах обработки устанавливается «легковесное» клиентское API, позволяющее им получать очередные задания на обработку и отчитываться о ходе их выполнения.

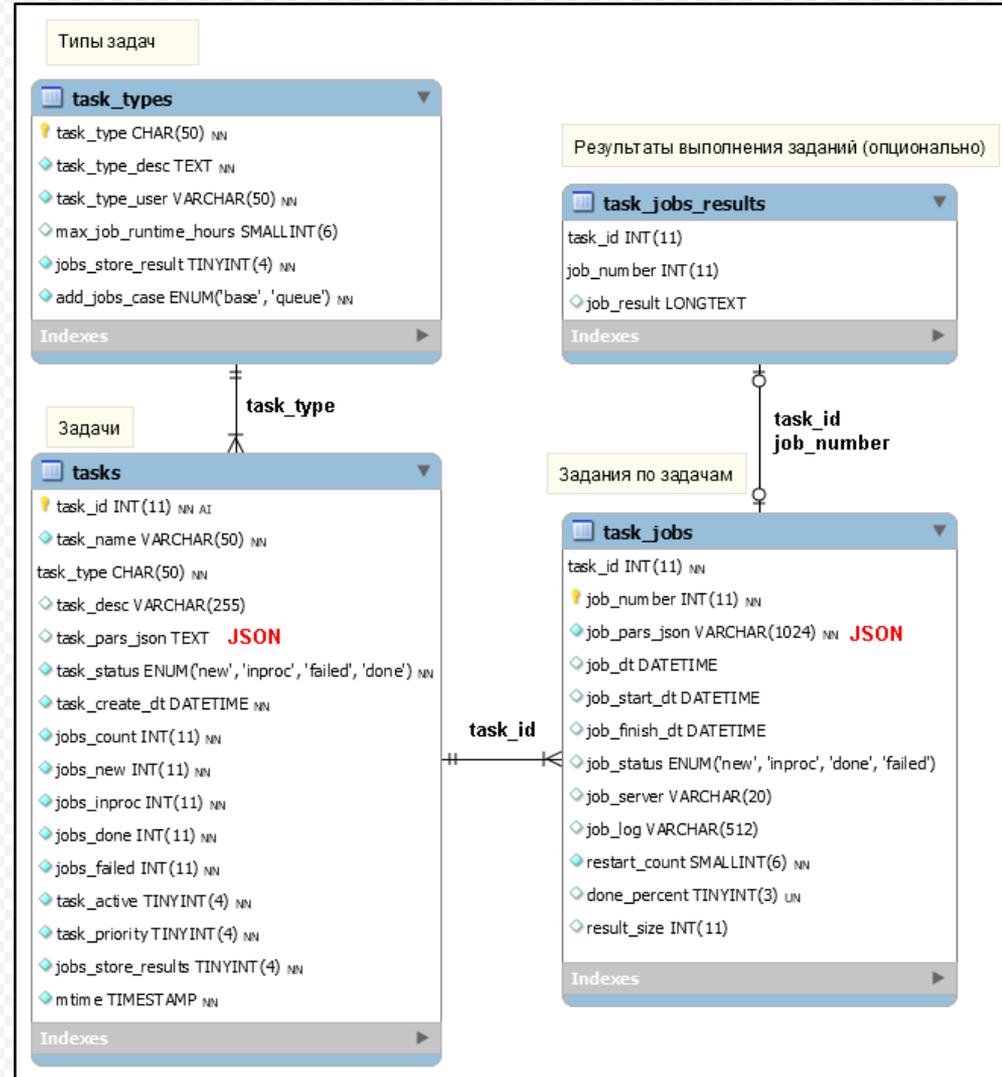
# Принципиальная схема функционирования системы TASK JOBS



## Структура БД task\_jobs

Централизованное хранение информации в системе **TASK JOBS** реализовано на основе использования БД **task\_jobs**, в структуру которой входят четыре основных таблицы:

1. таблица **task\_types**, описывающая типы задач и общие для их выполнения параметры;
2. таблица **tasks** содержит параметры задач, а также целый ряд показателей, описывающих ход их выполнения;
3. таблица **task\_jobs** содержит конфигурацию отдельных заданий, а также информацию об их выполнении;
4. таблица **task\_jobs\_results** позволяет сохранять в БД текстовые результаты выполнения заданий.



# Пример использования веб-интерфейсы системы TASK JOBS

## [TASK\_JOBS] Интерфейс системы управления выполнением комплексных задач обработки данных

Выбор типа задач:

### Тип задач (task\_types)

Тип задач task_type	Описание task_type_desc	Ответственный task_type_user	Максимальное время выполнения задания max_job_runtime_hours	Задания сохраняют результат в БД jobs_store_results	Вариант добавления заданий add_jobs_case	Изменить	Удалить
web_data_queue	Очередь заданий для получения данных в WEB интерфейсы	Прошин	1	1	queue		-

Выбор задачи:

### Задача (tasks)

ID task_id	Название task_name	Описание task_desc	Параметры task_pars_json	Время создания task_create_dt	Статистика по заданиям					Приоритет task_priority	Статус task_status	Активность task_active	Изменить	Удалить
					jobs_count	jobs_new	jobs_inproc	jobs_done	jobs_failed					
16	get_unisat_point_iki	Получение рядов данных по точке (ИКИ)		2022-09-20 15:15:53	290	0	1	182	107	1	failed	активна		-

### Задания (task\_jobs)

Фильтр по статусу:

Номер job_number	Параметры job_pars_json	Добавление job_dt	Начало job_start_dt	Окончание job_finish_dt	Сервер job_server	Протокол (ошибка) job_log	Перезапусков restart_count	Статус job_status	Процент вып. done_percent	Результат	Время выполнения	Отметить
293	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-11-02 16:16:57"	2022-11-02 16:16:57	2022-11-02 16:17:01	2022-11-02 16:17:07	192.168.200.61	-	0	done	-	2.7Kb	6 sec	<input type="checkbox"/>
292	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-27 10:20:00", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-11-02 14:12:24	2022-11-02 18:16:48	2022-11-02 18:16:50	192.168.200.61	-	1	inproc	4	-	2 sec	<input type="checkbox"/>
291	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-11-02 13:18:41", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-11-02 13:18:41	2022-11-02 16:04:49	2022-11-02 16:04:53	192.168.200.61	-	24	done	-	2.7Kb	4 sec	<input type="checkbox"/>
290	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-11-02 10:43:02", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-11-02 10:43:02	2022-11-02 16:04:37	2022-11-02 16:04:49	192.168.200.61	-	1	done	-	6Kb	12 sec	<input type="checkbox"/>
289	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-11-02 10:42:51", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-11-02 10:42:51	2022-11-02 10:43:00	2022-11-02 10:43:06	192.168.200.61	-	0	done	-	2.5Kb	6 sec	<input type="checkbox"/>
288	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-11-02 10:41:32", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-11-02 10:41:32	2022-11-02 18:18:02	2022-11-02 18:19:04	192.168.200.61	-	1	done	-	2.5Kb	1 min 2 sec	<input type="checkbox"/>
287	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-31 09:35:45", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-10-31 09:35:45	2022-11-02 18:18:02	2022-11-02 18:18:58	192.168.200.61	-	2	done	-	3.3Kb	56 sec	<input type="checkbox"/>
286	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-28 13:39:53", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-10-28 13:39:53	2022-11-02 18:18:02	2022-11-02 18:18:50	192.168.200.61	-	1	done	-	3.5Kb	48 sec	<input type="checkbox"/>
285	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-28 13:39:48", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-10-28 13:39:48	2022-11-02 18:18:01	2022-11-02 18:18:44	192.168.200.61	-	3	done	-	2.6Kb	43 sec	<input type="checkbox"/>
284	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-28 13:39:45", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-10-28 13:39:45	2022-11-02 18:18:01	2022-11-02 18:18:39	192.168.200.61	-	11	done	-	2.7Kb	38 sec	<input type="checkbox"/>
283	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-28 13:19:09", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-10-28 13:19:09	2022-11-02 18:18:01	2022-11-02 18:18:34	192.168.200.61	-	5	done	-	12Kb	33 sec	<input type="checkbox"/>
282	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-27 13:33:04", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-10-27 13:33:04	2022-11-02 13:11:00	2022-11-02 13:11:04	192.168.200.61	-	4	done	-	781b	4 sec	<input type="checkbox"/>
281	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-27 13:32:29", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-10-27 13:32:29	2022-11-02 18:18:01	2022-11-02 18:18:08	192.168.200.61	-	1	done	-	781b	7 sec	<input type="checkbox"/>
280	{ "DB_PKG_MODE": "comp", "device": "MODIS", "dt": "2022-10-27 13:32:08", "dt_from": "2022-10-25", "object_type_id": "point", "product": [ "comp_ndvi_7dc" ] }	2022-10-27 13:32:08	2022-11-02 18:18:00	2022-11-02 18:18:05	192.168.200.61	-	1	done	-	781b	5 sec	<input type="checkbox"/>

## Сведения о реализации

- Основной код программной реализации системы TASK JOBS разработан под операционной системой UNIX на языке программирования Perl. В качестве СУБД используется MySQL, а качестве веб-сервера - Apache
- Реализовано два варианта клиентского API, соответственно на языках программирования Python и Perl, которое может использоваться как для UNIX, так и для Windows серверов. Существенно, что клиентское API для взаимодействия с реализованным REST API является достаточно простым, что позволяет легко адаптировать его под любую программную платформу

## Заключение

В настоящее время система TASK JOBS внедрена для решения ряда задач, связанных с получением долговременных рядов данных по спутниковым наблюдениям, которые используются в подсистеме анализа данных в виде графиков. В перспективе планируется использовать ее для всех специфических задач по обработке данных.

Работы по развитию технологий управления обработкой данных сверхбольших архивов спутниковой информации с 2019 года выполняются в рамках темы Минобрнауки РФ «Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера» (№122042500019-6) с использованием возможностей ЦКП «ИКИ-Мониторинг» (<http://ckp.geosmis.ru/>).