

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

СОГЛАСОВАНО

Генеральный директор
ЗАО «АйТи»


Бакиев О.Р.
2011 г.

УТВЕРЖДАЮ

Ректор НИУ ИТМО


Васильев В.Н.
2011 г.

МНОГОПРОФИЛЬНАЯ ИНСТРУМЕНТАЛЬНО-
ТЕХНОЛОГИЧЕСКАЯ ПЛАТФОРМА СОЗДАНИЯ
И УПРАВЛЕНИЯ РАСПРЕДЕЛЕННОЙ СРЕДОЙ
ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ CLAVIRE

ПРОГРАММНЫЙ КОМПОНЕНТ ХРАНЕНИЯ ПРОФИЛЕЙ
ИСПОЛНЕНИЯ WF CLAVIRE/PROVENANCE


ОПИСАНИЕ ПРОГРАММЫ

ЛИСТ УТВЕРЖДЕНИЯ

RU.СНАБ.80066-06 13 32-ЛУ

Представители
Организации-разработчика

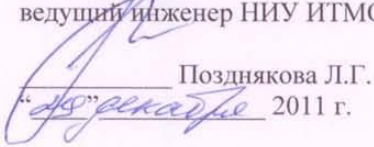
Руководитель разработки,
профессор НИУ ИТМО


Бухановский А.В.
"20" декабря 2011 г.

Ответственный исполнитель,
с.н.с. НИУ ИТМО


Луценко А.Е.
"20" декабря 2011 г.

Нормоконтролер
ведущий инженер НИУ ИТМО


Позднякова Л.Г.
"20" декабря 2011 г.

2011

Ине.№ подл.	Подп. и дата
Взам.ине.№	Ине.№ дубл.
Подп. и дата	Подп. и дата

УТВЕРЖДЕН
RU.СНАБ.80066-06 13 32-ЛУ

**МНОГОПРОФИЛЬНАЯ ИНСТРУМЕНТАЛЬНО-
ТЕХНОЛОГИЧЕСКАЯ ПЛАТФОРМА СОЗДАНИЯ
И УПРАВЛЕНИЯ РАСПРЕДЕЛЕННОЙ СРЕДОЙ
ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ CLAVIRE**

**ПРОГРАММНЫЙ КОМПОНЕНТ ХРАНЕНИЯ ПРОФИЛЕЙ
ИСПОЛНЕНИЯ WF CLAVIRE/PROVENANCE**

ОПИСАНИЕ ПРОГРАММЫ

RU.СНАБ.80066-06 13 32

ЛИСТОВ 12

Инв.№ подл.	Подп. и дата	Взам. инв.№	Инв.№ дубл.	Подп. и дата

АННОТАЦИЯ

Документ содержит описание программного компонента хранения профилей исполнения WF CLAVIRE/Provenance RU.СНАБ.80066-06 01 32. Программный компонент разработан в ходе выполнения проекта «Создание распределенной вычислительной среды на базе облачной архитектуры для построения и эксплуатации высокопроизводительных композитных приложений» (Договор № 21057 от 15 июля 2010 г., шифр 2010-218-01-209) в рамках реализации постановления Правительства РФ № 218 «О мерах государственной поддержки развития кооперации российских высших учебных заведений и организаций, реализующих комплексные проекты по созданию высокотехнологичного производства».

СОДЕРЖАНИЕ

2.	ОБЩИЕ СВЕДЕНИЯ	4
2.	ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ	4
3.	ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ	4
3.1.	Архитектура и принципы функционирования	4
3.2.	Основные классы	6
3.2.1.	Интерфейс IStorage	7
3.2.2.	Интерфейс IProvenanceService	7
4.	ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА	8
5.	ВЫЗОВ И ЗАГРУЗКА	8
6.	ВХОДНЫЕ ДАННЫЕ	9
7.	ВЫХОДНЫЕ ДАННЫЕ	9
	ПЕРЕЧЕНЬ СОКРАЩЕНИЙ	11

2. ОБЩИЕ СВЕДЕНИЯ

Программный компонент хранения профилей исполнения WF CLAVIRE/Provenance RU.СНАБ.80066-06 01 32 отвечает за накопление и хранение информации о работе платформы, а также предоставляет доступ к хранимой информации другим компонентам платформы.

Данный программный компонент разработан на языке C# 4.0 в виде web-сервиса с применением технологии WCF платформы .NET. Для функционирования сервиса необходимо наличие на целевой машине следующего программного обеспечения: .NET версии не ниже 4.0, Microsoft Internet Information Services (IIS), MongoDB с версией не ниже 1.6.5.

2. ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ

Программный компонент предназначен для хранения разнородной информации о работе системы и предоставления интерфейса для доступа к этой информации других компонентов платформы. Основным видом хранимой информации являются результаты и профили исполнения WF. Интерфейс компонента состоит из набора операций, предназначенных для сохранения информации, а также набора операций, позволяющих производить выборку необходимых данных по определенным критериям.

Хранимую информацию можно использовать для:

- 1) определения происхождения (provenance) тех или иных артефактов, полученных после исполнения WF (файлов, параметров);
- 2) отладки, профилировки платформы и выявления узких мест в ее работе;
- 3) статистического анализа использования ресурсов, сервисов, работы пользователей.

3. ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ

3.1. Архитектура и принципы функционирования

Программный компонент логически делится на два модуля: хранилище данных об исполнении WF, интерфейсный web-сервис доступа к хранилищу. В качестве хранилища данных используется внешняя документно-ориентированная БД на основе СУБД

RU.СНАБ.80066-06 13 32 **Ошибка! Источник ссылки не найден.**

MongoDB (компонент включает в себя только сервис доступа и работает поверх внешнего хранилища).

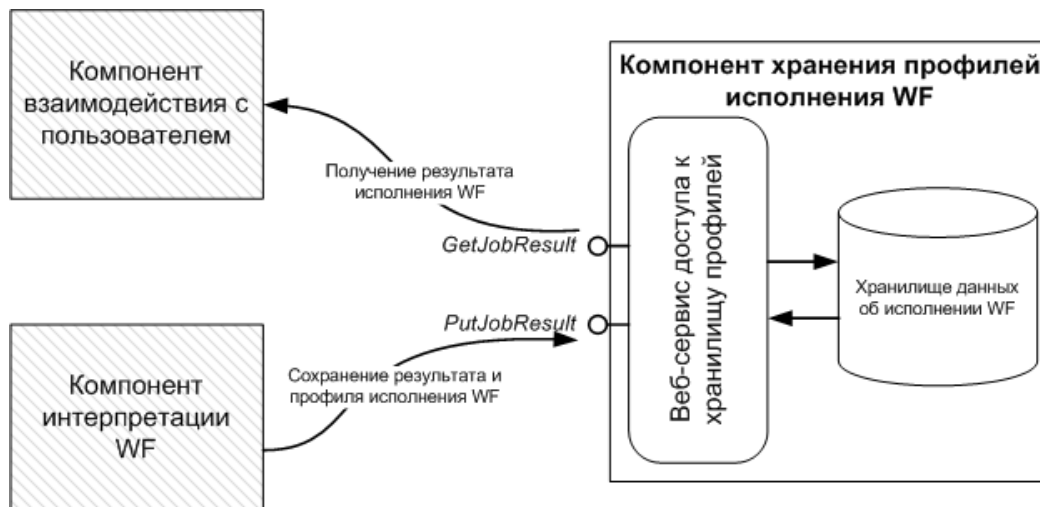


Рисунок 3.1 – Архитектура компонента

На рис. 3.1 представлена схема архитектуры компонента и приведено два варианта использования другими компонентами платформы. Web-сервис необходим для предоставления интерфейса для доступа другим компонентам платформы, а также для унификации структур данных об исполнении WF, которые используются между различными частями платформы. Хранилище данных используется для долгосрочного хранения и предоставления информации по требованию.

Профиль исполнения WF — это информационная структура, содержащая: результат выполнения WF; характеристики работы WF (такие как время выполнения отдельных шагов, временные издержки работы управляющей инфраструктуры при подготовке и исполнении WF); информацию об использованных аппаратных ресурсах; информацию о ходе исполнения WF. В основе схемы хранимой информации лежит объектная структура *JobDescription* (см. рис. 3.2), которая содержит в себе как информацию о результате выполнения WF, так и профилировочную информацию.

RU.СНАБ.80066-06 13 32 **Ошибка! Источник ссылки не найден.**

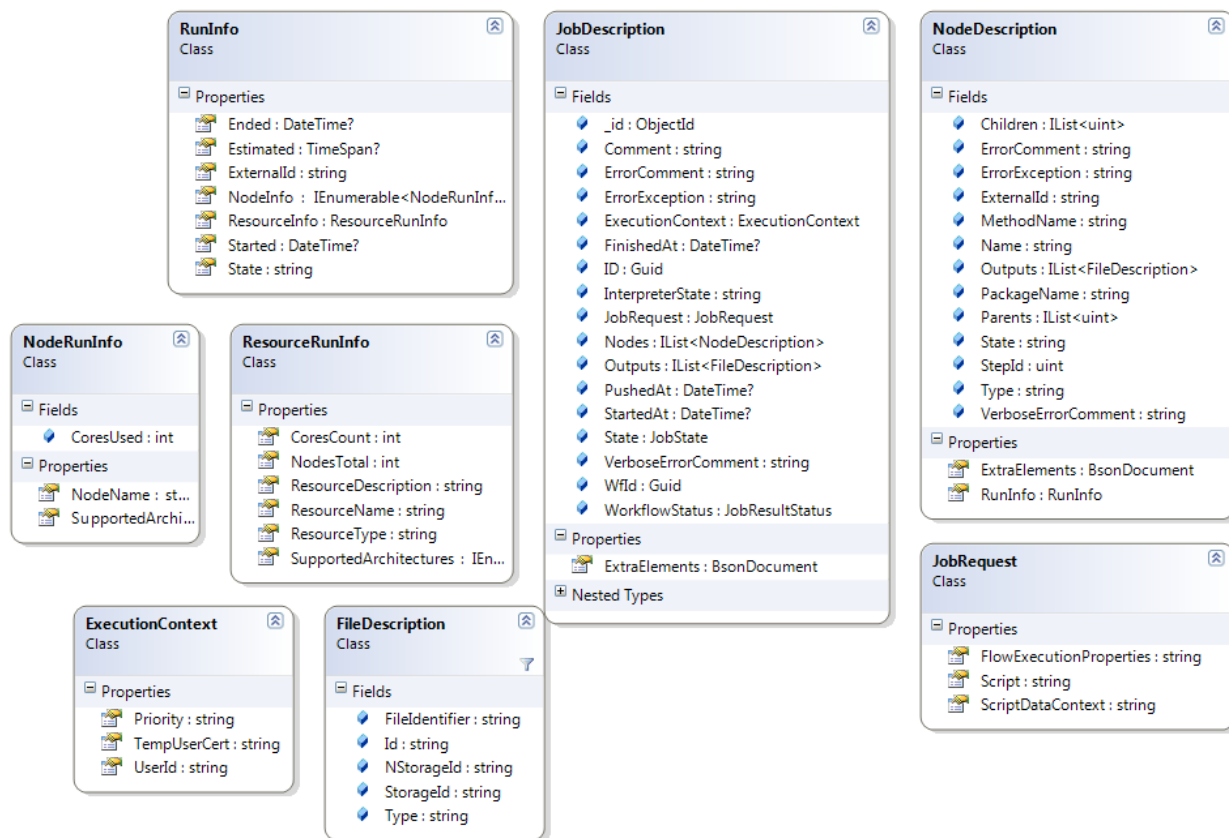


Рисунок 3.2 – Состав класса *JobDescription* и агрегированных им классов

Компонент Provenance используется:

- 1) компонентом взаимодействия с пользователем RU.СНАБ.80066-06 01 21 для отображения результатов выполнения WF;
- 2) компонентом исполнения WF RU.СНАБ.80066-06 01 29 для получения истории запусков задач, а также для предварительного поиска идентичных заданий из набора уже выполненных шагов (в качестве кэша результатов);
- 3) компонентом интерпретации WF RU.СНАБ.80066-06 01 20 для сохранения результатов и профиля исполнения WF.

Provenance позволяет использовать накопленную информацию системным администраторам для профилировки работы платформы, оптимизации, а также производить анализ статистических данных об использовании ресурсов, вычислительных пакетов, активности пользователей и т.д.

3.2. Основные классы

Ниже приводятся сокращенные описания структуры и методов основных классов компонента хранения профилей исполнения WF.

RU.СНАБ.80066-06 13 32 **Ошибка! Источник ссылки не найден.**

3.2.1. *Интерфейс IStorage*

Интерфейс хранилища данных компонента. Включает методы сохранения и извлечения данных.

Методы

- string [] GetCollectionNames () – получить список коллекций. Возвращает список коллекций.
- string GetDataInJson (StoragePath storagePath) – получение данных по пути хранения в виде JSON. Аргументы: storagePath – путь до объекта хранения.
- T GetDataObject< T > (StoragePath storagePath) – получение объекта из хранилища (десериализация). Аргументы: T – тип объекта, storagePath – путь до объекта хранения.
- void SetDataObject< T > (StoragePath storagePath, T obj, TimeSpan actualFor) – сохранение объекта в хранилище. Аргументы: storagePath – путь до объекта хранения, obj – объект, actualFor – до какого момента данный объект будет актуальным, T – тип объекта.
- void SetFields< T > (StoragePath storagePath, IDictionary< string, T > fields, TimeSpan actualFor) – установить значения полей объекта хранения. Аргументы: T – тип поля, storagePath – путь до объекта хранения, fields – словарь полей, actualFor – до какого момента информация будет актуальной.

3.2.2. *Интерфейс IProvenanceService*

Контрактный интерфейс WCF-сервиса компонента хранения профилей исполнения WF. Реализован в классе ProvenanceService.

Методы

- void PutJobResult(JobDescription jd) – сохранение профиля исполнения WF в формате JobDescription в хранилище. Аргументы: jd – информация об исполненном WF.
- void PutJobResultInBson(byte[] jr) – сохранение профиля исполнения WF в формате Bson в хранилище. Аргументы: jr – информация об исполненном WF в виде сериализованного объекта Bson (JobDescription).
- JobDescription GetJobResult(Guid jid) – получение информации об исполненном WF. Аргументы: jid – идентификатор WF.

4. ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА

Программный компонент Provenance реализован в виде web-сервиса платформы .NET 4.0 на языке C# с использованием технологии WCF (Windows Communication Foundation) на основе стандарта SOAP. Хостинг сервиса производится в IIS (Internet Information Services). Компонент предназначен для функционирования на аппаратных системах со следующими характеристиками:

- архитектура процессора – x86, x86_64, IA64;
- минимальный объем оперативной памяти – 1 ГБ;
- минимальный объем свободного пространства на жестком диске для – 30 ГБ (для хранилища);
- минимальная тактовая частота процессора – 2 ГГц.

Необходимо учитывать, что объем хранилища данных постоянно растет со временем использования платформы, поэтому при выборе аппаратной платформы важными являются характеристики, связанные с работой подсистемы хранения данных (скорость доступа к жесткому диску, объем свободного пространства на жестком диске, фрагментированность файловой системы). Динамическое увеличение объема хранилища поддерживается за счет масштабирования MongoDB, которое может быть произведено путем добавления новых узлов MongoDB (sharding).

5. ВЫЗОВ И ЗАГРУЗКА

Программный компонент реализован в виде SOAP WCF-сервиса платформы .NET. Для запуска сервиса используется стандартный механизм используемого web-сервера. Загрузка сервиса в этом случае выполняется web-сервером автоматически по мере поступления запросов от клиентов сервиса.

Provenance предоставляет интерфейс в виде WCF-сервиса, реализованного классом ProvenanceService (см. спецификацию IProvenanceService в разделе 3.2.2). Вызов сервиса производится стандартным для технологии WCF способом: по опубликованному описанию сервиса (WSDL) необходимо создать прокси-класс, через который осуществляется взаимодействие с сервисом путем вызова необходимых операций (методов). Процедура создания прокси-класса зависит от того, на базе каких технологий строится клиентское приложение. Если выбран язык программирования C# и платформа .NET, построение клиента производится за счет вызова служебного программного

RU.СНАБ.80066-06 13 32 **Ошибка! Источник ссылки не найден.**

средства svcutil.exe, распространяемого в составе платформы .NET, либо за счет создания ссылки на сервис в среде Microsoft VisualStudio.

В качестве хранилища данных веб-сервис использует нереляционную БД на основе СУБД MongoDB. Хранилище данных, входящее в состав компонента, может быть установлено удаленно по отношению к машине, на которой развернут интерфейсный сервис. Желательно (но необязательно) установить MongoDB в виде сервиса операционной системы. Для работы компонента не требуется предварительной настройки базы данных, кроме указания реквизитов доступа в конфигурационных файлах.

6. ВХОДНЫЕ ДАННЫЕ

Входными данными для компонента является адрес хранилища данных об исполнении WF. Адрес задается в стандартном для MongoDB виде – строке подключения (рис. 6.1).

```
<?xml version="1.0"?>
<configuration>
...
<applicationSettings>
  <Easis.Provenance.Properties.Settings>
    <setting name="ResultsConnectionString" serializeAs="string">
      <value>mongodb://localhost/results</value>
    </setting>
  </Easis.Provenance.Properties.Settings>
</applicationSettings>
</configuration>
```

Рисунок 6.1 – Фрагмент конфигурационного файла web.config, содержащий адрес хранилища данных

Входными данными со стороны других компонентов платформы являются данные в виде объектной структуры *JobDescription* (рис. 3.2).

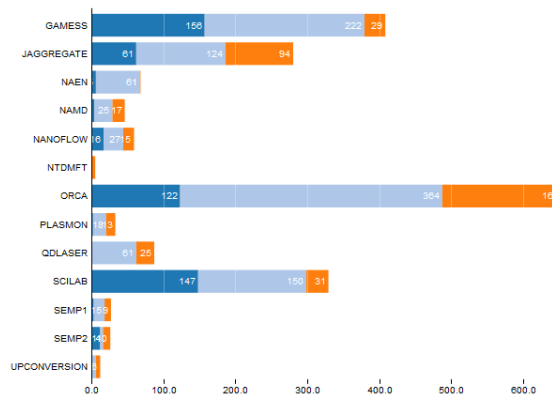
7. ВЫХОДНЫЕ ДАННЫЕ

Выходными данными Provenance являются объекты .NET, структура которых представлена на рис. 3.2. Используя данные, полученные от компонента, возможно получить необходимую статистическую информацию о работе платформе и, в частности, о запусках пакетов. Пример статистической обработки данных, хранящихся в компоненте, представлен на рис. 7.1.

RU.СНАБ.80066-06 13 32 Ошибка! Источник ссылки не найден.

Статистика запусков

name	success	failed	not finished
GAMESS	222	29	156
JAGGREGATE	124	94	61
NAEN	61	1	5
NAMD	25	17	3
NANOFLOW	27	15	16
NTDMFT	0	4	0
ORCA	364	166	122
PLASMON	18	13	1
QDLASER	61	25	0
SCILAB	150	31	147
SEMP1	15	9	2
SEMP2	4	10	11
UPCONVERSION	5	6	0



Синий - незавершенные по тем или иным причинам запуски,
 голубой - успешные запуски,
 оранжевый - запуски с ошибками.

Количество запущенных задач по дням

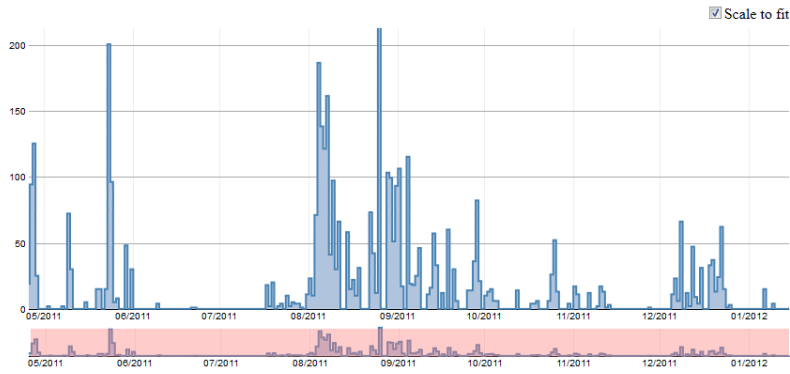


Рисунок 7.1 – Пример статистической обработки данных Provenance

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

SOAP	Simple Object Access Protocol (простой протокол доступа к объектам)
WCF	Windows Communication Foundation (программный фреймворк)
WF	Workflow (поток заданий)
БД	База данных
МИТП	Многопрофильная инструментально-технологическая платформа
СУБД	Система управления базами данных

