

Deductor 5.3: новые ВОЗМОЖНОСТИ



BaseGroup Labs
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

Направления развития

1. От отчетов к глубокому анализу
2. Управление качеством данных
3. Аналитика как сервис
4. Большие данные

Тенденции

1. Доступ к данным
2. Алгоритмы
3. Визуализация
4. Веб-сервисы
5. Масштабируемость



BaseGroup Labs
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

Доступ к данным

Новые источники



Microsoft®
Excel 2010

1С:ПРЕДПРИЯТИЕ 8

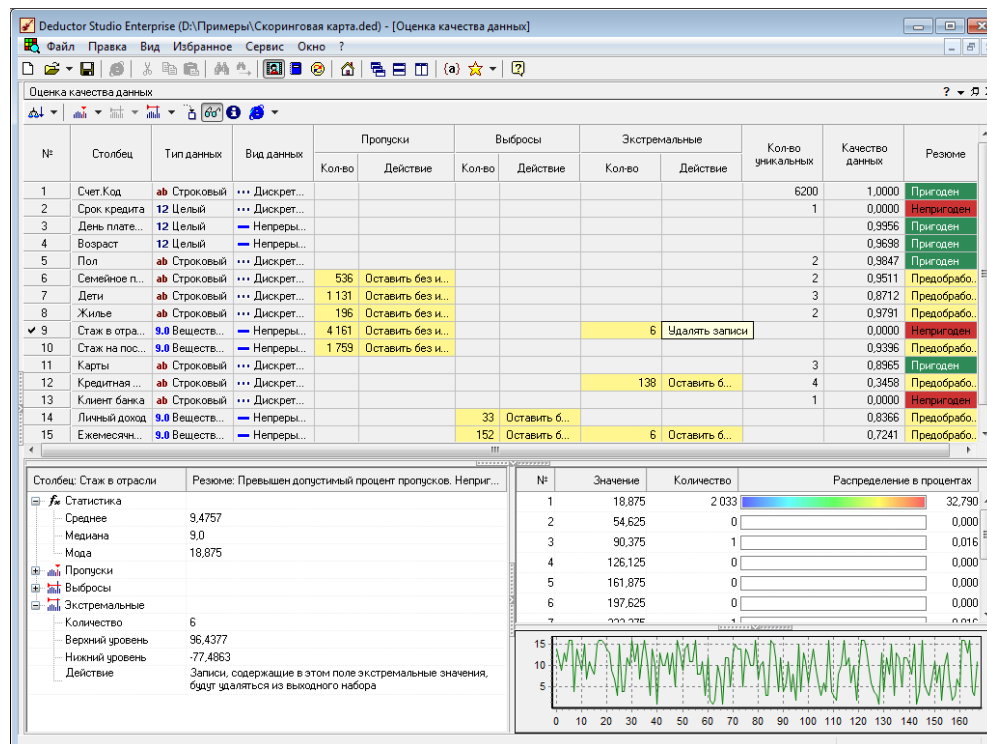




BaseGroup Labs
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

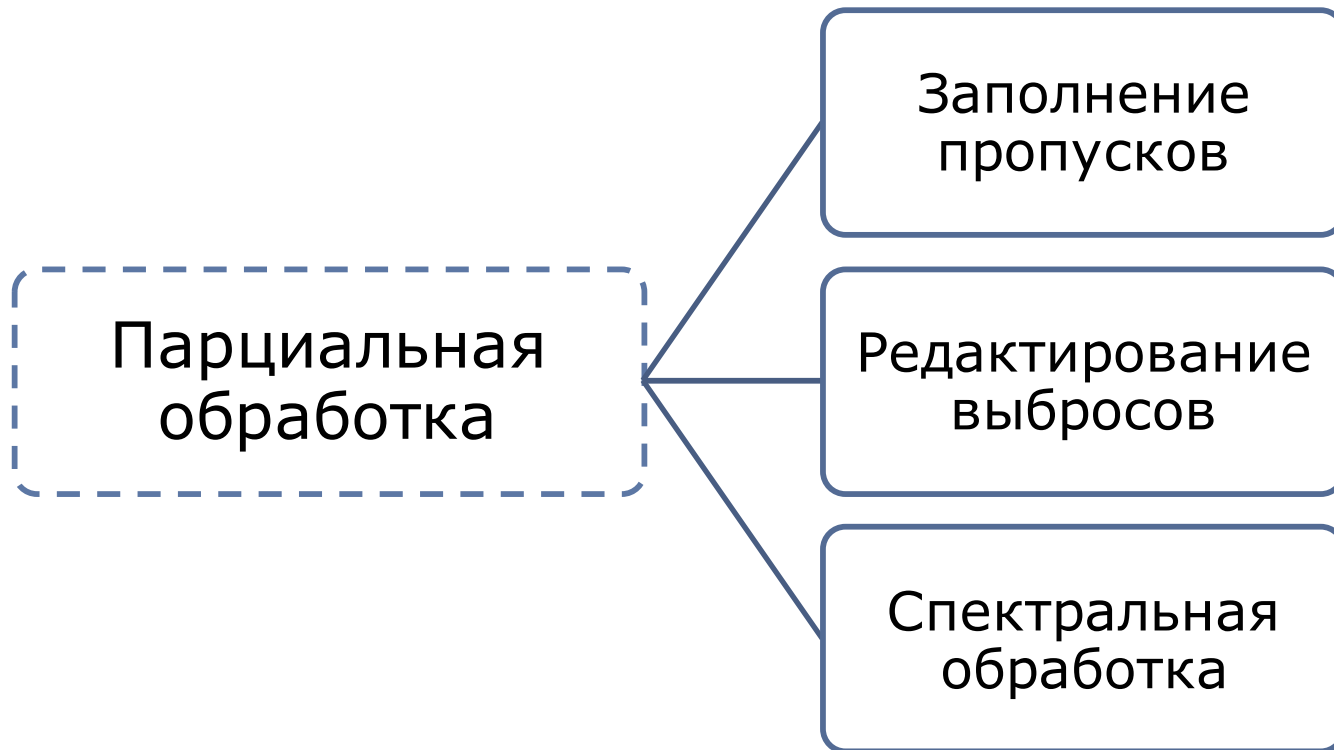
Алгоритмы обработки

Оценка качества данных

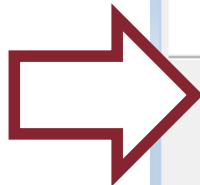




Очистка данных



Оценка качества + очистка



Мастер обработки - Редактирование выбросов (1 из 5)

Редактирование выбросов и экстремальных значений

Общие настройки набора данных

☐ Использовать данные узла оценки качества данных

☐ Обращивать как упорядоченный набор данных

Метод определения выбросов и экстремальных значений

☒ Стандартное отклонение от среднего

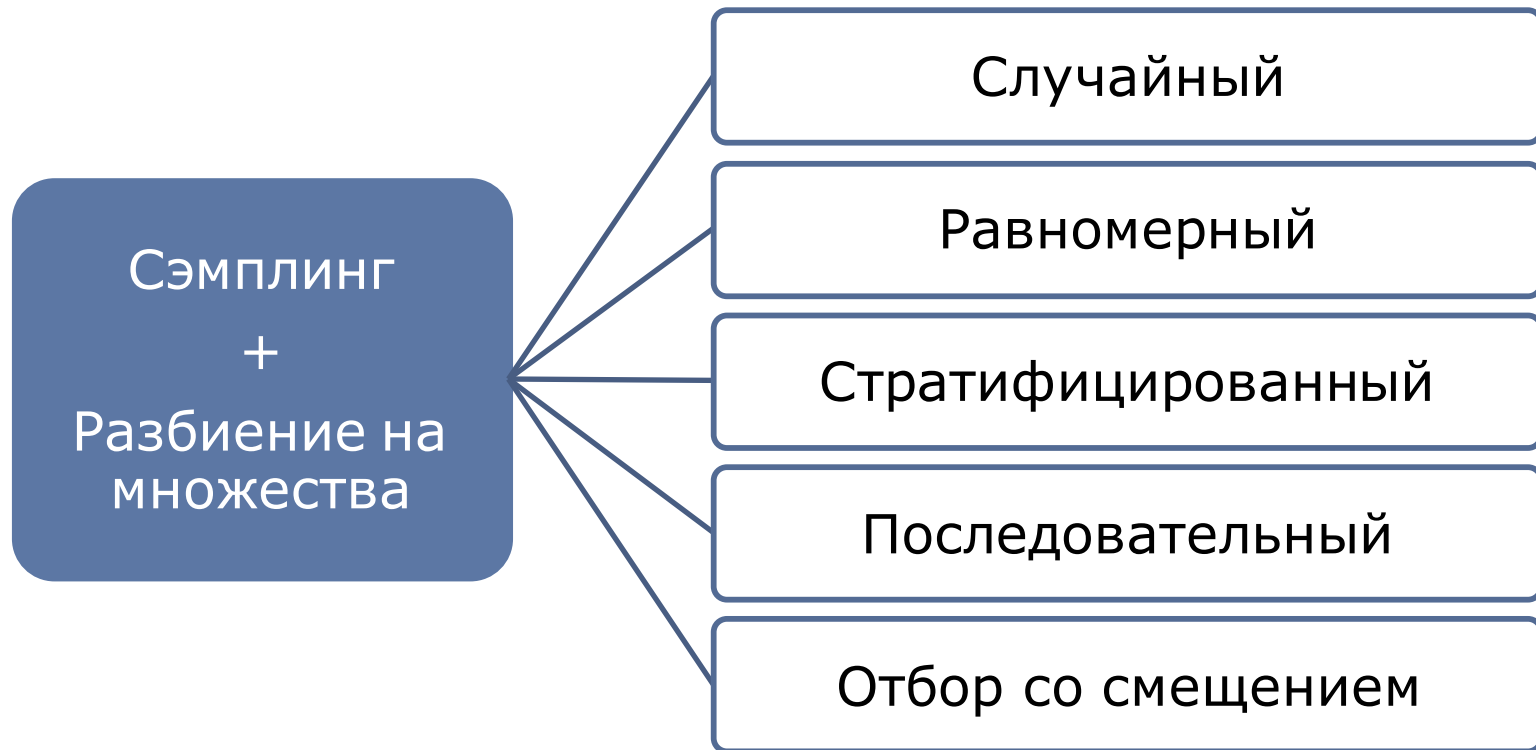
Для выбросов Для экстремальных значений

☐ Интерквартильная ширина

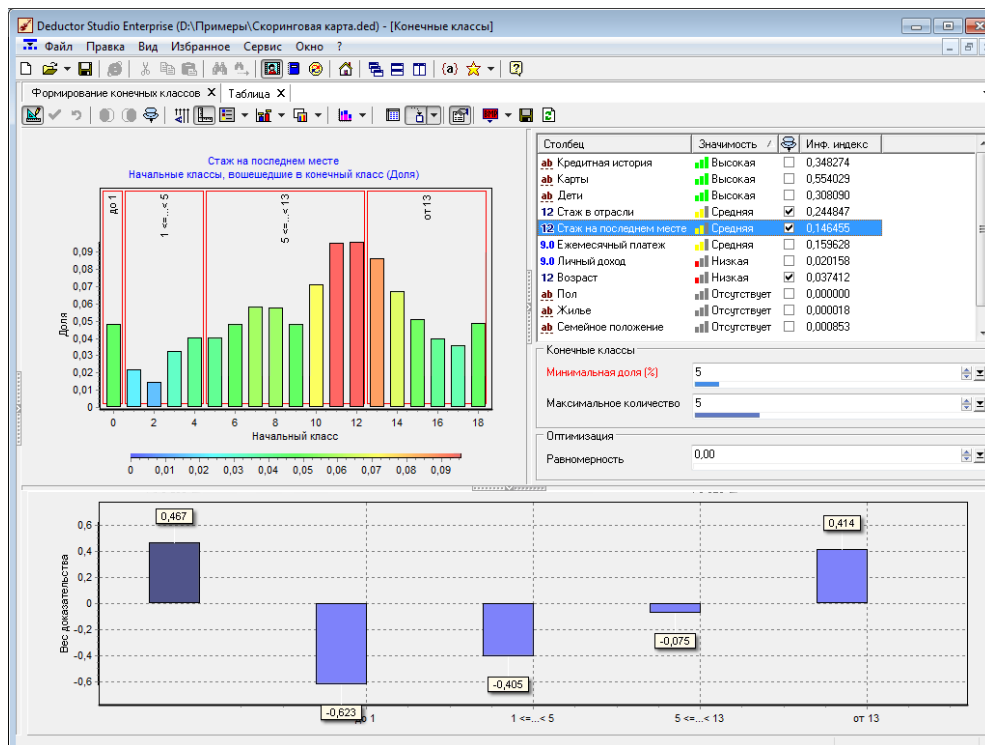
Для выбросов Для экстремальных значений

< Назад Далее > Отмена

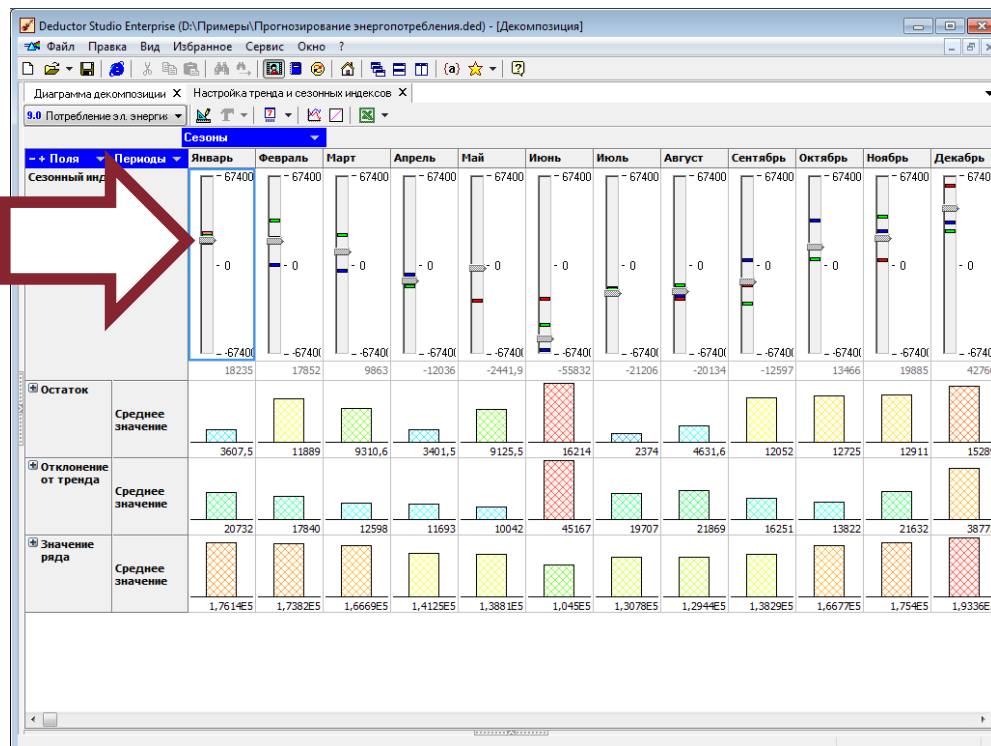
Формирование выборок



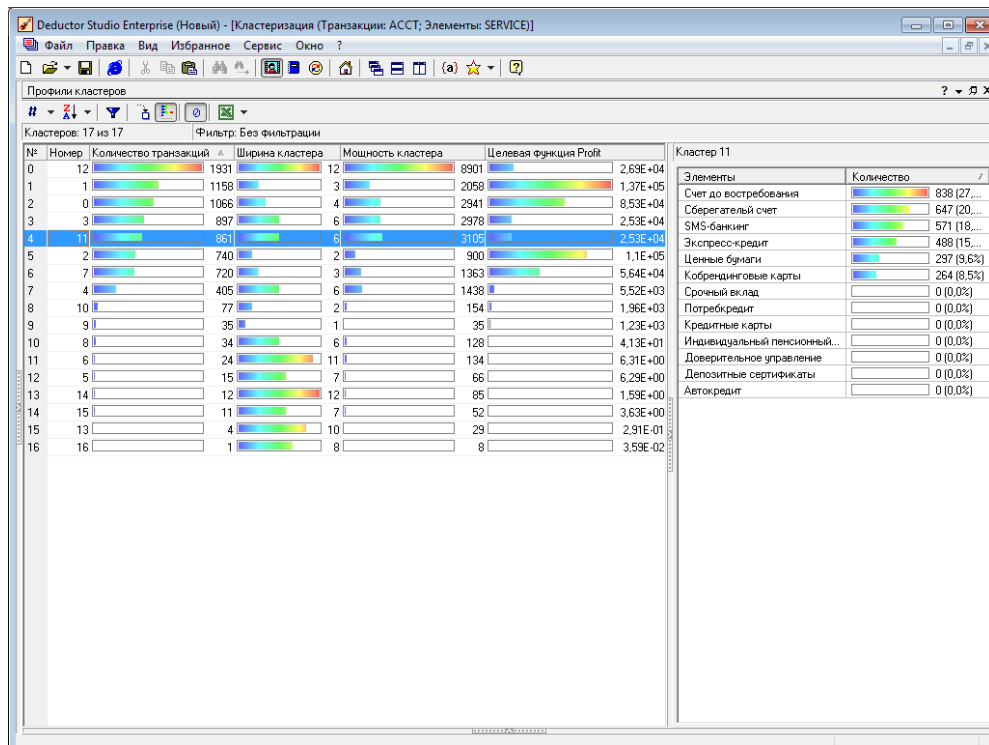
Конечные классы



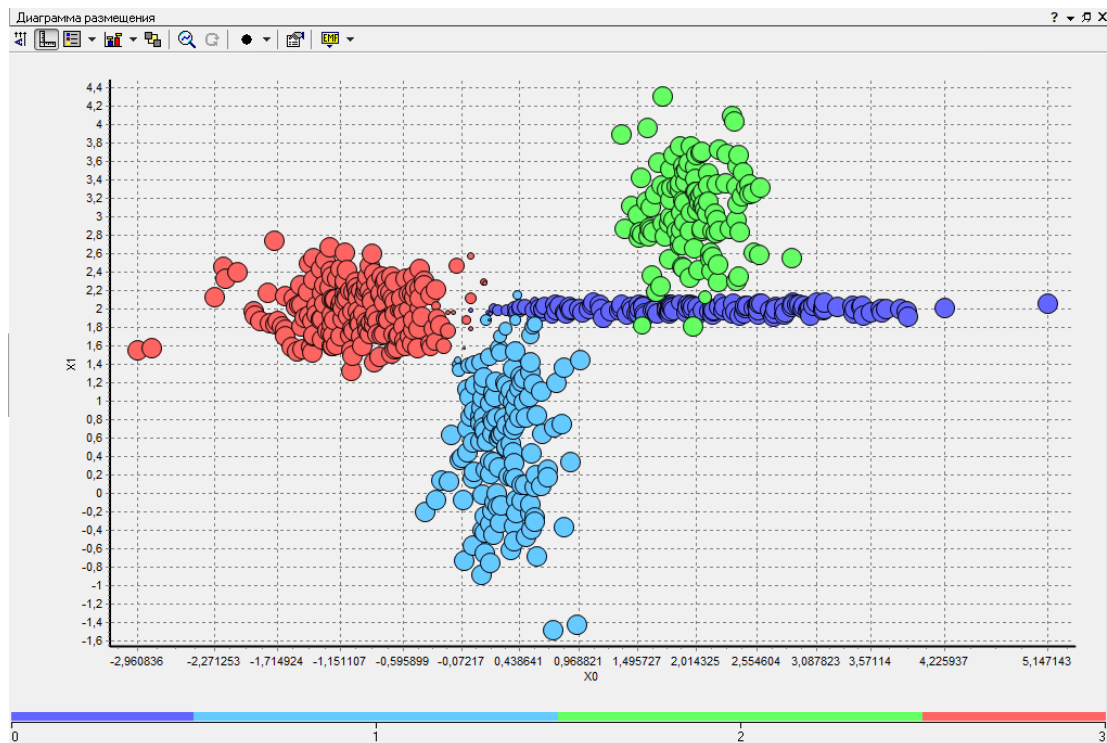
Декомпозиция временного ряда



Кластеризация транзакций



ЕМ-кластеризация



Улучшены обработчики

- Логистическая регрессия
- Факторный анализ
- Калькулятор
- Линейная регрессия
- Групповая обработка



BaseGroup Labs
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

Визуализация

Отчет по регрессии

Отчет по регрессии							
Модель		-2 Log Likelihood					
Базовая информация		Лог-регрессия "Финальная"					
★ Финальная		-2 Log Likelihood	R ² МакФадден	Хи-квадрат	Число степеней свободы	AIC	AICc
○ Нулевая		1 776,678	0,200	444,385	17	1 814,678	1 815,045
Шаги построения		2 221,063					Значимость
○ 1 (Нулевая)		2 221,063					Метод отбора переменных
							Полное включение
Коэффициенты лог-регрессии							
Фактор	Коэффициент	Стандартная ошибка	Коэффициент Вальда	Значимость	Отношение шансов	95% доверительный интервал отношения шансов	
						Минимум	Максимум
Константа	3,62233						
Возраст = 25 <=...< 40		0,0000			1,0	1,0	1,0
Возраст = 40 <=...< 50	0,08529	0,1486	0,3295	0,5659	1,08903	0,8139	1,45715
Возраст = до 25	-0,10957	0,2569	0,1820	0,6697	0,89622	0,54171	1,48274
Возраст = от 50	0,3471	0,1482	5,4852	0,0192	1,41496	1,05825	1,89191
Дети = [1; 2 и более]	0,0	0,0000			1,0	1,0	1,0
Дети = [отсутствуют; Класс пропусков]	-0,92723	0,1515	37,4432	0,0000	0,39565	0,29399	0,53247
Стаж в отрасли = 5 <=...< 10	0,0	0,0000			1,0	1,0	1,0
Стаж в отрасли = Класс пропусков	-0,41106	0,2298	3,1987	0,0737	0,66295	0,42251	1,04021
Стаж в отрасли = до 5	-0,45228	0,2836	2,5432	0,1108	0,63618	0,3649	1,10914
Стаж в отрасли = от 10	0,28978	0,3010	0,9271	0,3356	1,33614	0,74074	2,41011

Качество классификации

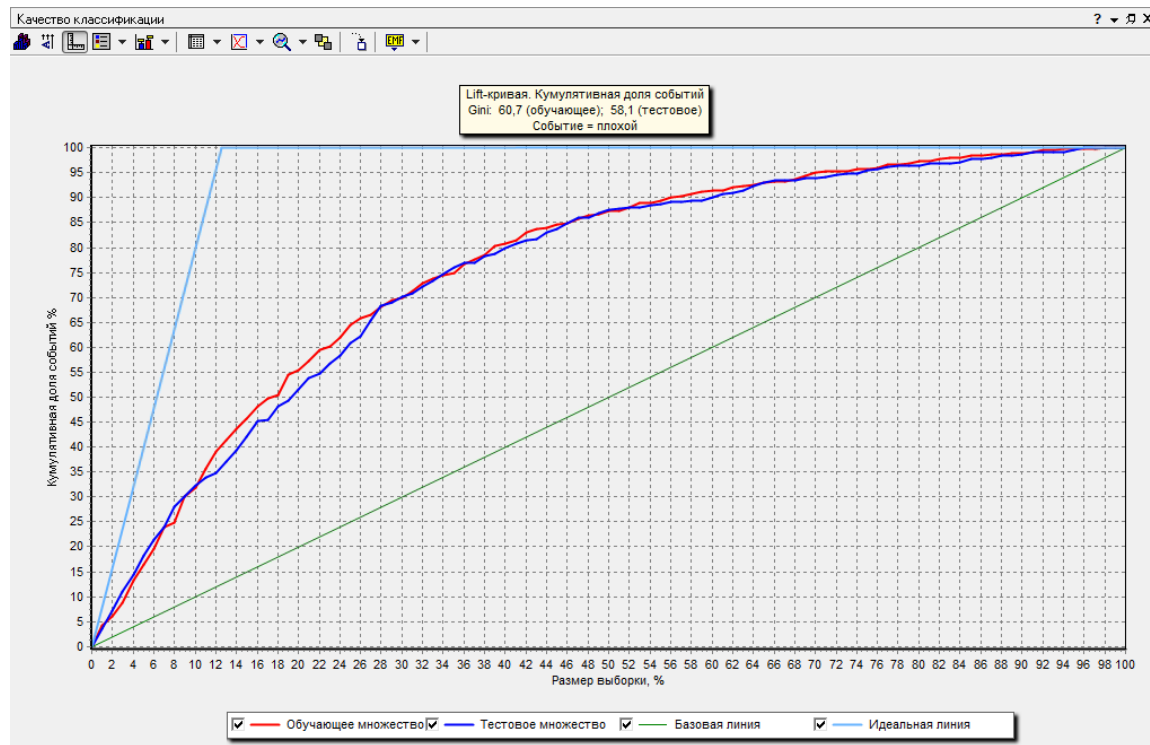
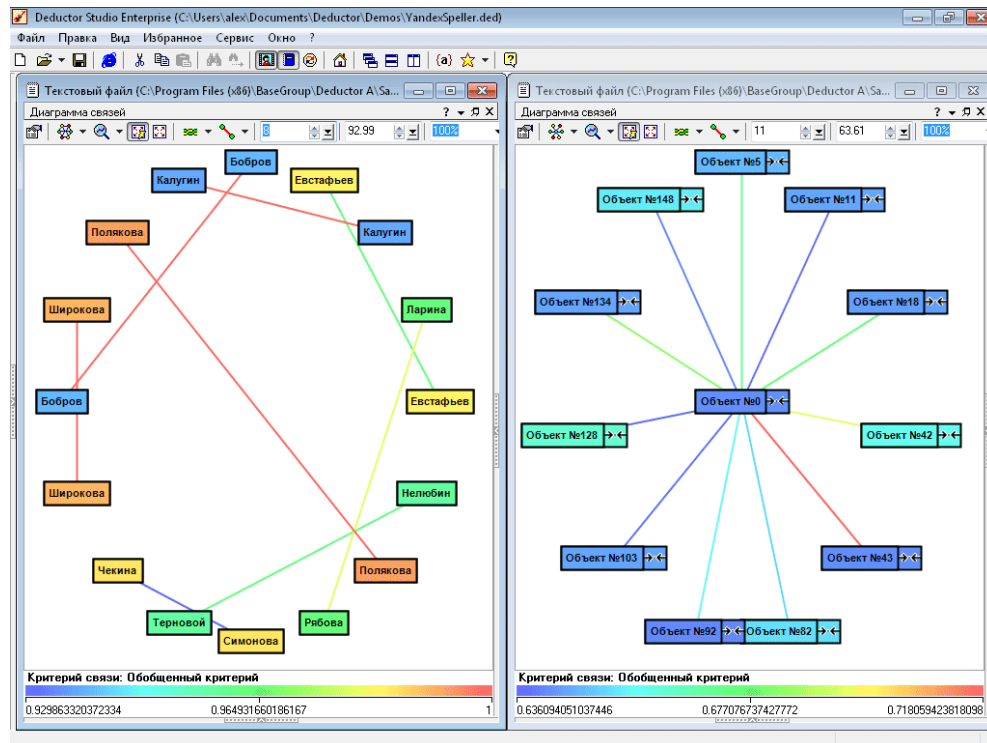
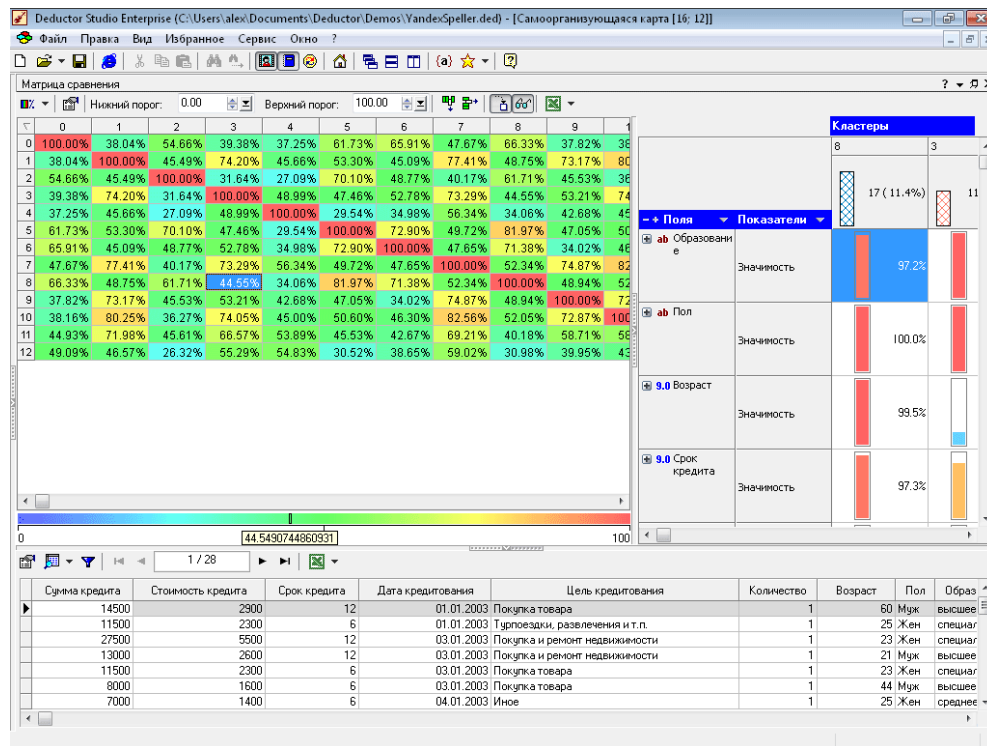


Диаграмма связей



Матрица сравнения





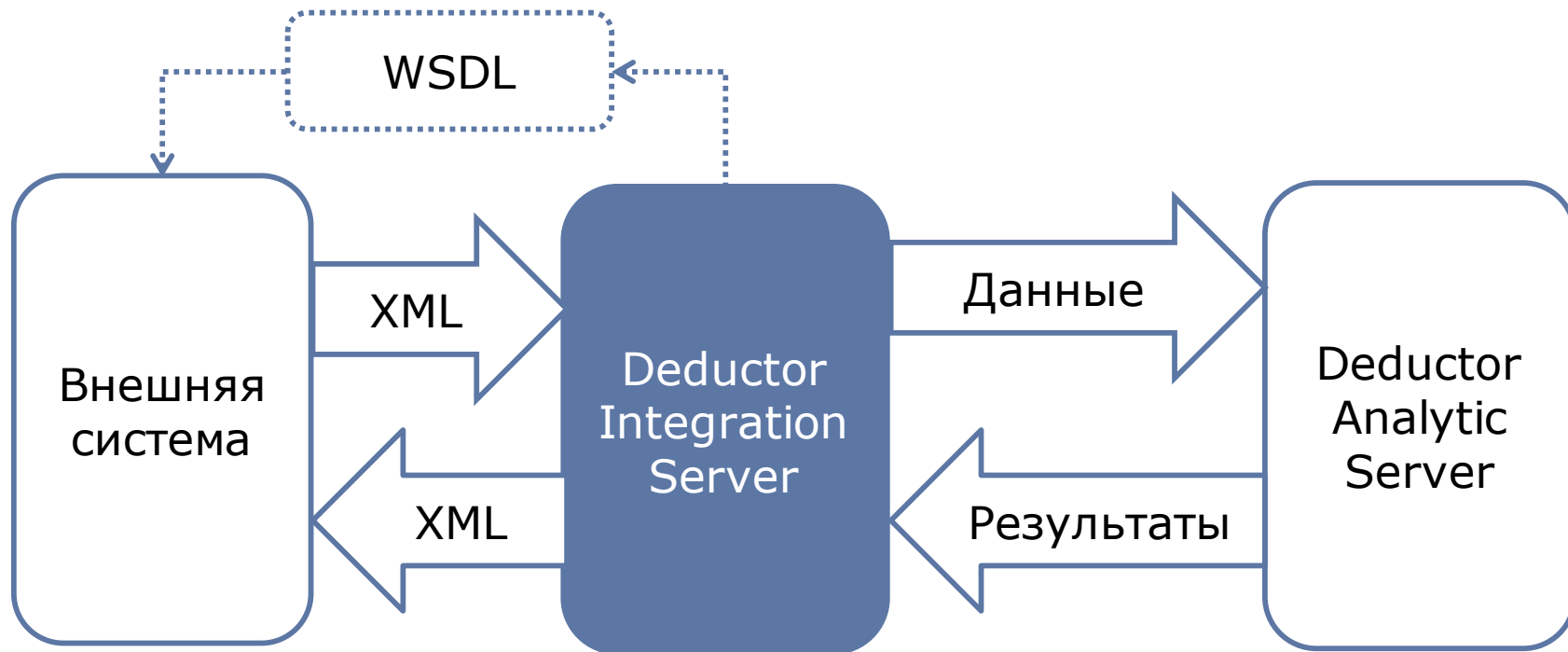
BaseGroup Labs
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

Веб-сервисы

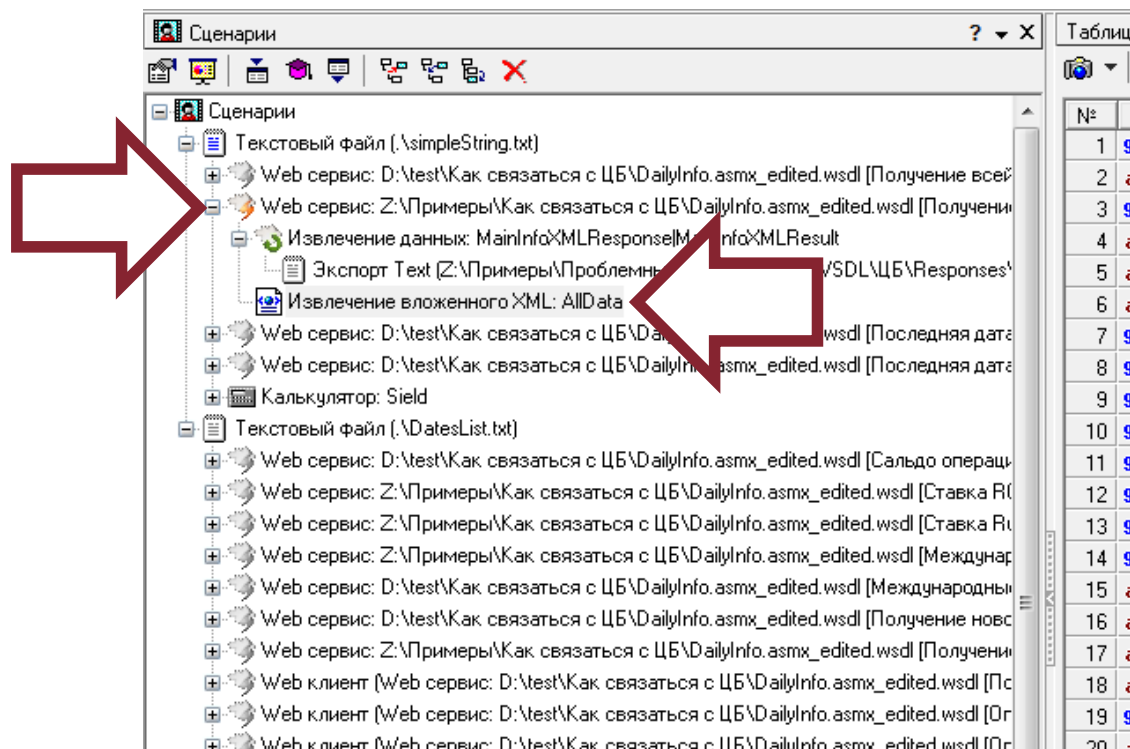
Веб-сервисы в Deductor:

- Транспорт – HTTP / HTTPS
- Формат обмена – XML
- Описание данных – WSDL

SOAP-сервер



SOAP-клиент



Сценарии

- Текстовый файл (.simpleString.txt)
- Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Получение всей информации)
- Web сервис: Z:\Примеры\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Получение информации)
- Извлечение данных: MainInfoXMLResponseMainInfoXMLResult
- Экспорт Text (Z:\Примеры\Проблемные\WSDL\ЦБ\Responses\MainInfoXMLResult.txt)
- Извлечение вложенного XML: AllData
- Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Последняя дата)
- Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Последняя дата)
- Калькулятор: Sield
- Текстовый файл (.DatesList.txt)
- Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Сальдо операции)
- Web сервис: Z:\Примеры\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Ставка Рубля)
- Web сервис: Z:\Примеры\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Ставка Рубля)
- Web сервис: Z:\Примеры\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Международный рубль)
- Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Международный рубль)
- Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Получение новостей)
- Web сервис: Z:\Примеры\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Получение новостей)
- Web клиент (Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Получение новостей))
- Web клиент (Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Получение новостей))
- Web клиент (Web сервис: D:\test\Как связаться с ЦБ\DailyInfo.asmx_edited.wsdl (Получение новостей))

№	Таблиц
1	g
2	€
3	g
4	€
5	€
6	€
7	g
8	g
9	g
10	g
11	g
12	g
13	g
14	g
15	€
16	€
17	€
18	€
19	g
20	€



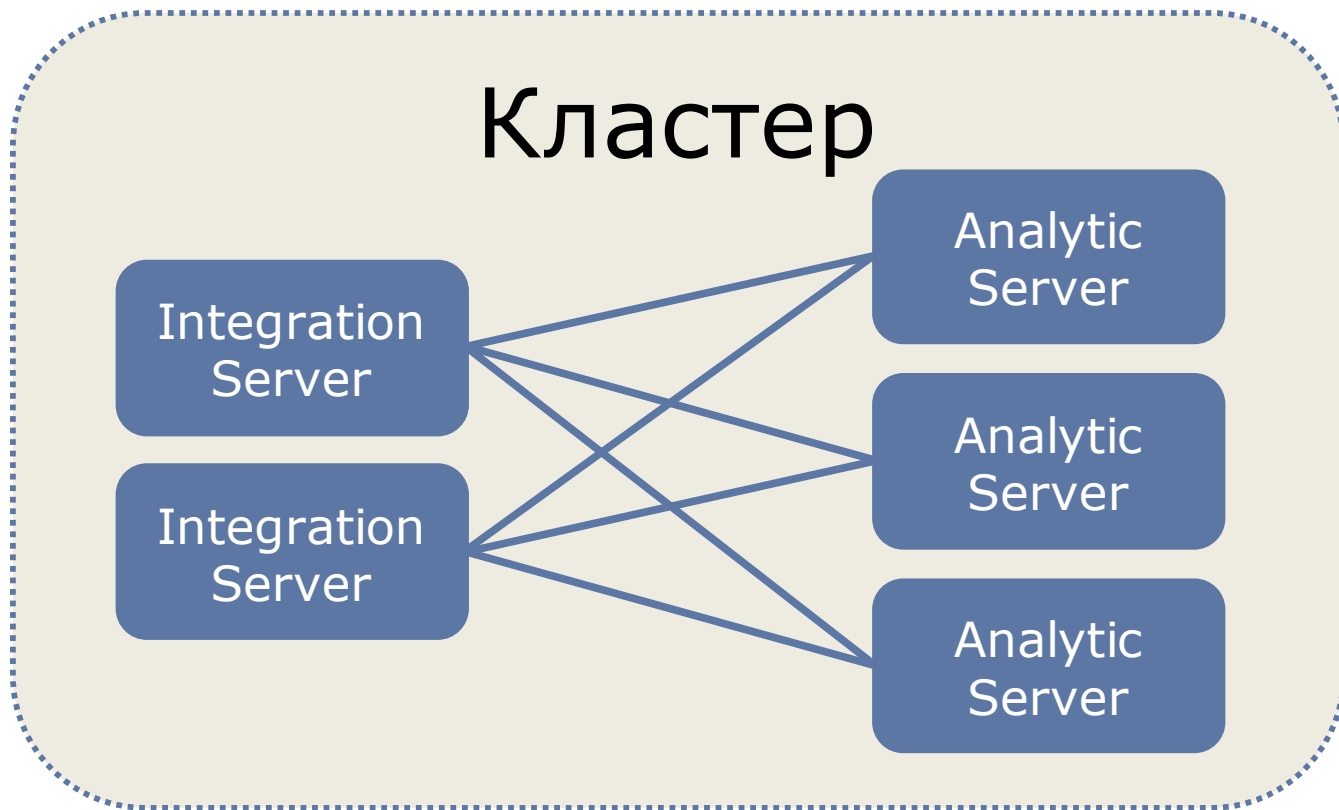
BaseGroup Labs
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

Масштабируемость

Повышение производительности

1. Не считать лишнего
2. Не считать повторно
3. Использовать эвристики
4. Распараллелить расчеты

Организация кластера



Преимущества кластера

1. Горизонтальное масштабирование
2. Балансировка нагрузки
3. Отказоустойчивость
4. Горячая замена компонентов



BaseGroup Labs
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

Защита от копирования

Удобная защита



Компактный
ключ



Ключ на
время



Работа без
драйверов





BaseGroup Labs
ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

basegroup.ru