PMO @ R

MEDIA-TEL by devoteam

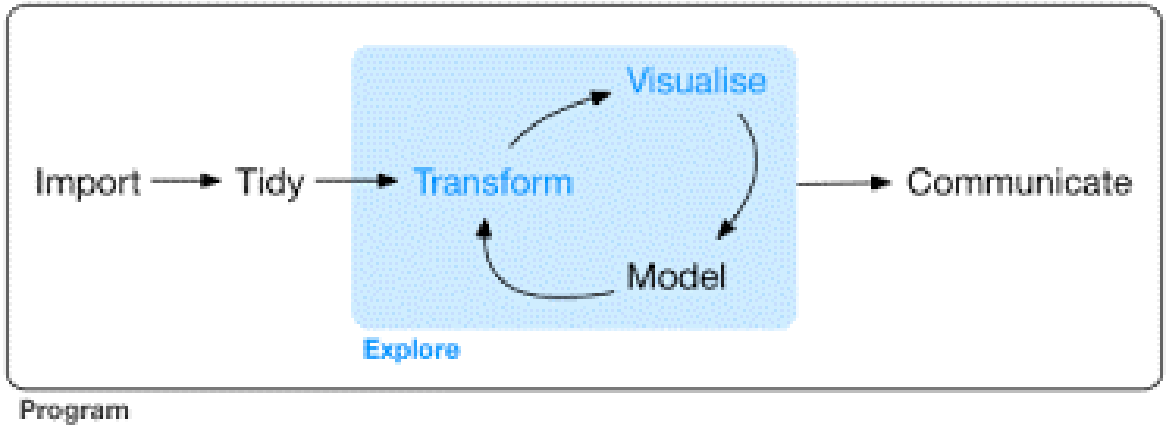
Модули и open-source инструменты
24 March 2021

Специфика аудита бизнес-процессов

- В задачах аудита бизнес-процессов, как правило, требуется выполнение следующих требований:

  - скорость (= деньги) проведения аудита;
  - возможность самостоятельного подключения любых источников данных;
  - возможность самостоятельного проведения любой сложности трансформации данных;
  - возможность быстрого проведения аналитических итераций;
  - возможность повторного проведения аналитики с получением идентичных результатов;
  - представление результатов аудита красивом виде и в различных форматах.

Типовой сценарий в РМО



Последовательность в РМО такая же как во всех активностях по работе с данными

MEDIA-TEL
by devoteam

6 основных «кубиков»

MEDIA-TEL
by devoteam

Кубик #1

Методика формирования исходных данных

MEDIA-TEL
by devoteam

Формирование правильного лога

- Для формирования событийного лога Бизнес система(-ы) должны формировать технический лог, удовлетворяющий определенным требованиям.

- Есть апробированный набор рекомендаций.

Требования к протоколированию бизнес-значимых событий Scan&Go

1. Протоколирование должно вестись в файлы формата Log4j в текстовом формате UTF-8 на жесткий диск back-end сервера Scan&Go
2. Механизм ротации лог файлов должен обеспечивать целостность лога текущего дня в рамках каждого отдельного магазина/сессии. Оптимально 1 день = 1 файл, допустимо сегментирование по кластерам магазинов. Недопустима жесткая ротация по объему, с разрывом посреди стрим/сессий
3. Глубина хранения истории протоколирования 30 дней (допустимо архивирование логов прошлых периодов)
4. Для целей бизнес мониторинга работы Scan&Go должно поддерживаться попадание всех бизнес значимых событий (см. таблицу со списком бизнес значимых событий ниже)
5. Должно обеспечиваться соответствие 1 событие - 1 логическая строка лога
6. Потери информации (полное отсутствие записи о событии, неполная запись о событии) в рамках работы механизма попадания на сервер Scan&Go не должны превышать 0,01% от общего количества событий за период
7. В лог должны записываться как персональные события по клиентской сессии (канал клиентское приложение - сервер приложений), так и все значимые бизнес события, происходящие при взаимодействии сервера приложений с иными внутренними или внешними компонентами и системами (ошибки, фиксированные, авторизация, и т.д.)
8. Любые типы данных, представление которых может или зависит от локали (форматы дат, чисел, валюта и т.д.) должны быть принудительно унифицированы и не зависеть от установки локали сервера/клиента
9. В случае отсутствия явного указания таймзоны, время считается по TZ "Europe/Moscow"
10. Строка в логе должна начинаться с однозначно определяемого идентификатора
11. Для удобства восприятия и ручного анализа лог файла необходимо использовать в качестве разделителя полей одиночный пробел
12. Для всех событий каждая строка лога обязательно должна содержать следующие поля
    a. **Timestamp** записи строки лога с точностью до 1 миллисекунды в формате ISO 8601
    b. Timestamp начала события с точностью до 1 миллисекунды в формате ISO 8601

MEDIA-TEL by devoteam

Кубик #2

Импорт данных

MEDIA-TEL
by devoteam

Импорт данных

Формат	Пакеты
.csv (zip csv)	readr, vroom, data.table
.xlsx	readxl, openxlsx
.json	jsonlite, jqr
.xml	xml2
.docx	docxtractor
REST API	httr
БД (MS SQL, PostgreS, Oracle, MySQL, Access, Redis, Clickhouse, …)	DBI + driver
Google Sheets	googlesheets4
…	…

MEDIA-TEL
by devoteam

Импорт из csv: пример

```r
df <- read_csv("./data/pmo/pmo_sales.csv")

df
```

```
## # A tibble: 12 x 4
##    date       item     unitprice weight
##    <date>     <chr>        <dbl>  <dbl>
##  1 2020-09-01 Арбуз         19.5   17.2
##  2 2020-09-02 Арбуз         19.5    6.8
##  3 2020-09-03 Арбуз         19.5   34.7
##  4 2020-09-05 Арбуз         19.5   12.6
##  5 2020-09-06 Арбуз         19.5   17.2
##  6 2020-09-07 Арбуз         19.5   54.8
##  7 2020-09-01 Помидоры      90      5.7
##  8 2020-09-02 Помидоры      90      2.4
##  9 2020-09-03 Помидоры      90     15.9
## 10 2020-09-04 Помидоры      90     23.6
## 11 2020-09-05 Помидоры      90     24.8
## 12 2020-09-06 Помидоры      90     16.8
```


MEDIA-TEL by devoteam

Импорт из Excel: пример

```
df <- read_excel("./data/pmo/pmo_sales.xlsx", sheet = "Данные здесь")

df
```

```
## # A tibble: 12 x 4
##    date                item     unitprice weight
##    <dttm>              <chr>    <chr>     <chr>
##  1 2020-09-01 00:00:00 Арбуз    19.5      17.2
##  2 2020-09-02 00:00:00 Арбуз    19.5      6.8
##  3 2020-09-03 00:00:00 Арбуз    19.5      34.7
##  4 2020-09-05 00:00:00 Арбуз    19.5      12.6
##  5 2020-09-06 00:00:00 Арбуз    19.5      17.2
##  6 2020-09-07 00:00:00 Арбуз    19.5      54.78
##  7 2020-09-01 00:00:00 Помидоры 90        5.7
##  8 2020-09-02 00:00:00 Помидоры 90        2.4
##  9 2020-09-03 00:00:00 Помидоры 90        15.9
## 10 2020-09-04 00:00:00 Помидоры 90        23.6
## 11 2020-09-05 00:00:00 Помидоры 90        24.8
## 12 2020-09-06 00:00:00 Помидоры 90        16.8
```


MEDIA-TEL by devoteam

Кубик #3

Трансформация данных. «Глаголы tidyverse»

MEDIA-TEL
by devoteam

Глагол mutate – создание колонки

```
df <- read_csv("./data/pmo/pmo_sales.csv") %>%
  # считаем выручку по позициям
  mutate(amount = unitprice * weight)
df
```

```
## # A tibble: 12 x 5
##    date       item     unitprice weight amount
##    <date>     <chr>        <dbl>  <dbl>  <dbl>
##  1 2020-09-01 Арбуз         19.5   17.2   335.
##  2 2020-09-02 Арбуз         19.5    6.8   133.
##  3 2020-09-03 Арбуз         19.5   34.7   677.
##  4 2020-09-05 Арбуз         19.5   12.6   246.
##  5 2020-09-06 Арбуз         19.5   17.2   335.
##  6 2020-09-07 Арбуз         19.5   54.8  1068.
##  7 2020-09-01 Помидоры      90      5.7   513
##  8 2020-09-02 Помидоры      90      2.4   216
##  9 2020-09-03 Помидоры      90     15.9  1431
## 10 2020-09-04 Помидоры      90     23.6  2124
## 11 2020-09-05 Помидоры      90     24.8  2232
## 12 2020-09-06 Помидоры      90     16.8  1512
```


MEDIA-TEL by devoteam

Глагол group_by – группировка по колонкам, глагол summarise – расчет подытога

```
# считаем выручку по товарам
df %>%
  group_by(item) %>%
  summarise(sum(weight), sum(amount))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 3
##   item     `sum(weight)` `sum(amount)`
##   <chr>            <dbl>         <dbl>
## 1 Арбуз            143.          2794.
## 2 Помидоры          89.2         8028
```


MEDIA-TEL
by devoteam

Глагол select – выбор и переименование колонок

```
df %>%
  select("Дата" = date, "Выручка, руб" = amount, item)
```

```
## # A tibble: 12 x 3
##    Дата       `Выручка, руб` item
##    <date>              <dbl> <chr>
##  1 2020-09-01           335. Арбуз
##  2 2020-09-02           133. Арбуз
##  3 2020-09-03           677. Арбуз
##  4 2020-09-05           246. Арбуз
##  5 2020-09-06           335. Арбуз
##  6 2020-09-07          1068. Арбуз
##  7 2020-09-01           513  Помидоры
##  8 2020-09-02           216  Помидоры
##  9 2020-09-03          1431  Помидоры
## 10 2020-09-04          2124  Помидоры
## 11 2020-09-05          2232  Помидоры
## 12 2020-09-06          1512  Помидоры
```


MEDIA-TEL by devoteam

Глагол filter – выбор строк по условию

```
df %>%
  filter(amount > 1000, item == "Арбуз")
```

```
## # A tibble: 1 x 5
##   date       item  unitprice weight amount
##   <date>     <chr>     <dbl>  <dbl>  <dbl>
## 1 2020-09-07 Арбуз      19.5   54.8  1068.
```


MEDIA-TEL
by devoteam

Глагол arrange – сортировка строк по колонкам

```
df %>%
  arrange(date, desc(amount))
```

```
## # A tibble: 12 x 5
##    date       item     unitprice weight amount
##    <date>     <chr>        <dbl>  <dbl>  <dbl>
##  1 2020-09-01 Помидоры      90      5.7   513
##  2 2020-09-01 Арбуз         19.5   17.2   335.
##  3 2020-09-02 Помидоры      90      2.4   216
##  4 2020-09-02 Арбуз         19.5    6.8   133.
##  5 2020-09-03 Помидоры      90     15.9  1431
##  6 2020-09-03 Арбуз         19.5   34.7   677.
##  7 2020-09-04 Помидоры      90     23.6  2124
##  8 2020-09-05 Помидоры      90     24.8  2232
##  9 2020-09-05 Арбуз         19.5   12.6   246.
## 10 2020-09-06 Помидоры      90     16.8  1512
## 11 2020-09-06 Арбуз         19.5   17.2   335.
## 12 2020-09-07 Арбуз         19.5   54.8  1068.
```


MEDIA-TEL by devoteam

Форматный вывод

```
df %>%
  group_by(item) %>%
  gt(rowname_col = "date")
```

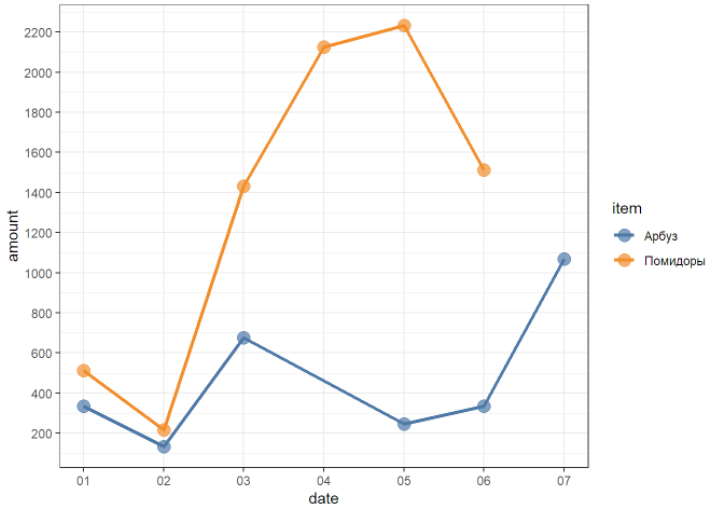
	unitprice	weight	amount
Арбуз			
2020-09-01	19.5	17.20	335.40
2020-09-02	19.5	6.80	132.60
2020-09-03	19.5	34.70	676.65
2020-09-05	19.5	12.60	245.70
2020-09-06	19.5	17.20	335.40
2020-09-07	19.5	54.78	1068.21
Помидоры			
2020-09-01	90.0	5.70	513.00
2020-09-02	90.0	2.40	216.00
2020-09-03	90.0	15.90	1431.00
2020-09-04	90.0	23.60	2124.00
2020-09-05	90.0	24.80	2232.00
2020-09-06	90.0	16.80	1512.00

MEDIA-TEL
by devoteam

Графический вывод

```
gp <- ggplot(df, aes(date, amount, color = item, fill = item)) +
  geom_point(size = 4, shape = 19, alpha = 0.7) +
  geom_line(lwd = 1.1) +
  scale_x_date(date_breaks = "1 day", date_minor_breaks = "1 day", date_labels = "%d") +
  scale_y_continuous(breaks = scales::pretty_breaks(10)) +
  ggthemes::scale_color_tableau() +
  ggthemes::scale_fill_tableau() +
  theme_bw()

gp
```



MEDIA-TEL
by devoteam

Разговорная 5-ти минутка

Применяем «глаголы» трансформации к тематике process mining

Импорт лога событий

```r
df <- read_csv("./data/pmo/pmo_school.csv")

df
```

```
## # A tibble: 14 x 5
##    timestamp           activity resource part       user
##    <dttm>              <chr>    <chr>    <chr>      <chr>
##  1 2020-09-01 07:57:01 Открыл   Дверь    Входную    Коля
##  2 2020-09-01 08:12:24 Открыл   Дверь    Входную    Маша
##  3 2020-09-01 08:14:47 Открыл   Окно     Фрамугу    Маша
##  4 2020-09-01 08:27:29 Вытер    Доску    <NA>       Митя
##  5 2020-09-01 08:18:22 Закрыл   Окно     Фрамугу    Митя
##  6 2020-09-01 09:15:35 Взял     Учебник  География  Коля
##  7 2020-09-01 10:05:13 Положил  Учебник  География  Коля
##  8 2020-09-01 15:12:24 Закрыл   Дверь    Входную    Маша
##  9 2020-09-02 07:43:15 Открыл   Дверь    Входную    Коля
## 10 2020-09-02 08:03:14 Открыл   Дверь    Входную    Митя
## 11 2020-09-02 08:05:25 Открыл   Окно     Форточку   Маша
## 12 2020-09-02 10:25:41 Взял     Учебник  Математика Коля
## 13 2020-09-02 15:19:36 Закрыл   Окно     Форточку   Митя
## 14 2020-09-02 17:12:24 Закрыл   Дверь    Входную    Коля
```


MEDIA-TEL by devoteam

Прямые расчеты по логу

- Подсчет числа вхождений композитной активности на основе activity и resourse

```
df %>%
  mutate(new_activity = glue("{activity} - {resource}")) %>%
  count(new_activity, sort = TRUE)
```

```
## # A tibble: 7 x 2
##   new_activity          n
##   <glue>            <int>
## 1 Открыл - Дверь        4
## 2 Взял - Учебник        2
## 3 Закрыл - Дверь        2
## 4 Закрыл - Окно         2
## 5 Открыл - Окно         2
## 6 Вытер - Доску         1
## 7 Положил - Учебник     1
```

- Какая активность была последней и в какой час она происходила?

```
df %>%
  mutate(hr = hour(timestamp), date = as_date(timestamp)) %>%
  group_by(date) %>%
  # оставляем самое последнее действие
  filter(timestamp == max(timestamp)) %>%
  ungroup() %>%
  select(date, hr, everything(), -timestamp, -part)
```

```
## # A tibble: 2 x 5
##   date          hr activity resource user
##   <date>     <int> <chr>    <chr>    <chr>
## 1 2020-09-01    15 Закрыл   Дверь    Маша
## 2 2020-09-02    17 Закрыл   Дверь    Коля
```


MEDIA-TEL by devoteam

Кубик #4

Набор инструментов для проведения process mining исследований

MEDIA-TEL
by devoteam

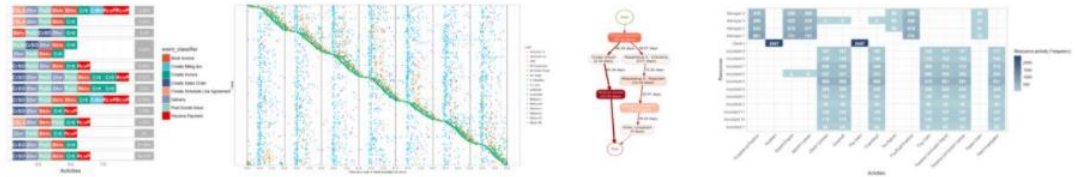
Пакет bupaR: набор инструментов для PM

Home | About | Get started | Import | Preprocessing | EDA | Process Discovery | Conformance Checking | Packages

Business Process Analysis

bupaR

bupaR is an open-source, integrated suite of R-packages for the handling and analysis of business process data. It currently consists of 8 packages, including the central package, supporting different stages of a process mining workflow.



bupaR provides support for different stages in process analysis, such as importing event data, calculating descriptives, process monitoring and process visualization. The central package, bupaR includes basic functionality for creating event log objects in R. It contains several functions to get information about an event log and also provides specific event log versions of generic R functions.

https://www.bupar.net/index.html

MEDIA-TEL by devoteam

Пример - событийный лог взаимодействия с пациентами

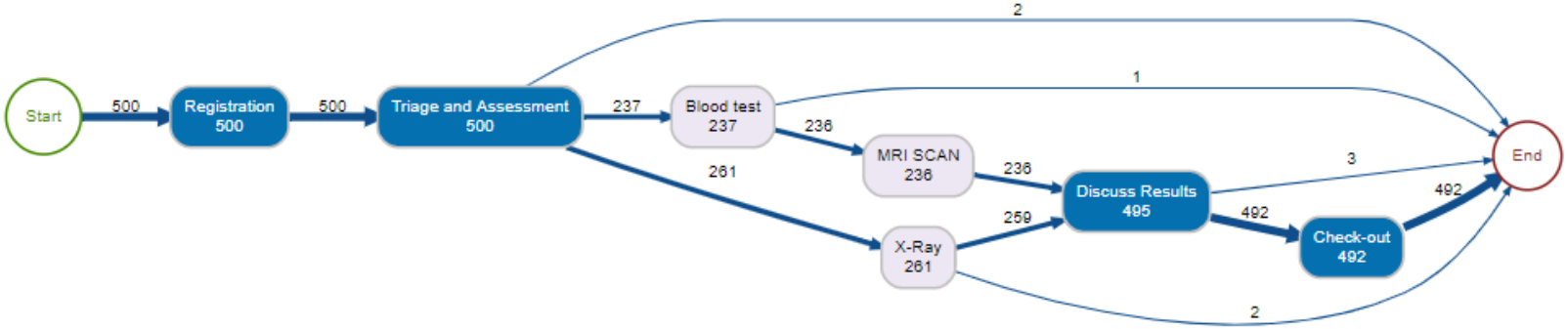
```
patients
```

```
## Log of 5442 events consisting of:
## 7 traces
## 500 cases
## 2721 instances of 7 activities
## 7 resources
## Events occurred from 2017-01-02 11:41:53 until 2018-05-05 07:16:02
##
## Variables were mapped as follows:
## Case identifier:     patient
## Activity identifier:     handling
## Resource identifier:     employee
## Activity instance identifier:    handling_id
## Timestamp:           time
## Lifecycle transition:        registration_type
##
## # A tibble: 5,442 x 7
##    handling patient employee handling_id registration_ty~ time
##    <fct>    <chr>   <fct>    <chr>       <fct>            <dttm>
##  1 Registr~ 1       r1       1           start            2017-01-02 11:41:53
##  2 Registr~ 2       r1       2           start            2017-01-02 11:41:53
##  3 Registr~ 3       r1       3           start            2017-01-04 01:34:05
##  4 Registr~ 4       r1       4           start            2017-01-04 01:34:04
##  5 Registr~ 5       r1       5           start            2017-01-04 16:07:47
##  6 Registr~ 6       r1       6           start            2017-01-04 16:07:47
##  7 Registr~ 7       r1       7           start            2017-01-05 04:56:11
##  8 Registr~ 8       r1       8           start            2017-01-05 04:56:11
##  9 Registr~ 9       r1       9           start            2017-01-06 05:58:54
## 10 Registr~ 10      r1       10          start            2017-01-06 05:58:54
## # ... with 5,432 more rows, and 1 more variable: .order <int>
```


MEDIA-TEL by devoteam

Реконструируем карту процесса (DFG)

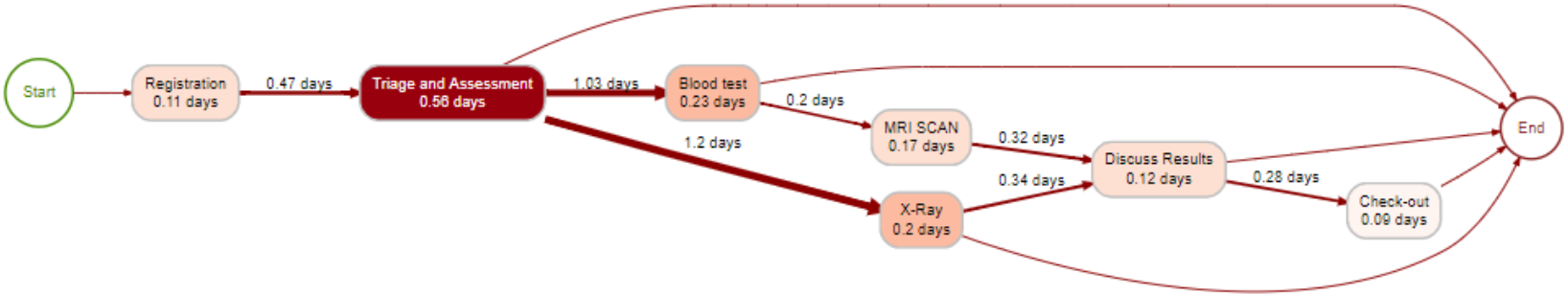
```
patients %>%
    process_map()
```



MEDIA-TEL by devoteam

«Достаем» метрики производительности процесса

```
patients %>%
    process_map(performance(median, "days"))
```

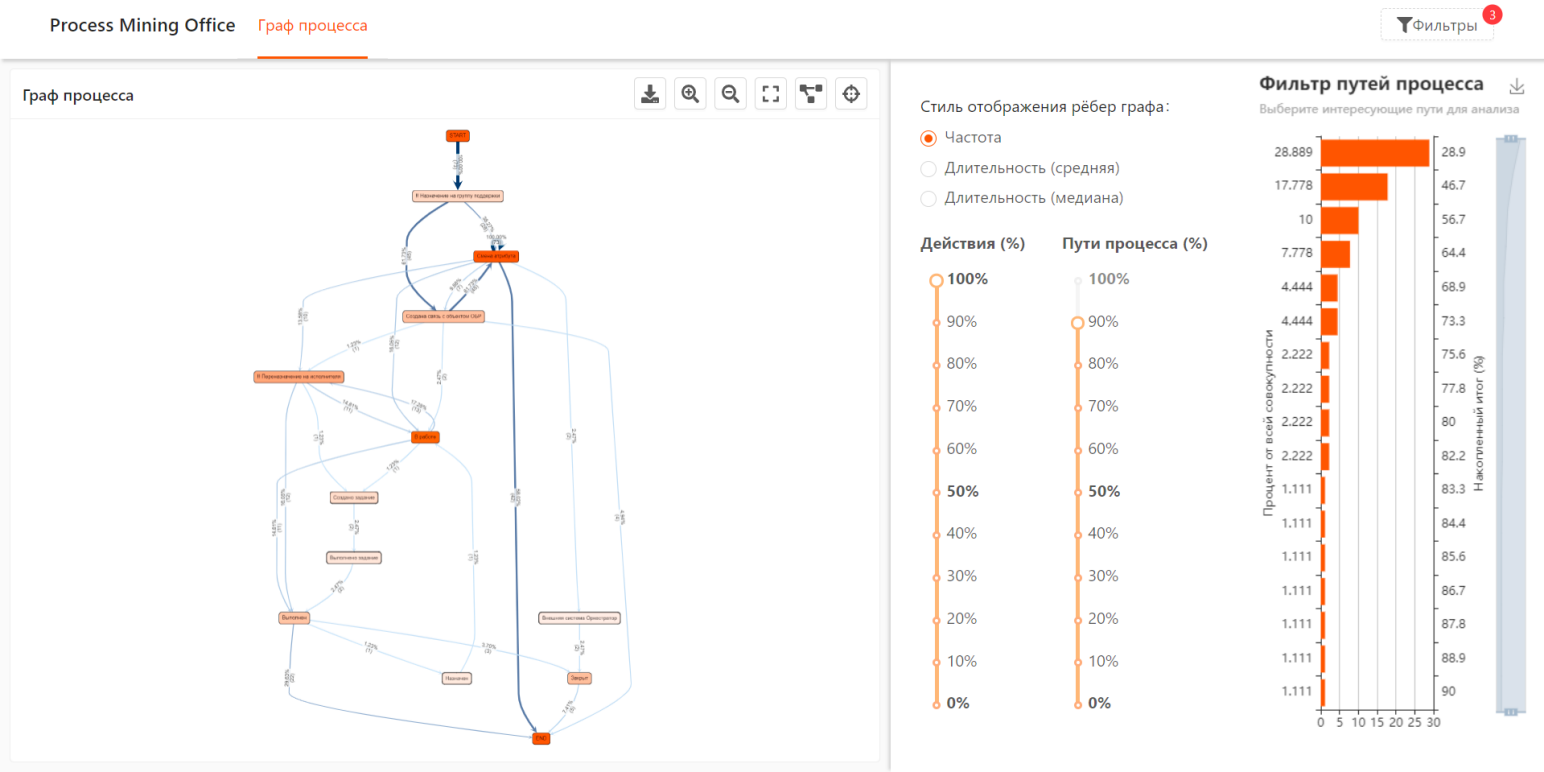


MEDIA-TEL by devoteam

«Traces» или «Customer Journey»

MEDIA-TEL by devoteam

Облегченный АРМ для анализа DFG



MEDIA-TEL by devoteam

Кубик #5

Представление для пользователей

MEDIA-TEL
by devoteam

Популярные средства коммуникации

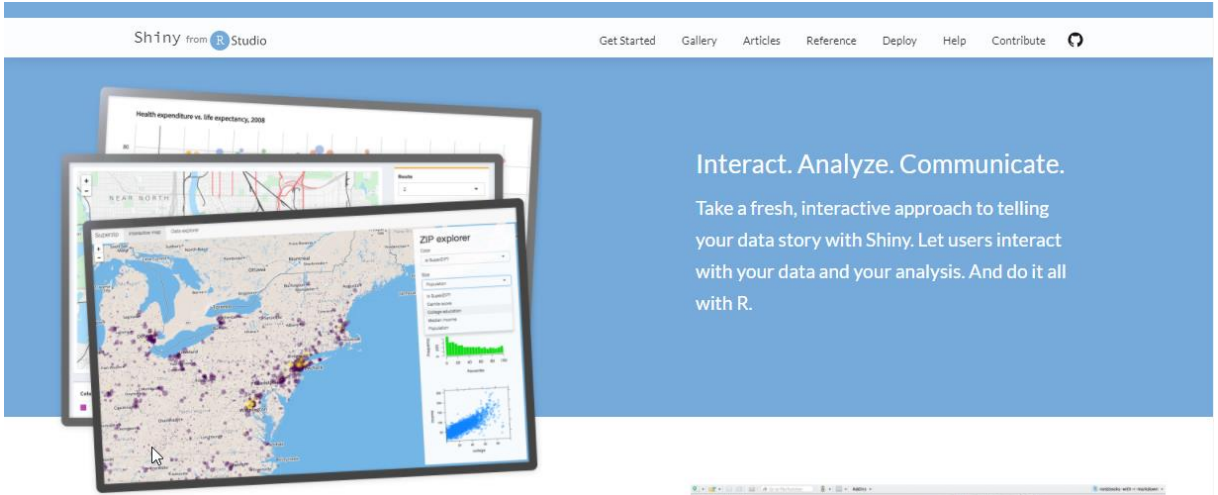
- Story-telling отчеты. Генерация по требованию или по расписанию текстово-графического html документа, последовательно освещающего все важные аспекты для пользователя. Возможно включение элементов интерактива (js виджеты).

- Дашборд на «телевизор в холле». Периодически обновляемый набор метрик (таблично-графический)

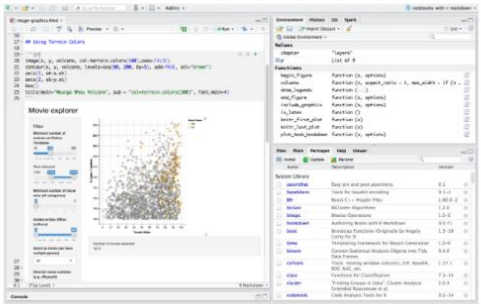
- Специализированный АРМ продвинутого пользователя (аналитика) для проведения анализа данных.

- Выгрузки во внешний мир.

Open-source Shiny from RStudio

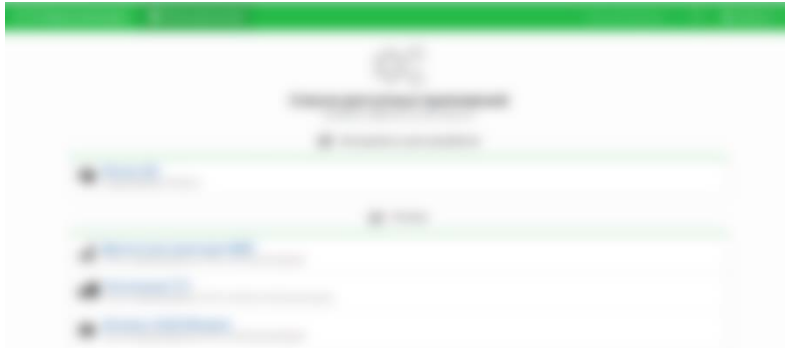


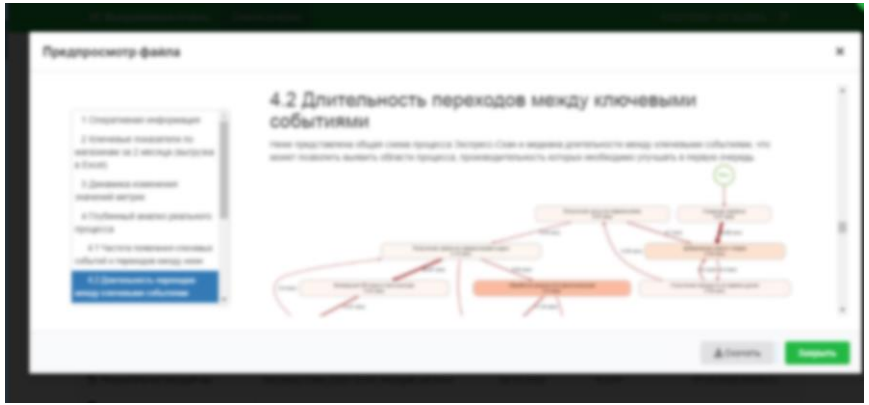
Shiny is an R package that makes it easy to build interactive web apps straight from R. You can host standalone apps on a webpage or embed them in R Markdown documents or build dashboards. You can also extend your Shiny apps with CSS themes, htmlwidgets, and JavaScript actions.



MEDIA-TEL
by devoteam

Примеры story-telling отчетов





MEDIA-TEL by devoteam

Примеры дашбордов

MEDIA-TEL
by devoteam

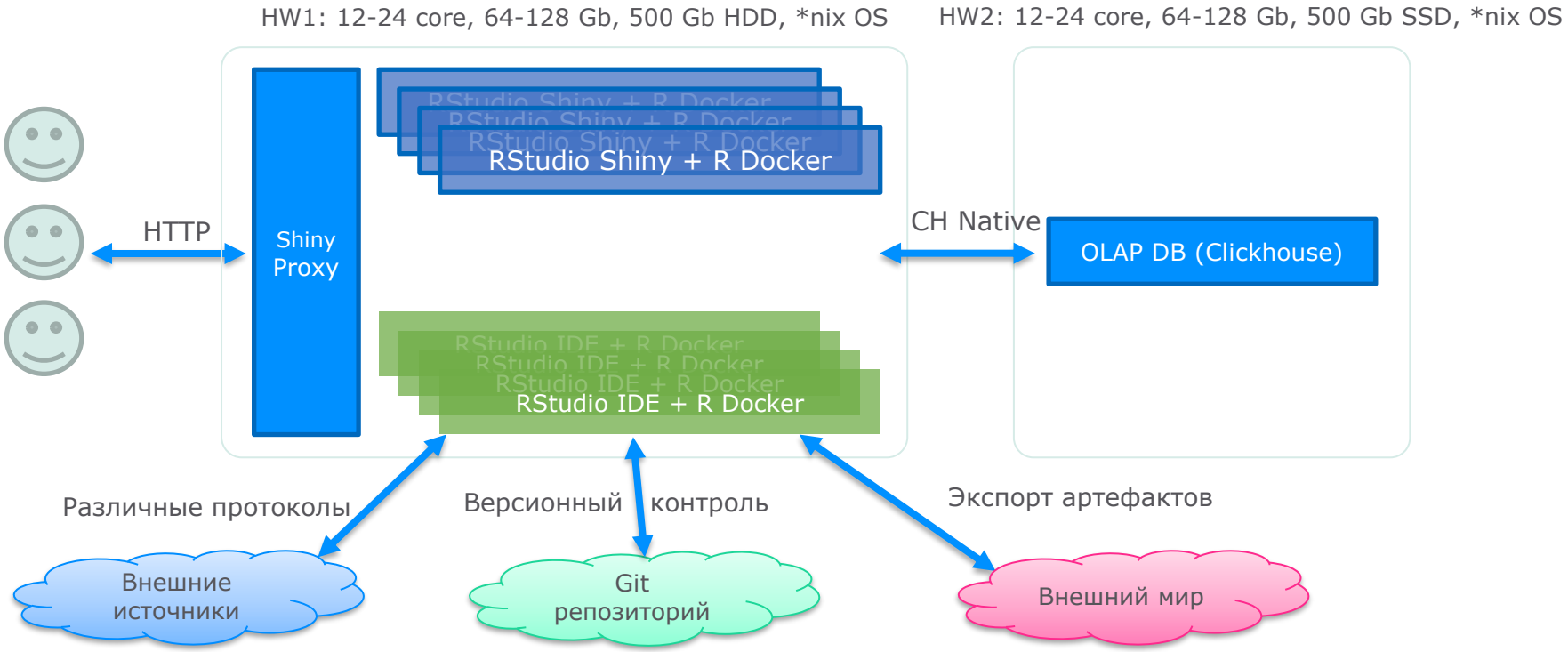
Примеры различных АРМ

MEDIA-TEL
by devoteam

Кубик #6

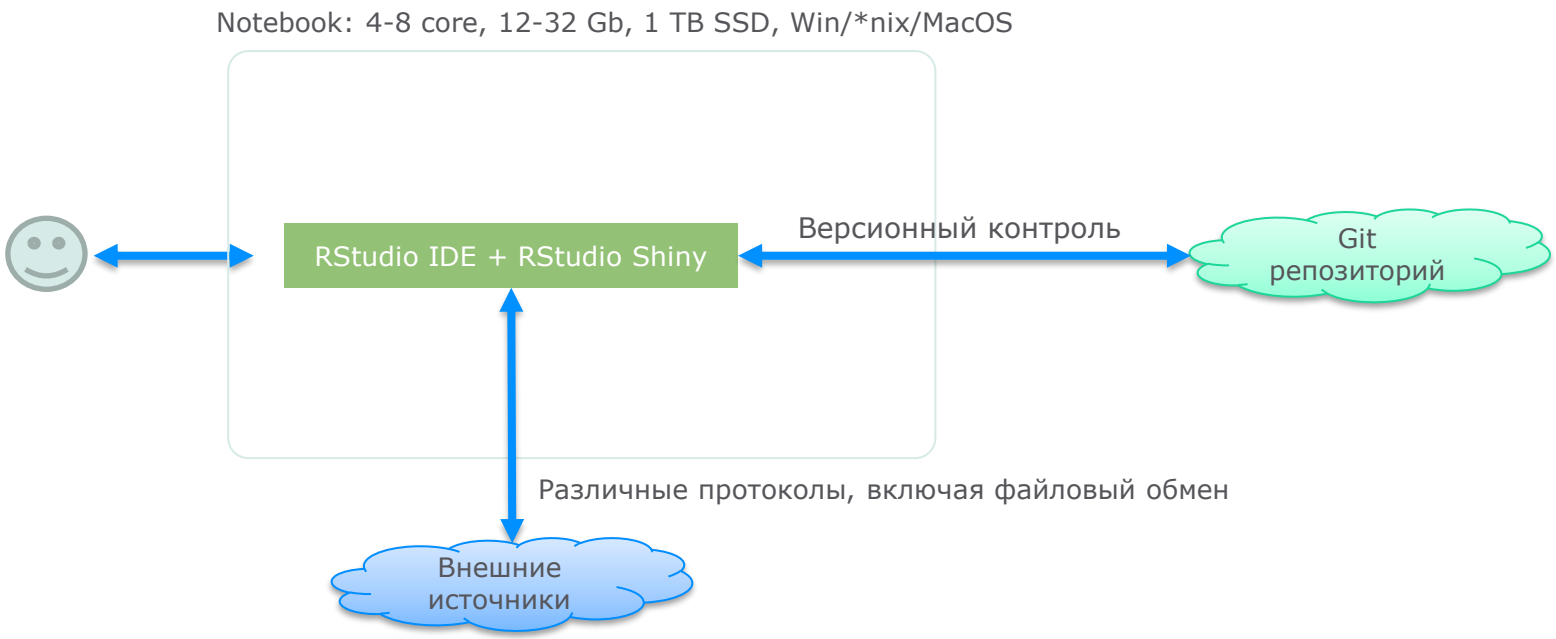
Программно-аппаратный комплекс

Типовая минималистичная архитектура



MEDIA-TEL by devoteam

ИЛИ даже так



MEDIA-TEL by devoteam

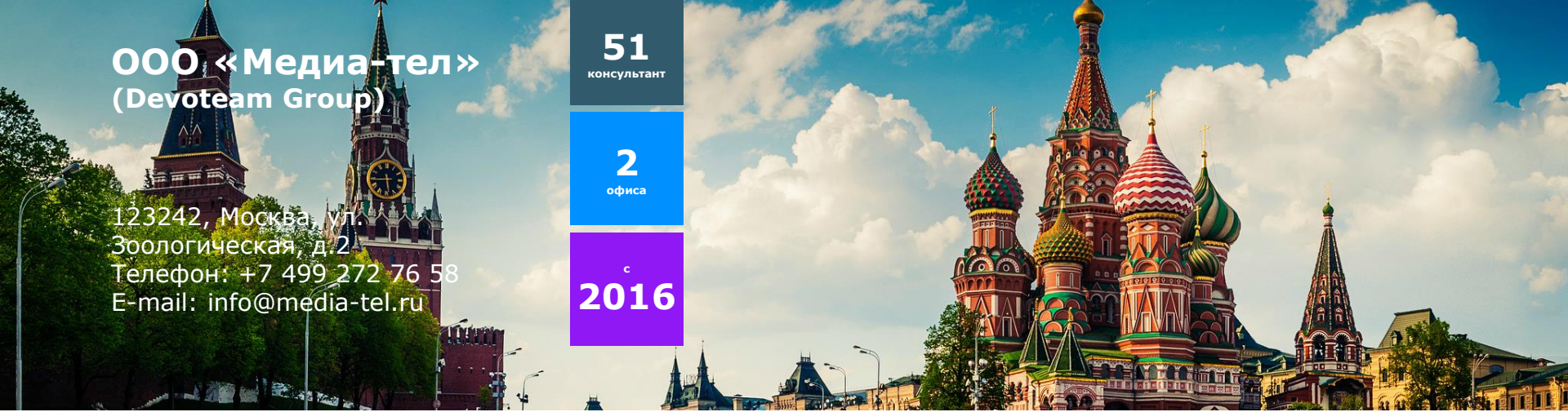
Некоторые материалы для расширенного изучения

- RStudio IDE (https://rstudio.com/products/rstudio/) – среда разработки
- RStudio Shiny (https://shiny.rstudio.com/) – создание web APM
- RMarkdown (https://rmarkdown.rstudio.com/) – компоненты для создания html отчетов
- Yandex Clickhouse (https://clickhouse.tech/) – колоночная БД
- CRAN (https://cran.r-project.org/) – язык R и репозиторий пакетов
- Tidyverse (https://www.tidyverse.org/) – фреймворк для эффективной обработки и визуализации данных в табличном представлении
- Tidymodels (https://www.tidymodels.org/) – фреймворк для эффективного использования алгоритмов машинного обучение в tidyverse стиле
- Htmlwidgets (http://gallery.htmlwidgets.org/) – репозиторий визуальных расширений, включая интерактивные инструменты

MEDIA-TEL by devoteam

Публикации

- Как в enterprise приручить при помощи R технологии process mining?
  - https://habr.com/ru/post/526194/
- Бизнес-процессы в enterprise компаниях: домыслы и реальность. Проливаем свет с помощью R
  - https://habr.com/ru/post/461463/



ООО «Медиа-тел»
(Devoteam Group)

123242, Москва, ул. Зоологическая, д.2
Телефон: +7 499 272 76 58
E-mail: info@media-tel.ru

51 консультант

2 офиса

с **2016**

Константин Левчин
konstantin.levchin@devoteam.com
Тел.: +7 985 767 79 42

Кирилл Кушнарев
kirill.kushnarev@devoteam.com
Тел.: +7 916 160 12 16

Татьяна Орлова
tatiana.orlova@devoteam.com
Тел.: +7 919 106 52 15

Innovate Technology Consulting for business