ASTRAVERGE / PGDEVELOPER

# Аналитика R: работа, консоль и модели

## AstraVerge PGDeveloper 0.7.0

Практическое руководство · 11 ноября 2026

Рабочая среда для PostgreSQL: от структуры базы до кода приложения.

Команды и имена файлов сохраняют техническое имя av-pgdeveloper. Параметры и границы применимости относятся к версии 0.7.0.

# Содержание

R Analytics через AV-Agent 3

Покупатель: активация и первый запуск . . . 3

Состав поставки и администрирование . . . 3

Безопасность SQL и точность . . . 4

Коммерческая поставка . . . 4

Проверки . . . 5

Консоль R, модели и фоновые отчёты 5

Выбор пакетов . . . 5

Сохранение и расписание . . . 6

Версия 0.7.0. Единое руководство; содержание исходных разделов сохранено.

# R Analytics через AV-Agent

В контекстном меню Result Grid доступно Analyze: Summary, Group By, Frequency, Distribution, Correlation, Data Quality, AI Data, Time Series, Streaming summary, Clustering, Decision tree, R console и Analyze on PostgreSQL. Результаты открываются в отдельном Analysis Window (Summary/Table/Chart/Messages/R console), с выбором колонок, настройками, Progress/Cancel и экспортом CSV/JSON/PNG. Повторная команда из неизменённого грида использует прежнее окно/dataset. После правки грида следующий Analyze снимает новые данные. Аналитические результаты не редактируют SQL Result.

## Покупатель: активация и первый запуск

Сертификаты, TLS, адреса и пароли подключения вручную настраивать не нужно.

1. Установить сборку PGDeveloper с модулем R Analytics и активировать приобретённый модуль обычным ZIP-файлом AV-Guard из покупки. Установщик поставки может заранее поместить его в `<профиль>/licenses/av-pgdeveloper-r.zip`; иначе PGDeveloper предложит выбрать файл активации. Дополнительная покупка AV-Agent не требуется. 2. Выполнить запрос и выбрать Analyze → Summary. Программа сама установит локальные компоненты, создаст защищённое подключение и запустит существующий AV-Agent. При следующем входе в Windows он запускается автоматически в текущем профиле. 3. Если R отсутствует, подтвердить установку бесплатной среды R либо выбрать уже установленный R. После установки исходный Analyze продолжится без перезапуска IDE.

Настройки находятся в отдельном меню Settings справа от Tools: AV-Agent - состояние, «Настроить автоматически», параметры подключения и диагностика; R Runtime - проверка/выбор среды R и официальный установщик; Licensing - подписанный ZIP активации R Analytics и проверка доступа модуля. Технические поля доступны через AV-Agent → «Настройки подключения…» для внешнего AV-Agent. Автоматический режим всегда локальный HTTPS/mTLS; никакого отключения проверки сертификатов или перехода на открытый HTTP нет. Клиентский ключ автоматически защищается DPAPI текущего пользователя Windows.

R устанавливается отдельно, системно в `C:/Programs/R/R-4.6.2` со штатным UAC. Официальный HTTPS CRAN и закреплённый SHA-256 проверяются автоматически. R не входит в дистрибутив PGDeveloper; лицензия GPL-2 | GPL-3. Settings → R Runtime → Install analysis packages устанавливает выбранный набор CRAN в отдельную библиотеку модуля.

## Состав поставки и администрирование

`prepareAnalyticsBundle` помещает в сборку только av-agent.exe, av-pgdeveloper-r.exe, av-pgdeveloper-r-setup.exe и официальный plugin.avdoc. В поставке нет лицензии разработчика, сертификатов, закрытых ключей или R Runtime. Без исходников, Go/Python/PowerShell- скриптов разработки у покупателя устанавливается одноразовым av-pgdeveloper-r-setup.exe. Он использует обычный AV-Agent `-check` для проверки купленной лицензии Guard и подписи AV-R; недействительный комплект не становится активной установкой. Новый сервер/транспорт/supervisor не создаются: ими остаются штатные компоненты Agent.

Локальный профиль модуля - `%LOCALAPPDATA%/AstraVerge/av-pgdeveloper-r`, с закрытой ACL, динамически выбранным loopback-портом и `connection.json`. Метаданные ссылаются только на фиксированные локальные файлы; попытка перенаправить ключ на удалённый адрес или произвольный путь отклоняется. Подключение для профиля PGDeveloper не требует analyticsAgentUrl/CA/PKCS-12/token. Старые явно заданные подключения сохраняются для совместимости; после успешной проверки локального модуля «Среда R» включает автоматическое подключение.

В Settings → R Runtime всегда доступна единая «Среда R». Без лицензии аналитики показана только активация пакетом покупки. После подтверждения доступа показаны настройки среды R и пакетов; активация скрыта. Адреса, сертификаты и параметры транспорта настраиваются автоматически. Отдельных пунктов AV-Agent и Licensing нет. Правила регистрации, scopes и лицензирования: README AV-R. Автоматический установщик первой версии предназначен для Windows; альтернативные ОС используют существующую операторскую настройку Agent.

## Безопасность SQL и точность

Loaded Rows: только имеющиеся строки, включая локальные правки, без помеченных удалёнными; частичная загрузка явно показана в заголовке/Messages. Нет повторного SELECT, Commit/Rollback, JDBC-отмены, закрытия reader или переключения транзакций. Full Result после отдельного подтверждения повторяет SELECT в независимой read-only транзакции с forward-only JDBC reader и пишет все строки в дисковый снимок. Грид, его курсор и транзакция остаются прежними; незакоммиченные изменения и локальные правки в новую выборку не включаются. Ограничения max_rows/max_bytes провайдера сохраняются.

Типизированный снимок сохраняет порядок колонок, NULL, пустые строки, decimal/int8, даты и offsets. До 256 KiB/10 тыс. строк используется существующий JSON, далее - локальный файл с явной схемой, hash/owner/root validation. Это выбор по памяти, не привязка к временному ограничению размера запроса Agent. Удалённые большие наборы пока отклоняются. numeric не округляется при передаче; арифметика R требует явного checkbox согласия, если double не представляет значения точно. Границы Data Quality и Frequency не требуют округления. Бинарные колонки надо исключить SQL-проекцией; сложные типы анализируются как текст. NaN/Infinity и даты BC/infinity пока не поддержаны.

Даты сравниваются в UTC, timestamp без зоны интерпретируется как UTC только для анализа. SD/variance выборочные, quantile type 7; IQR 1.5, Z-score >3. Skewness и excess kurtosis - моменты относительно sample SD без поправки на смещение. Корреляции не определяют автоматически пригодность признаков и не доказывают причинность. AI Data - обычная статистика, без LLM/обучения ML. Ollama не используется.

Один dataset привязан к principal/product/session и boot плагина. После аварии сообщается dataset_lost; Reload snapshot явно загружает тот же снимок. Новые строки анализировать повторным Analyze из грида. Операции неизвестного исхода не повторяются. При закрытии окна сессия освобождается; есть TTL и очистка abandoned plugin work. Результат <=100 тыс. строк/16 MiB; исходник по умолчанию <=256 MiB/2 млн строк; R читает данные в память, реальный RSS выше размера файла. CSV - первая таблица, NULL `\N` (для lossless/all tables выбрать JSON), PNG - последний график.

## Коммерческая поставка

Personal 999 ₽ без коммерческого использования; Professional 5 000 ₽ с коммерческим; Enterprise от 7 000 ₽/место, минимум 10 мест. AV-R дополнительно 999 ₽ за каждое лицензированное место любой редакции. Один бинарник для всех редакций.

Право AV-R реально проверяется существующим AV-Agent/AV-Guard: подписанный av_agent с plugins[id=av-pgdeveloper-r] и подписанный release executable. Оно включается в bundle покупки PGDeveloper+модуль; отдельная покупка Agent не вводится. Guard core не изменён. Регистрацию товаров/связанный выпуск bundle и коммерческую подпись надо выполнить штатной системой перед поставкой; тестовые лицензии для этого непригодны. Общего verifier базовых редакций в Java ранее не было: его подключение через официальный AV-Guard SDK остаётся отдельным этапом, локальный флаг не заменяет лицензию.

### Проверки

`gradlew test` - точность snapshot, NULL/empty, partial/full, локальные правки, отмена и защита транспортных настроек. `../AV-Agent-Providers/av-pgdeveloper-r/tools/build.ps1` - Go/R наборы. `../AV-Agent-Providers/av-pgdeveloper-r/tools/integration.ps1` - настоящий подписанный Agent → AV-R → R, API, перезапуск/устаревшие IDs и Swing Result Grid на 30 тыс. строк с heartbeat EDT, Summary → Correlation → Group By без повторной загрузки. БД не используются. Протокол: AV-R CONTRACT.

Результаты проверки и изменённые компоненты: R-ANALYTICS-VALIDATION.md.

Автоматическая установка проверяется `tools/analytics-first-use-smoke.ps1 -ActivationPackage <ZIP>`: новый профиль, только купленный файл активации, затем настоящий Swing Analyze на 30 тыс. строк без ввода технических параметров. Тест не меняет автозапуск рабочего профиля.

## Консоль R, модели и фоновые отчёты

Result Grid → Analyze → R console открывает ручной редактор над отдельным снимком. Execute R selection / script исполняет выделение, иначе весь скрипт. `resultset` - data.frame; `resultset_raw` сохраняет исходные строки и NULL, `resultset_schema` - имена и типы. Decimal и большие int8 остаются текстом до явного преобразования. Переменные сохраняются между командами одного dataset; Reset R workspace удаляет их. Ошибка скрипта возвращается как ошибка операции; предыдущий workspace сохраняется.

```
head(resultset)
mean(as.numeric(resultset$amount), na.rm = TRUE)
plot(as.numeric(resultset$amount))
```

Консоль работает с обычными правами учётной записи AV-Agent: это явно разрешённый пользователем R-код, а не sandbox. Для внешнего Agent требуется отдельное разрешение `permissions[].console=true`; новая локальная установка выдаёт его своему desktop principal. Неизвестный исход никогда не вызывает автоматический повтор команды. Workspace ограничен 64 MiB, отображаемый вывод - 2000 строк/65536 символов.

Use chunked resultset (по умолчанию при более 100000 строк) даёт ленивый объект. После смены режима сбросьте workspace. `head(resultset)` читает начало, `read_resultset(offset,n)` - срез со смещением от нуля, `as.data.frame(resultset)` явно загружает всё. Прямой `$column`/`[[...]]` у ленивого объекта выдаёт объяснение, а не тихий NULL; `names`/`ncol` работают без чтения данных. Однопроходная обработка ограничивает память размером порции:

```
total <- 0
for_each_chunk(function(chunk, offset) {
  total <<- total + sum(as.numeric(chunk$amount), na.rm = TRUE)
}, chunk_size = 10000)
total
```

Streaming summary рассчитывает count/NULL/empty, sum/min/max/mean, выборочные SD/variance устойчивым объединением моментов. Квантили не рассчитываются. Analyze on PostgreSQL выполняет summary/group/frequency/missing на сервере с собственной numeric-арифметикой. В IDE поступает только агрегированный результат; лимит 100000 групп/16 MiB проверяется с ошибкой, без скрытого усечения.

### Выбор пакетов

Универсального лучшего алгоритма нет; набор покрывает разные свойства данных:

| Задача | Реализация | Когда выбирать |
|---|---|---|
| k-means | stats | Компактные числовые кластеры; nstart=25, seed |
| PAM / CLARA | cluster | Медоиды; CLARA использует выборки для больших наборов |
| Иерархическое дерево кластеров | stats hclust, Ward.D2 | Дендрограмма и разрез на k кластеров |
| DBSCAN / HDBSCAN | dbscan | Кластеры по плотности, выделение шума; HDBSCAN не требует epsilon |
| Дерево классификации/регрессии | rpart | CART, читаемые условия, importance и CP/CV |
| Отрисовка дерева | rpart.plot | Подписи условий и листьев; предусмотрен base R fallback |

Settings → R Runtime позволяет проверить реальную загрузку и установить эти четыре пакета с обязательными зависимостями из официального HTTPS-зеркала CRAN ETH Zürich. Системная и пользовательская библиотеки R не перезаписываются. Проверенные версии: cluster 2.1.8.3, dbscan 1.2.7, rpart 4.1.27, rpart.plot 3.1.5, R 4.6.2.

Кластеризация использует выбранные числовые признаки, полные конечные строки и опциональную стандартизацию. PAM/Ward ограничены 5000 строк из-за квадратичной памяти; для больших наборов используйте CLARA/k-means/плотностные методы. Модель использует полный допустимый dataset; назначения в таблице показаны до 100000 строк, scatter - до 10000, silhouette - до 5000, что явно подписано. Это не замена полной выгрузки назначений из собственной консоли.

rpart показывает узлы, importance, CP/CV, training accuracy/RMSE и график. Training метрика не выдаётся за независимое качество. Глубина, minsplit, cp и seed задаются в форме; целевая колонка исключается из предикторов.

## Сохранение и расписание

Saved R analyses поддерживает правку источника/параметров, копирование, обновление схемы SELECT через LIMIT 0, историю и восстановление версий. Колонки сопоставляются по имени и типу. Сценарий консоли сохраняет полный текст редактора и порционный режим, но не текущие переменные workspace.

Create scheduled analysis task копирует определение в задачу и задаёт HTML/JSON отчёт. Последующая правка сценария не меняет задачу. При редактировании задачи можно явно выбрать новую версию; digest расписания потребует повторной регистрации. Цепочка работает при закрытой IDE через Windows Task Scheduler для вошедшего пользователя. Неудачная предварительная проверка, ошибка R или неизвестный исход блокируют повтор до проверки. Отчёт записывается через временный файл; старый файл сохраняется при отмене. SQL и параметры подключения в отчёт не добавляются; собственный print/cat скрипта может, разумеется, выводить выбранные пользователем данные.

Обезличенный журнал ошибок сохраняется между запусками с ID операции, классом, SQLSTATE и методами приложения. Сообщения исключений, SQL и значения не сохраняются. Help → Collect diagnostics показывает содержимое перед экспортом; отправки нет.