

Журнал событий (Event Log)

Синонимы: Лог событий, Процессный лог, Журнал цифровых следов

Журнал событий — это перечень упорядоченных во времени событий (действий, статусов), выполняемых в рамках реализации некоего бизнес-процесса.

Примеры источников журналов событий:

- Информационные системы;
- Сетевое оборудование;
- Трекинг посетителей сайтов;
- Устройства интернета вещей;
- Мониторинг действий сотрудников на своих АРМ.

Данные из журналов событий являются источником информации для анализа по технологии Process Mining. События могут храниться в таблицах баз данных, журналах регистрации сообщений, архивах электронных писем, журналах транзакций. Типичными форматами для хранения логов событий являются формат CSV или Excel (для небольших логов), а также XES (eXtensible Event Stream) и MXML (Mining eXtensible Markup Language).

Для Process Mining большое значение имеет качество логов событий, так как результат анализа напрямую зависит от входных данных.

Существует несколько критериев оценки качества данных о событиях:

- События должны быть **достоверными**, то есть нужно быть уверенными, что записанные события действительно случились и что атрибуты событий являются верными;
- Лог событий должен быть **завершенным**, то есть предоставлять полную картину действий;
- У любых записываемых событий должна быть четко определенная **семантика**, то есть разнотипные события должны различаться.

Журналы событий можно рассматривать как совокупности экземпляров процессов, которые в свою очередь являются последовательностью событий в рамках выполнения одного экземпляра процесса.

Обязательные поля журнала событий:

1. **Process_ID** — определяет экземпляр процесса, т.е. перечень событий, объединенных одной меткой ID экземпляра процесса.
2. **Event** — упорядоченный перечень событий, выполняемых в рамках экземпляров процесса. В рамках лога «Event» является неделимой сущностью. События также

часто называют шагом процесса. Есть логи, в которых события заменены статусом, то есть результатом выполнения события. Например: товар принят, документы оформлены и прочее.

3. **Timestamp** — момент начала совершения события.

Фрагмент журнала событий с базовыми полями

Process_ID	Event	Timestamp
1	Приемка ж/д терминал	03.06.2021 13:59
1	Оформление приходных документов	03.06.2021 14:05
1	Оформление документов контрагентов	03.06.2021 14:26
1	Ввод данных контрагента	03.06.2021 14:35
1	Оформление приходных документов	04.06.2021 09:47
1	Заменить	04.06.2021 14:05
1	Оформление приходных документов	04.06.2021 14:05
1	Заменить	05.06.2021 08:28
1	Определение локации	05.06.2021 08:28
1	Вызов погрузчика	05.06.2021 08:29
1	Размещение груза	05.06.2021 13:40
1	Груз размещен	05.06.2021 13:41
2	Приемка ж/д терминал	04.06.2021 02:28
2	Оформление приходных документов	04.06.2021 03:54
2	Груз размещен	04.06.2021 05:09
3	Приемка	04.06.2021 03:02
3	Ввод данных контрагента	04.06.2021 04:45

Лог также должен соответствовать критериям качества. Некоторые из них:

- Лог должен быть отсортирован. Как правило, лог сортируется сначала по Process_ID, затем по Timestamp.
- Не должно быть пропусков указанных ключевых полей. Иначе потребуются или перевыгрузка, или удаление всего Process_ID с пропусками.

- Если один Process_ID содержит только одно событие, это служит «желтым» сигналом: необходимо перепроверить качество логирования.

Каждое событие имеет один или несколько атрибутов. Расширенный лог может содержать формально не ограниченный перечень дополнительных полей — атрибутов. Из значимых полей в логе могут быть:

- **Timestamp Fin** — момент завершения события
- **User** — пользователь, совершивший действие
- **Feature** — продукт, канал коммуникации, город и прочее

По журналам событий можно построить граф процесса и провести аналитику по процессу, сравнить процесс AS IS с проектом этого процесса и т.д.