

History Support

This page outlines the history support that is anticipated to be part of Activiti 5.0.

Note

This page currently describes a proposal for the history support only and does not reflect the current state of the implementation in Activiti.

Purpose

The history module provides access to historic process execution data and acts as a data provider for basic process execution reports and statistics. In an initial version, the history module should just provide enough information to feed simple statistical reports and queries about process instance / activity instance executions and their execution times.

Semantics: History vs. Audit

While history is focussing on the fact that and when a specific process or activity has been executed and how long it took to execute it, audit aims for providing comprehensive information on what exactly has been done during a process execution. Typically, the data provided by history builds the base for creating reports and statistics while audit aims for satisfying the regulatory/legal requirements of providing traceability and information on who changed what and from what value to what new value. Due to this different requirements, the natural data model for history and audit differ fundamentally: a history data model contains a consolidated view on executed process steps, optimized for querying, an audit data model in contrast focus on persisting the actual events together with the context data (user, condition evaluations, process data, ...) needed to trace the change.

Scope of the initial history support is limited to reporting/statistics, not audit. Based on the current process event bus infrastructure, an audit component could be built without any dependency or relation to the history support.

Reports

The following reports should be included in Activiti 5.0:

- # of process instances created within a time range, grouped by process definitions
- minimal, maximal, and average execution process instance execution time, grouped by process definitions
- minimal, maximal, and average execution time for activity instances for a specific process definition

Reports are provided by a specific report service, that in turn uses the history mechanism for accessing basic historic data.

These reports should be made available as a chart within Activity Probe or Explorer (TBD).

Data Model

The history data model consists of two main relations:

HistoricProcessInstance

- processInstanceId
- processDefinitionId

- `startTime`
- `endTime`
- *duration*

HistoricActivityInstance

- `activityInstanceid`
- `activityId`
- *activityName*
- *activityType*
- `processInstanceid`
- *processDefinitionId*
- `startTime`
- `endTime`
- *duration*

Italic attributes are there for providing a denormalized data view optimized for querying from external tools, directly based on the history data tables (e.g. activities can be grouped by `processDefinitionId` without the need for performing a join on the historic process instance table). As a consequence, the two relations are not linked in any way (i.e. no foreign key from historic activity instances to historic process instances). The same attributes and structure are reflected in the entities.

The history data model is built and maintained by specific event consumers that listen to process instances started/ended and activity instance started/ended event propagated by the process event bus.

Service

A historic data service will provide access to historic data in a read-only way. The history data service is configured globally per process engine and is accessible via the process engine configuration.

Query API

The query API allows for describing and executing queries against historic process and activity instances. The query API is designed in a "fluent" way, similar to the query APIs for tasks or jobs. Access to the query object is provided through the history data service. The query API is not intended to supersede direct access to the historic data tables, but provides access to historic data in a convenient way for programmatic queries.

The query API supports the following features:

- restrictions: by `processInstanceid`, `processDefinitionId`, `activityId`, `activityType` (combinations of restrictions possible, depending on historic data type), date ranges (start / end)
- sorting: by `startDate`, `processInstanceid` (single sorting only)
- grouping: by `processDefinitionId`