Tenant Id and Module Name Restrictions

<table>
<thead>
<tr>
<th>Status</th>
<th>IN PROGRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholders</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
</tr>
<tr>
<td>Due date</td>
<td></td>
</tr>
<tr>
<td>Owner</td>
<td>Julian Ladisch</td>
</tr>
</tbody>
</table>

Truncation problem

PostgreSQL silently truncates identifiers after 63 bytes.

The PostgreSQL schema name is `<tenant id>_<module name>`. Truncation may result in the same schema name, for example for mod-inventory and mod-inventory-storage.

Name clash problem

The PostgreSQL schema name is `<tenant id>_<module name>` with minus/hyphen converted to underscore.

Tenant id `foo` and module name `bar-baz` result in schema name `foo_bar_baz`.

Tenant id `foo-bar` and module name `baz` result in the same schema name `foo_bar_baz` resulting in a name clash.

Module name uniqueness problem

A PostgreSQL schema name is case insensitive, and the module converts minus/hyphen to underscore because minus/hyphen is not allowed in a PostgreSQL schema name.

However, Okapi’s module name is case sensitive and may contain lower and upper case letters. And it allows both minus/hyphen and underscore.

The module names `mod-foo` and `Mod_Foo` result in the same PostgreSQL schema name when created for the same tenant.

Reserved key word problem

PostgreSQL has a few reserved key words with underscore:

- CURRENT_CATALOG
- CURRENT_DATE
- CURRENT_ROLE
- CURRENT_TIME
- CURRENT_TIMESTAMP
- CURRENT_USER
- SESSION_USER

Trying to use them as a schema name results in a syntax error.

Example: If a module has name `user` then enabling it for tenant id `current` or `session` will fail.

Kubernetes and AWS ECS label problem

Kubernetes label names must follow the DNS label standard:

- contain at most 63 characters
- contain only lowercase alphanumeric characters or `.`
- start with an alphabetic character
- end with an alphanumeric character

Labels like `mod-inventory-storage-23-0-2` are commonly used and may exceed the maximum length if a long module name is combined with a long version number.

AWS ECS has these restrictions:

- Service Name: Up to 255 letters (uppercase and lowercase), numbers, hyphens, and underscores are allowed.
- TargetGroup Name: A maximum of 32 alphanumeric characters including hyphens are allowed, but the name must not begin or end with a hyphen.
The existing module name mod-data-import-converter-storage already exceeds this 32 character limit. Sysops have assigned a special TargetGroup Name for this module, such a workaround should be avoided.

Other restrictions

The problems stated above need to be solved.

More restrictions exist that are already enforced:

- Tenant id must not begin with a digit because PostgreSQL schema names must not begin with a digit. This is already enforced by PostgreSQL.
- Okapi restricts tenant id letters to be lower case and a-z. Accented letters and unicode characters are not allowed. (Okapi's regexp)
- Okapi doesn't allow a module name to contain a minus/hyphen followed by a digit because this starts the version suffix (Okapi's ModuleId parsing).

Solution

Proposed solution:

Limit tenant id to 31 bytes. Disallow underscore in tenant id. Regexp: `[a-z][a-z0-9]{0,30}`

Restrictions for back-end modules:

- Module name can contain only lowercase letters, digits and minus. Start with a letter. Disallow minus followed by a digit or a minus.
- Limit module name to 31 bytes. Disallow uppercase letters. Disallow underscore.
- Regex: `[a-z]([a-z0-9][-?[a-z]]){0,30}
- Disallow these module names: catalog, date, role, time, timestamp, user

Migration

Some FOLIO installations use tenant ids with underscore that is no longer allowed.

One back-end module fails the new module name restrictions:

- mod-data-import-converter-storage is 33 bytes long exceeding the length limit of 31 bytes.

Tenant name migration

To reduce the downtime in multi-tenant installations it must be possible to migrate one tenant at a time.

Okapi and the modules should provide APIs and/or scripts to do the migration.

**mod-data-import-converter-storage rename**

A shorter name for this repository and module name might be mod-data-import-conv-storage with only 28 bytes.

GitHub allows to rename a repository.

When the renamed module executes the tenant upgrade it checks whether the schema name with the old module name still exists. If yes it is renamed, the PostgreSQL ROLE is renamed and a new ROLE password is assigned if needed.

References

<table>
<thead>
<tr>
<th>Key</th>
<th>Summary</th>
<th>T</th>
<th>Created</th>
<th>Updated</th>
<th>Due</th>
<th>Assignee</th>
<th>Reporter</th>
<th>P</th>
<th>Status</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>OKAPI-1082</td>
<td>Restrict module names</td>
<td>🟢</td>
<td>Mar 04, 2022</td>
<td>Jun 13, 2022</td>
<td>Adam Dickmeiss</td>
<td>Adam Dickmeiss</td>
<td>🔄</td>
<td>BLOCKED</td>
<td>Unresolved</td>
<td></td>
</tr>
<tr>
<td>OKAPI-1081</td>
<td>Reject invalid tenant ids</td>
<td>🔴</td>
<td>Mar 03, 2022</td>
<td>Jun 23, 2022</td>
<td>Adam Dickmeiss</td>
<td>Julian Ladisch</td>
<td>🔄</td>
<td>BLOCKED</td>
<td>Unresolved</td>
<td></td>
</tr>
<tr>
<td>FOLIO-3450</td>
<td>SPIKE: Decide on the syntax tenantID and moduleName</td>
<td>✓</td>
<td>Mar 16, 2022</td>
<td>Jun 21, 2022</td>
<td>Julian Ladisch</td>
<td>Dilshod Khusanov</td>
<td>🔄</td>
<td>CLOSED</td>
<td>Done</td>
<td></td>
</tr>
</tbody>
</table>

3 issues