KIP-203: Add toLowerCase support to sasl.kerberos.principal.to.local rule

This page is meant as a template for writing a KIP. To create a KIP choose Tools->Copy on this page and modify with your content and replace the heading with the next KIP number and a description of your issue. Replace anything in italics with your own description.

Status

Current state: *Accepted*

Discussion thread: *here*

JIRA:

> ⚠ *Unable to render Jira issues macro, execution error.*

Please keep the discussion on the mailing list rather than commenting on the wiki (wiki discussions get unwieldy fast).

Motivation

From Kafka documentation

> "By default, the SASL *user name* will be the primary part of the Kerberos principal. One can change that by setting sasl.kerberos.principal.to.local.rules to a customized rule in server.properties. The format of sasl.kerberos.principal.to.local.rules is a list where each rule works in the same way as the auth_to_local in Kerberos configuration file (krb5.conf)
>
> Each rules starts with RULE: and contains an expression in the format **[n:string](regexp)s/pattern/replacement/g."**

Currently, username translations based on rules are done maintaining the case of the input principal. For example, given the rule

```
"RULE:[2:$1@$0](.*@EXAMPLE.COM)s/@.*//"
```

```
If the source string is joe-qa/host@EXAMPLE.COM, the result is joe-qa
If the source string is JOE-QA/host@EXAMPLE.COM, the result is JOE-QA
If the source string is Joe-QA/host@EXAMPLE.COM, the result is Joe-QA
```

However, this may not be desired given how different operating system handle usernames, where some are case-sensitive and some are case-insensitive. For example, Linux is case-sensitive and Windows is case-insensitive. To help with this issue, the rule can support to force the translated result to be all lower case.

Proposed Changes

We propose to extend "sasl.kerberos.principal.to.local.rules" config rule format/syntax by supporting an optional "/L" (toLowerCase) at the end of the rule. Since usernames are locale insensitive strings, we propose to use Locale.ENGLISH for the case conversion

 For example, given the rules

```
"RULE:[2:$1@$0](joe-qa-.*@EXAMPLE.COM)s/.*/JOE-QA//L",
"RULE:[2:$1@$0](JOE-QA-.*@EXAMPLE.COM)s/.*/JOE-QA-UPPER//L",
"RULE:[2:$1@$0](.*@EXAMPLE.COM)s/@.*///L"
```

```
If the source string is joe-qa-cl1/host@EXAMPLE.COM, the result will be joe-qa
If the source string is JOE-QA-cl1/host@EXAMPLE.COM, the result will be joe-qa-upper
If the source string is joe_user/host@EXAMPLE.COM, the result will be joe_user
If the source string is JOE_USER/host@EXAMPLE.COM, the result will be joe_use
```

However, it must be noted that this does not affect how pattern matches on input and therefore that will still be case-sensitive.

Compatibility, Deprecation, and Migration Plan

- This is an extension to the existing sasl.kerberos.principal.to.local.rules format and there will be no impact on existing users.