

Evaluation and Review of Talend ESB at Vuclip



Table of Contents

1 Introduction.....	3
2 Architecture.....	3
2.1 Current situation.....	3
2.2 Identified risks.....	4
2.3 Recommendations.....	5
2.3.1 Use the locator service.....	5
2.3.2 Use the Talend Administration Center (TAC).....	5
3 Job design.....	7
3.1 Current design issues.....	7
3.2 Recommendations.....	7
4 SDLC support.....	9
4.1 Gaps and risks identified.....	9
4.2 Recommendations.....	9
5 Issue list.....	11
5.1 General issue list.....	11
5.2 Issues raised during the evaluation.....	21



1 Introduction

Vuclip has selected Talend ESB as their ESB platform and have been using the Talend Open Studio for ESB (free version) in production for some time. However, in light of recent issues that have occurred in the production environment and with an eye on the future in terms of business growth Vuclip is considering the Enterprise version of the product.

This document aims to address the issues and questions raised by Vuclip, but also at the same time review and recommend best practices in the following areas:

- Architecture
- Job design
- Process in terms of SDLC
- Address list of current issues

2 Architecture

Please note that the aim of this section is only to highlight the some of the most obvious architecture issues and their potential improvements for the production environment. It can by no means be considered a complete architecture redesign.

2.1 Current situation

The development team currently uses the Talend Open Studio for ESB for developing the ESB jobs and in production the ESB runtime that comes with the Open Studio for ESB is used.

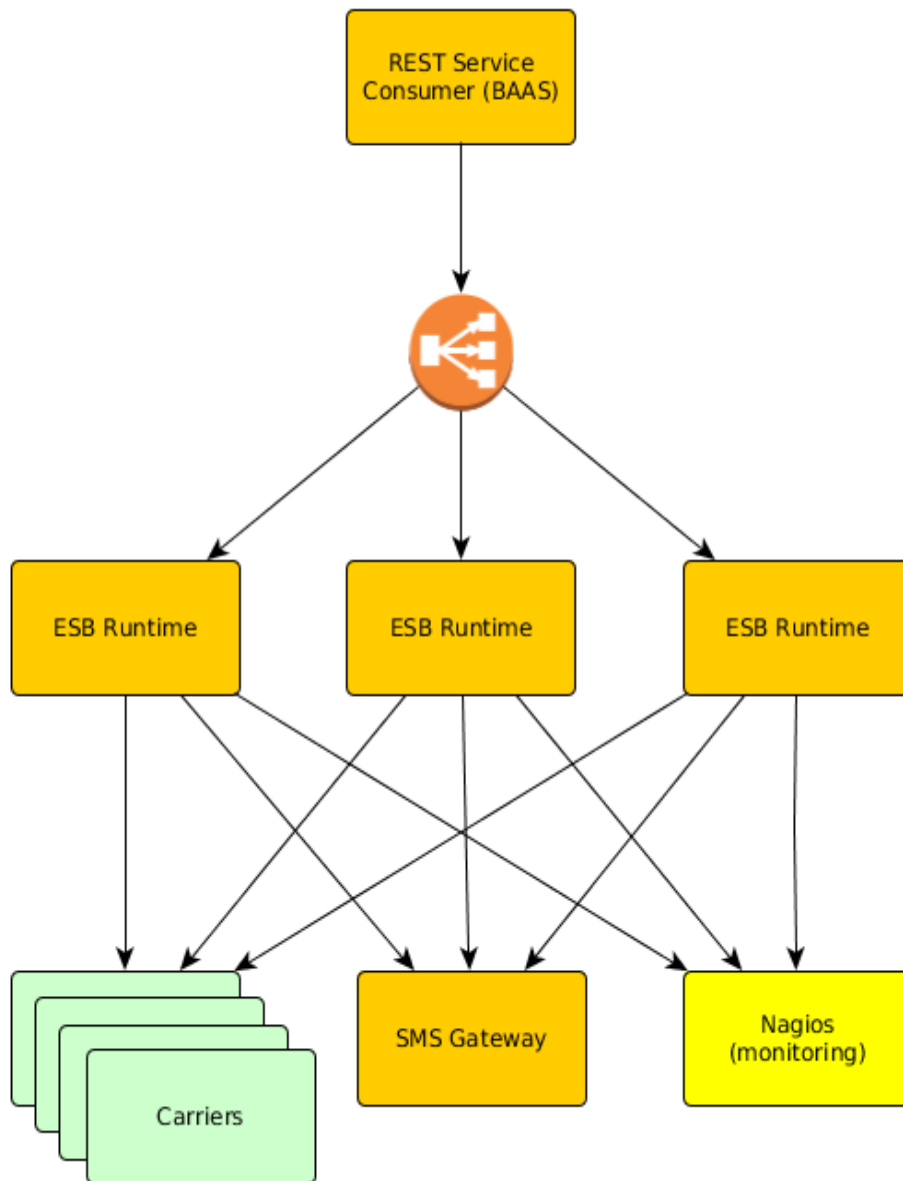
The internal BAAS system is the main consumer of the REST services exposed by jobs running on the Talend ESB. There are 3 runtimes fronted by a load balancer which dispatches requests coming from BAAS to the different runtimes. The 3 runtimes are identical in configuration and have the same jobs deployed.

The jobs in turn interface with external systems such as the carrier's services and the SMS gateway.

Monitoring happens on 2 levels.

First by the load balancer which monitors the ports on which the ESB servers are listening. If a port is not responding that server is considered down and all traffic is rerouted to the other two servers. Secondly be Nagios which monitors the individual REST services. If a service is not responding Nagios raises an alert and informs an administrator to take action.

The following diagram displays current high level architecture of the production environment.



2.2 Identified risks

The current architecture could lead to the following problems:

- When an individual jobs/service is down Nagios will raise an alert and will notify an administrator of the issue. It's then up to the administrator to take action to rectify the situation. Until then, the load balancer will continue to dispatch requests to the service even though the service is down. This could ultimately lead to a revenue loss.



- Different carriers have different levels of activity, but also different SLAs. Because all the jobs for all the carriers run on the same server, this means that activity for one carrier can impact the tps (transactions/sec) and potentially lead to fail agreed SLAs.
- Having the runtimes configured identically, with the same set of jobs running on each of the servers, could lead to scalability issues if the number of jobs deployed increases. In order to scale bigger or more powerful servers will be required.

2.3 Recommendations

2.3.1 Use the locator service

Talend ESB comes equipped with a locator service. The purpose of this service is to allow newly deployed REST services to register themselves with the locator service. A service consumer can then ask the locator service for the location of a particular service and the locator service can then redirect consumer to the service on a particular server.

Benefits:

- Minimum service downtime.
The locator service also monitor all registered services so as soon as a particular service stops responding the locator service will direct the consumers to services on different servers. This way immediate action is taken as soon as a service is down.
- Horizontal scalability.
Because with the service locator services can be deployed on any server the system can easily scale horizontally. And if a particular service needs to scale more than other services, additional instances can be deployed. For instance service A has 2 instances deployed and service B 4 instances because it's being hit more frequently.
- Multi-tenancy isolation.
Because service can now be deployed on any server, service related to a specific carrier can be pooled together on dedicated servers. This way the activity of one carrier does not impact the activity of another.

2.3.2 Use the Talend Administration Center (TAC)

The Talend Enterprise edition comes with TAC which is extremely well suited for managing and monitoring the ESB runtimes and the jobs deployed.



Not only does it make the deployment of jobs both manually and automatically much easier, it also integrates very well with the locator service to allow administrators to keep track of all services deployed. It's the best tool for managing the runtime and jobs in a production environment.

Some benefits that would benefit Vuclip in particular given the requirements:

- Full logging support. The Talend Enterprise edition has full Log4j support throughout the product suite up to component level, something that's missing in the Open Studio version. In addition, the Enterprise version comes with a log server comprised of Logstash, Elasticsearch and Kibana, taking trouble shooting and monitoring to a level required for production environments.
- Extensive (service activity) monitoring and statistics support. In conjunction with the service locator the Enterprise version supports full monitoring of all the endpoints deployed allowing details such as performance, request counters, response times, etc. to be tracked with ease on a per endpoint basis.



3 Job design

This chapter is to identify issues with the design of the current jobs and recommend some general best practices.

3.1 Current design issues

The following potential design issues have been identified:

- Chaotic design layout. The layout of most of the jobs are very chaotic in nature. It's very difficult to understand the design and follow the flow.
- Too much Java code. A lot of unnecessary Java code has been written.
- Use of a lot of unnecessary components have been included in the design creating unnecessary large designs regularly leading to the Java method size exceeded problem.
- Inefficient job constructs.
 - Job parts have been converted into REST services to avoid exceeding the Java method size limit.
 - Unusual components have been used to solve common problems. For instance tFlowToIterate (which breaks the flow) has been used to implement a simple join with results from a database query.
 - tJavaRow used when none were needed.
 - Copy paste of certain job constructs within the same job in particular the logging.
- Missing error flows. There are lots of places in the flows where errors can occur, however not many precautions have been taken to handle these potential errors.

3.2 Recommendations

- The design of a job is in a large part also its documentation. As such it's very important to pay attention to its flow layout and naming convention. As a rule of thumb the design should flow from top left to right bottom, with the main flow (happy path) being the longest horizontal flow. The components and connections should be named to match its functionality contributing to understanding the design.
- Reduce the amount of custom Java code.



- Reduce the use of tJava, tJavaRow and tJavaFlex components.
- Prefer the use of tJavaFlex over the use of tJavaRow.
- Extract reusable pieces of Java code into routines.
- Reduce the number of components as much as possible. Use the full functionality of the tMap and tXMLMap. Joins are the speciality of the tMaps, however this functionality is hardly utilised. Using the full functionality of the tMaps can drastically reduce the number of components.
- Copy paste logging constructs can easily be implemented in a single construct using either tLogCatcher or tAssert.
- Move functionality abstracted into REST services into subjobs and execute them using tRunJob.
- Create error flows. In particular in the tMap and tXMLMap utilise the reject flows to handle error flows. Another option is to use tFilterRow to capture exceptions and create error flows.
- Perform schema and value checks at the beginning of the flow only once and not multiple times throughout the flow such as null and empty string checks. In combination with this it's also good practice to standardise the data such as make everything lower case, or use single value where multiple are possible. For instance use only "F" for all potential values identifying a female.



4 SDLC support

4.1 Gaps and risks identified

- No proper version / release management
The team uses the Talend Open Studio for the development of the jobs which lacks support for a version control repository such as SVN or Git. As a result the developers develop the jobs in isolation and on store the final built of the jobs in share folder. This makes it difficult to keep track of the history and allows on one developer working on one job.
- No code sharing and developer collaboration
Because the team is not using a central repository such as SVN or Git there is no collaboration or shared code between the developers. It's not possible for multiple developers to work on the same jobs which seriously limits the productivity of the team.
- No use of reusable artefacts
Several parts of the jobs developed by Vuclip contain repeatable parts such as subjobs, schemas, Java code, etc. However, these parts have only been copied and pasted across jobs and not abstracted in to reusable assets.
- No testing and use of test cases
For none of the jobs developed test cases had been created. This seriously compromises the correctness, accuracy and stability of the jobs.
- No continuous integration support
All the jobs develop were manually built, deployed and tested. No automatic process is in place to perform these tasks.
- No vendor support. Critical issues that might appear in production environments may not be very easy to solve without the help of the Talend's technical team.

4.2 Recommendations

To fill the gaps identified in the previous paragraph it's strongly recommended to upgrade all environments to the enterprise version. The enterprise version has features filling the gaps identified above. The Talend enterprise edition comes with the following features:

- Support for Subversion and Git repositories
Allowing full version control of the jobs and collaboration within the team. Multiple developers can work on the same jobs.



- The Talend Studio repository containing reusable assets such as routines, schemas, connections, contexts, etc. can be easily shared and reused across multiple developers.
- The enterprise edition comes with remote commandline support which is ideal for continuous integration support, allowing build, test and deploy remotely from the commandline. In addition it also comes with several Maven plugins which can be used to build, test and deploy jobs in conjunction with a CI server.
- Technical support. For any mission critical system it's of the greatest importance to have a support contract in place so that Vuclip has direct access to the Talend technical team to address and fix any issues that might occur in the production system based on SLAs.
- The enterprise version also comes with support for unit test support for certain components. This allows the developers to create unit tests in particularly useful for complex components such as tMaps.
- It's also a good practice to write test jobs around subjobs that can be run standalone. In this case the test jobs is nothing more than a wrapper job around the subjob supplying it with some test data and validating the output of the subjob.



5 Issue list

5.1 General issue list

As part of the engagement Vuclip wanted a list of issues to be addressed. The issues in this list can be categorised in one of the following 2 categories, problems, demos and howtos.

Issue	Memory leakage in ESB runtime			
	Priority	Type	Status	Resolution
	High	Bug	Workaround	Waiting for bugfix
Description	<p>This issue has been reproduced on the evaluation version of the enterprise edition and is due to a bug in the Apache CXF libraries used by Talend. The bug has been submitted to the Talend support desk and a fix is pending. The workaround is as follows:</p> <p>As part of the bundle a (OSGI-INF/blueprint/job.xml) is being generated. At the bottom of that file there is the following section:</p> <pre><!-- Wiring the counter repository --> <bean class="org.apache.cxf.management.counters.CounterRepository"> <property name="bus" ref="cxf" /> </bean></pre> <p>Removing this section seems to resolve the problem, or at least not cause the out of memory exception in the short testing period performed.</p>			

Issue	Install and configure the Enterprise version including the Talend Administration Centre			
	Priority	Type	Status	Resolution
	High	Task	Completed	Installed
Description	<p>The evaluation version 6.1.1 of the enterprise version has been installed on a test server on which we ran the tests.</p>			

Issue	How to configure multiple data sources?			
	Priority	Type	Status	Resolution
	High	Problem	Solved	Configuration
Description	<p>Steps on how to configure one or more data sources on Karaf.</p> <p>Form the Karaf console execute:</p>			



	<pre>feature:install jdbc install -s mvn:mysql/mysql-connector-java/5.1.18 install -s mvn:commons-dbcp/commons-dbcp/1.4</pre> <p>Edit the example file <ESB RUNTIME PATH>/add-ons/datasources/dataservice/datasource-mysql.xml and copy it into <ESB RUNTIME PATH>/container/deploy</p> <p>Check the Karaf console and log file that the datasource has been deployed correctly.</p>
--	--

Issue	How to auto-reconnect if the database connection has been lost in connection pools?			
	Priority	Type	Status	Resolution
	High	Problem	Solved	Configuration
Description	This can be fixed by configuring JNDI data sources (see previous item) and add the following configuration property: <code>testOnBorrow="true"</code>			

Issue	Java method size exceeded?			
	Priority	Type	Status	Resolution
	High	Problem	Workaround	Restructure the jobs
Description	<p>Talend generates Java code and without any special precautions the size of the methods in which Talend generates code can exceed the size limitation of Java. There are several ways we to avoid this:</p> <ul style="list-style-type: none">• Keep the jobs as small as possible. Avoid any redundant components and Java code.• Use “OnSubJobOk” or “OnComponentOk” to break flow into sub sections and therefore limit the method size.• Use tRunJob to create reusable subjobs.• Abstract large pieces of custom Java code in to routines.			

Issue	Cannot access Log4j logger or enable it in the components.			
	Priority	Type	Status	Resolution
	Medium	Problem	Solved	Upgrade to Enterprise edition
Description	Full (log4j) logging capabilities are only available in the Enterprise edition. Limited workarounds for the Open Studio are possible using community components or including custom code.			

Issue	Output files with a dynamic names using variables from the global map
-------	--



	result in file names containing null values.			
	Priority	Type	Status	Resolution
	Medium	Problem	Workaround	By design
Description	<p>The location in the flow where the file component is initialised is very important. It needs to be in a location where all the globalMap variables have already been resolved.</p> <p>Using the tFlowToIterate component after the globalMap variable has been initialised and write to the output files within these iterations.</p>			

Issue	tRESTResponse waits for the entire flow including logging to finish (require processing after response sent).			
	Priority	Type	Status	Resolution
	Medium	Problem	Workaround	By design
Description	<p>Any processing that should take place after the tRESTResponse should be executed using either “On Subjob OK” or “On Component OK”.</p>			

Issue	How to create a custom component/joblet?			
	Priority	Type	Status	Resolution
	Medium	Question	Demo	Enterprise edition
Description	Joblets are only available in the enterprise edition.			

Issue	How to monitor the ESB jobs?			
	Priority	Type	Status	Resolution
	Medium	Question	Demo	Enterprise edition
Description	<p>Extensive monitoring can be done using the Talend Administration Center which is only available in the enterprise edition.</p>			

Issue	How to write an event listener (i.e. listening to a queue)?			
	Priority	Type	Status	Resolution
	Medium	Question	Demo	See example job
Description	Created a sample job using ActiveMQ.			

Issue	How to do post processing after a response had been sent?			
	Priority	Type	Status	Resolution
	Medium	Question	Demo	See example job



Description	Use OnSubjobOk or OnComponentOk to process after the response.
-------------	--

Issue	How to process asynchronous (i.e. send an SMS)?			
	Priority	Type	Status	Resolution
	Medium	Question	Solved	See example job
Description	Created a sample job using ActiveMQ.			

Issue	Log4j not being initialised			
	Priority	Type	Status	Resolution
	Medium	Problem	Resolved	Enterprise edition
Description	Full Log4j is only available in the enterprise edition. To enable it, in project settings tick the “Activate log4j in components” tickbox.			

Issue	How to include conditional (query) parameters in trestclient?			
	Priority	Type	Status	Resolution
	Medium	Question	Resolved	See example job
Description	QueryParamTestJob. You can add conditional parameters to the querystring of URL. Same can be done for post parameters. Set the name of the (post) parameter also a globalMap variable. If the name does not exist in the globalMap it will not be added to the parameters.			

Issue	TRunJob does not execute in the Runtime when parallel execution ticked.			
	Priority	Type	Status	Resolution
	Medium	Problem	Resolved	Not applicable to ESB
Description	Do not use “Use an independent process to run subjob” in an ESB (runtime) environment.			

Issue	Installing new versions of the job (in TAC) does not work. Old versions are still active.			
	Priority	Type	Status	Resolution
	Medium	Problem	Resolved	Workaround
Description	First stop and undeploy, then perform the update. A bug had been raised in Jira with Talend.			

Issue	TRunJob – How to retrieve a response back from the subjob?
-------	---



	Priority	Type	Status	Resolution
	Medium	Question	Resolved	See example job
Description	Write the output of the subjob into a tBufferedOutput component and in the parent job in the advanced settings of the tRunJob component tick the “Propagate the child result to the output schema” checkbox.			

Issue	Deleting a link from a tXMLMap will delete the schema.			
	Priority	Type	Status	Resolution
	Medium	Problem	Workaround	Use repository schema.
Description	Is a bug. Yet to be reported with Talend.			

Issue	Using tMySQLInput, the flow get stuck when the query return no results - TPAY			
	Priority	Type	Status	Resolution
	Medium	Problem	Resolved	Restructure job
Description	Use tMap to perform the join (using tFlowToIterate for this is a bad practice). In the tMap you can the use “Catch inner join rejects” to handle error flows in case no results match with the database lookup.			

Issue	Jobs remain locked (visible in TAC locks section)			
	Priority	Type	Status	Resolution
	Medium	Problem	Resolved	Unlock project from within Studio
Description	Locks can remain open after Studio ungracefully shuts down.			

Issue	Context problem. The runtime always picks the default context. How to override this? How to select an appropriate profile?			
	Priority	Type	Status	Resolution
	Medium	Problem	Resolved	Configuration
Description	Create configuration file in etc folder with name <service name>.cfg, add a line “context = <Context name>”			

Issue	Stop / undeploy of an ESB job does not work in the new installation			
	Priority	Type	Status	Resolution
	Medium	Problem	Resolved	Required restart of Karaf
Description	After restarting Karaf, the runtime started to behave normally.			



Issue	Sometimes same deployment which works on one machine, it doesn't work on the other machine			
	Priority	Type	Status	Resolution
	Medium	Problem	Open	Not been able to reproduce.
Description				

Issue	When server is restarted SOAP services do not come up. (?)			
	Priority	Type	Status	Resolution
	Medium	Problem	Open	Not been able to reproduce.
Description				

Issue	How to configure the tRestClient timeout in the runtime?			
	Priority	Type	Status	Resolution
	Medium	Question	Resolved	Edit runtime configuration file
Description	See: https://help.talend.com/display/TalendESBContainerAdministrationGuide61EN/11.2.2+HTTP+Conduit+OSGi+Configuration+Parameters			

Issue	How to import 3rd Party libraries in routines?			
	Priority	Type	Status	Resolution
	Medium	Question	Resolved	Use “Edit Routine Libraries”
Description	Select the routine in question. Right mouse click and select “Edit Routine Libraries”. Add all libraries you need.			

Issue	ESB Client url change issue in the settings of tESBConsumer component			
	Priority	Type	Status	Resolution
	Low	Problem	Bug	Report bug with Talend
Description	Yet to report to Talend, but currently easy workaround available.			

Issue	How to import components into component designer			
	Priority	Type	Status	Resolution
	Low	Demo	Completed	N/A
Description				



Issue	Debugging through studio			
	Priority	Type	Status	Resolution
	Low	Demo	Completed	N/A
Description				

Issue	How to do logging?			
	Priority	Type	Status	Resolution
	Low	Demo	Completed	N/A
Description	Enterprise edition feature.			

Issue	How to create different flows based on http headers?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	See example job
Description	See HeadersTestJob. Fetch the headers in tJavaFlex and create multiple flows in a tMap based on the headers.			

Issue	How to do a dynamic SQL lookup in an inner join map?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	See example job
Description	DynamicSqlLookup (use “Load on every row”) in the tMap or tXMLMap.			

Issue	Bundling of jars inside workflow. This means multiple copies one inside each jar.			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	By design
Description	Process/job isolation by including a required Java libraries within the job is a common practice and not considered a bad practice.			

Issue	How to externalize the profile?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Use context variables instead
Description	External profiles were required because it was unclear how to deploy a job with a specific context. Now that it is clear that contexts can be selected deploy time, it is not required to externalise the profiles.			



Issue	In Java we don't have access to the Context.			
	Priority	Type	Status	Resolution
	Low	Problem	Resolved	Pass as parameter
Description	Routines can only declare static methods. To work with context variables, pass it as a parameter.			

Issue	Date format gets converted from json format to xml.			
	Priority	Type	Status	Resolution
	Low		Recommendation	Always use your own date parser
Description	Do not rely on Talend's default parser.			

Issue	What is the Log4j version used by Talend?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Version 1.2.x used throughout
Description				

Issue	tLogRow doesn't print timestamp.			
	Priority	Type	Status	Resolution
	Low	Problem	Resolved	Enterprise edition
Description	Upgraded version uses Log4j throughout which will print timestamps.			

Issue	How to import custom WSDLs? You don't get the input/output xmls. This means we have to hand-code them.			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	tXMLMap workaround
Description	This issue was due to Talend not supporting an older version of WSDL. The workaround is to manually compose the XML into a string.			

Issue	How to get local access to the log file?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Configuration
Description	Configure Log4j appender to write logs locally. This is an enterprise edition feature.			



Issue	QueryParam with '-' (dash) produce a compilation error.			
	Priority	Type	Status	Resolution
	Low	Problem	Resolved	By design
Description	Limitation of Java (not a limitation of Talend). Talend convert the parameters into Java variable names, but Java does not allow variable names with '-'. 			

Issue	Notification section in TAC not showing the task list			
	Priority	Type	Status	Resolution
	Low	Problem	Resolved	Not applicable to ESB
Description	These tasks are only applicable to the DI product.			

Issue	Commandline console in TAC not showing any data			
	Priority	Type	Status	Resolution
	Low	Problem	Resolved	Not applicable to ESB
Description	TAC does not use commandline commands to execute ESB tasks			

Issue	How to generate GUIDs?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Use java.util.UUID in routines
Description				

Issue	How to do an OR in a join in tMap?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Do the OR in SQL not tMap
Description	<p>For this particular problem the OR should be performed in the database query itself to solve the problem.</p> <p>However, if in case a more generic OR functionality is required, this can be achieved by using a Java boolean expression to match the records that need to be included in the results in the tMap itself.</p>			

Issue	When to use "Store temp data"?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	When the lookup table is large
Description	Normally the lookup table is stored in memory. Using this option will reserve			



	memory and store the data on disk. This obviously has a performance impact. You can specify the location in the component details.
--	--

Issue	Backward compatibility. Is it available?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Yes
Description	But please also check upgrade/migration documentation			

Issue	tJava - All custom code disappears each time.			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Upgrade to version 6.1
Description	Validate if this is resolved in the Enterprise version			

Issue	How to configure tStat data in TAC?			
	Priority	Type	Status	Resolution
	Low	Question	Open	
Description	License expired. Could not be configured any more.			

Issue	No way to set a null value. It sets it to a String "nul". This means each if check is =null OR = "null"			
	Priority	Type	Status	Resolution
	Low	Problem	Open	Could not be reproduced.
Description				

Issue	Custom exception catcher			
	Priority	Type	Status	Resolution
	Low		Waiting for Vuclip	
Description	Waiting for an example			

Issue	How to perform a schema check by regex? TschemaComplianceCheck does not seem to support regex.			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Use the tFilterRow component.
Description	Tick the advanced checkbox. There you can put the condition in the form of a			



	regular expression. In the output flow you then can capture the rejects.
--	--

Issue	What are SQL Templates used for?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	In DB transformations (ELT)
Description	These are used in by the tSQLTemplate* components which are primarily used for in database transformations (i.e. one table to another)			

Issue	How to do remote debugging with Talend Studio?			
	Priority	Type	Status	Resolution
	Low	Question	Resolved	Not possible with Talend Studio
Description	Maybe with another IDE			

5.2 Issues raised during the evaluation

During the evaluation of the enterprise edition of Talend ESB the following issues have been raised.

- Extremely slow saving, building and running of the ESB jobs.
The internet connection within Vuclip is fairly slow and with the TAC server sitting in the US this severely affects the save, build and execute time of the jobs in Talend Studio. The full round trip time of build, run and test increased a lot making it sometimes unworkable.
- Talend Studio completely freezes in Ubuntu if the internet connection breaks.
It actually completely froze Ubuntu. The desktop had to be rebooted in order to get things working again. The internet connection got lost on a regular basis.
- Deployment of some jobs on the evaluation runtime would fail even though they would run without problem locally.
With the help of Talend Support the problem was solved by restarting the server. However, this caused a bit of a concern for Vuclip if the deployment of jobs would require a server restart.
- Installing the ESB runtime as a service failed. From the Karaf command “feature install wrapper” would never finish. Talend support suggested complete re-installation of the server to fix the issue. However, it was already towards the end of the evaluation period and thus this task was not completed.