HPCC Systems 7.4.x Releases

Welcome to the Red Book for HPCC Systems® 7.4.x series. There are several sections in this Red Book as follows:

- General HPCC Systems core platform
- Significant new features core platform
- ECL IDE

Users may benefit from glancing at other Red Book entries when making a large jump between releases.

Here's how to contact us if you find an issue or want to add something to the Red Book:

- To raise an issue, use our Community Issue Tracker. Please create yourself an account if you don't already have one, to get automatic updates as your issue progresses through the workflow.
- To ask a developer a technical question about something you are doing or have encountered, post in the Developer Forum.
- To add a note to the RedBook, please contact Lorraine Chapman with full details.

General HPCC Systems core platform

New - Encountering an issue where embedded Java fails

The platform does not currently always enforce a Java Development Kit (JDK) dependency and the embedded Java plugin (javaembed) can fail if a JDK is not available. If you encounter this error, install the appropriate JDK for your Ubuntu distro or the *openjdk-devel* package for Centos.

https://track.hpccsystems.com/browse/HPCC-25879

Thor jobs locking up when using SMART join

This may happen in specific circumstances, when using a SMART join that has failed over to a local hash join and the LHS is stopped early. The workaround if this issue is being hit in builds prior to this fix (7.4.34), would be to switch from using a SMART join, to either a regular HASH join, or a standard JOIN.

Note: Because this only happens when it has already run out of memory (and failed over to a local hash join), coding specifically as a HASH join is no worse performance wise.

https://track.hpccsystems.com/browse/HPCC-23068

Whitelisting feature in DALI

The upcoming release of HPCC 7.4.0 Gold introduces a feature to DALI called Whitelisting. This feature is switched on by default. The effect and purpose of this new feature is to prevent unauthorized DALI clients communicating with DALI. If an unauthorized client tries to connect to DALI it will be rejected with an "Access denied" type error.

All components defined in the local environment (i.e. the environment that Dali belongs to), will automatically be whitelisted. So the feature should have no visible effect if there is zero access to DALI from clients outside the environment. DALI clients that are not defined in the environment, will need explicitly whitelisting.

Examples of DALI clients that fall into this category are:

1. daliadmin and other admin tools that may be used to access remotely from various machines.
   Please note: All admin tools are automatically whitelisted to run from the DALI node by default. If you need an DALI admin tool to be run from other nodes, then a whitelist entry should be added to the environment.
2. Foreign file access from engines in external environments, i.e. via queries with ~foreign:: directives. The components that need access will need defining and adding to the source DALI whitelist.
   Please note: ~foreign restrictions will not be enforced until the next point release (7.4.2), but it is better to add whitelist entries ahead of time in preparation of that release.
3. An Esp from other environments (if foreign files are in super files and accessed/viewed by Esp).
4. DFU server(s) from other environments (if used for remote logical file copies).

Please read the following documentation for DALI's Whitelisting:

HPCC Systems(r) Administrators Guide

Please note: Entries can be added/removed and the environment updated in DALI without restarting the environment.

https://track.hpccsystems.com/browse/HPCC-22030

HPCC Systems now supports Spark 2.4

Pyspark supports UDF aggregation through pandas including default aggregation functions which extend Spark's analytics capabilities.

https://track.hpccsystems.com/browse/HPCC-21936

HPCC Systems now supports DataSource API Version 1

This is of interest to those who are using Spark with HPCC Systems.

The HPCC Systems DataPatterns Bundle is now integrated into the Standard Library

The DataPatterns features provide basic profiling and research tools. Prior to HPCC Systems 7.4.0. these features were available as a bundle from our GitHub repository. These features have now been integrated into the HPCC Systems Standard Library and are available without the need to install the separate bundle.