

# PDP-4: Support for Kylin

## Motivation

- Multidimensional analysis on PNDA, of the kind typically required to support presentation of results via an interactive UI or BI dashboard has previously been implemented using Impala/Hive or exporting a result set to a RDBMS such as MySQL. Each of these approaches works well in certain conditions but also has downsides.
- Apache Kylin™ is an open source Distributed Analytics Engine designed to provide SQL interface and multi-dimensional analysis (OLAP) on Hadoop supporting extremely large datasets.
- Apache Kylin uses pre-computation to offer O(1) wrt. dataset size & typically sub-second SQL query performance over large datasets that will not fit in memory.
- This complements Impala or Hive which do not require pre-computation but take much longer (10s of seconds to minutes) for such queries and are much faster when operating on datasets that fit into memory.

For more background, review their website: <http://kylin.apache.org/docs21/gettingstarted/concepts.html>

## Proposal

Provide the Kylin web UI as part of PNDA and integrate it with the PNDA components such as the PNDA Console and the PNDA Deployment Manager.

## Plan

- Run the Kylin web UI in PNDA
- Extend the Deployment Manager with the ability to create Kylin cubes as part of deploying a PNDA application
- Create some example applications
- Add a platform test module that monitors the health of Kylin
- Add a Kylin block to the PNDA Console

## Interfaces

- TODO

## Compatibility

- TODO

## Alternatives

- TODO