

Behind the Curtain

Development and evaluation of a systematic process to transform a network of observational data into the OMOP Common Data Model to enable end-to-end reproducible research

INTRODUCTION

- Before research can be conducted, patient-level data sets need to be identified, understood, acquired, processed, standardized, and evaluated.
- While individual data sets can vary greatly, transforming a network of observational data sources into the OMOP Common Data Model (CDM) can be described as a systematic process.

METHODS

- We introduced a workflow management layer into our process leveraging the Jenkins project.
- We developed pipelines within Jenkins to manage the end-to-end processing of our patient level data source transformations.
- To support post-processing we deployed the transformed data using Amazon Spectrum, which reduces cost and simplifies the infrastructure deployment needs.

RESULTS

- Pipelines provide structure and visibility to the steps required for successful integration of each patient level data set update
- Automated process removes the requirement for manual intervention and allows for processing to continue across disparate systems.

A reproducible pipeline-driven process can be applied across multiple databases to increase efficiency and improve data and research quality

RESULTS

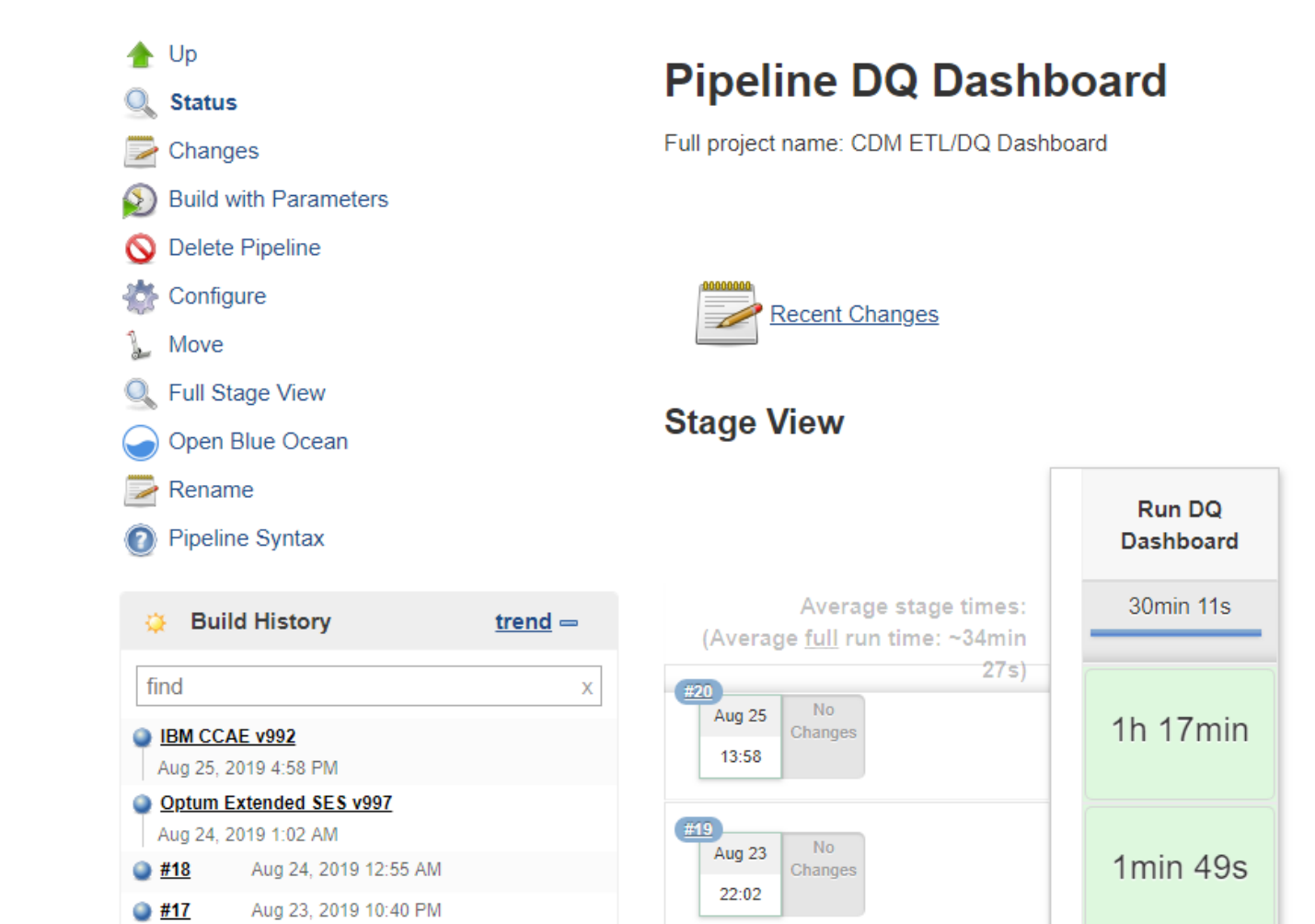
- The built-in performance and log capture allow for the review of build statuses, performance over time, and errors within the build steps
- Automating processes allows our team to focus on data quality and characterization efforts.



	Post processing: Achilles	Post processing: Achilles Export	Post processing: EARS
Average stage times: (Average full run time: ~7h)	30min 27s	4min 51s	5h 39min
Jan 27 10:17	219ms	203ms	1d 8h
Jan 27 09:31	219ms	204ms	12h 52min
Jan 27 14:47	42min 58s	24min 59s	227ms
Jan 27 14:09	2h 26min	11min 0s	
Jan 27 09:27	241ms	219ms	1s

DISCUSSION

- Jenkins pipelines leveraged our existing investment in cost-effective, serverless CDM builds
- We integrated our post processing steps to complete data source characterization and quality checks.



Frank J. DeFalco¹, Clair Blacketer¹, Ajit A. Londhe, MPH¹, Anthony Molinaro¹, Anton Ivanov²
¹Janssen Research & Development, Raritan, NJ; ²Arcadia, Saint Petersburg, Russia