

# CalPERS: Bringing Data Observability To World's Largest Retirement System

AND HELPING RETIREES RECEIVE THEIR PENSION BENEFITS ON TIME



# Overview

CalPERS is the California Public Employees Retirement System, and it's the largest pension fund in the US, with over six million participants. The system had around three terabytes of transactional data and hundreds of millions of records. It had to process pension based on data that was accumulated over decades of payroll history..

CalPERS had undergone multiple data conversions over the past 60 years including the one that was completed in the past year.

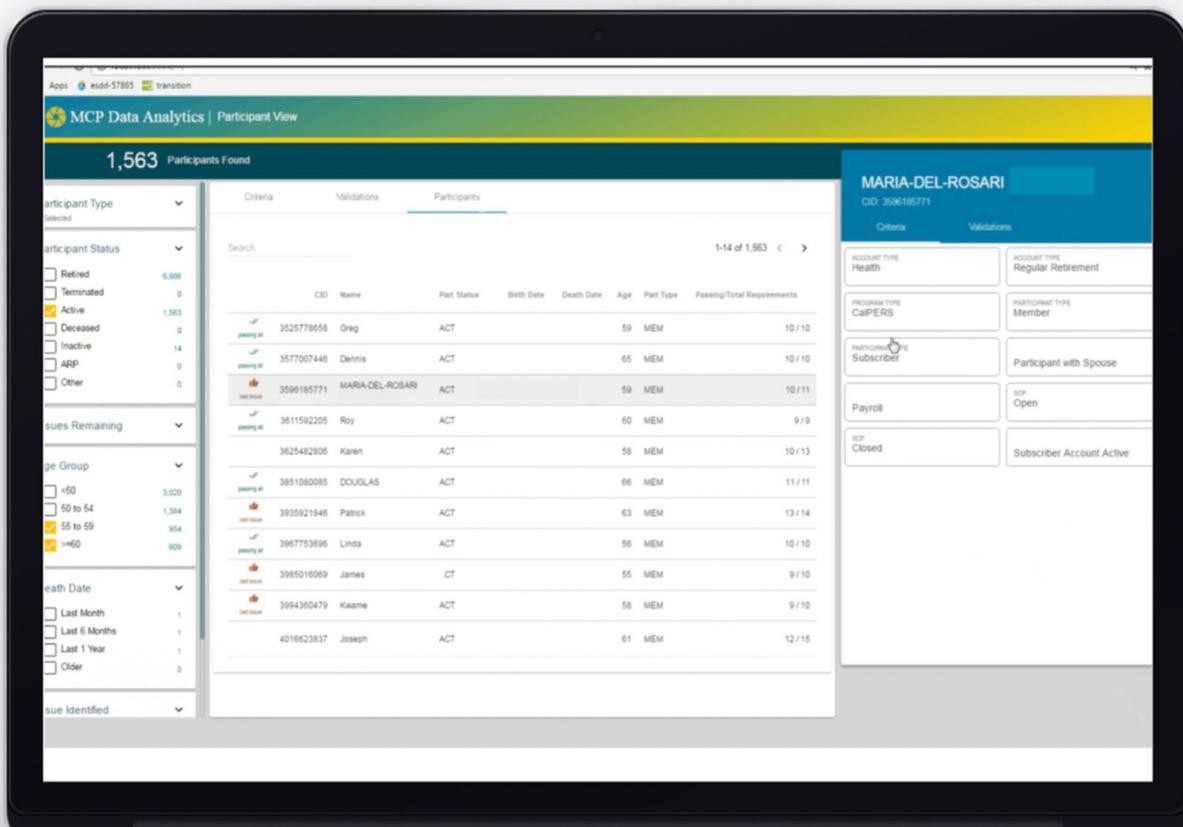
The CALPERS system can process cases with data entered through it (as that data is validated on data entry) but would not process cases that had data issues as those cases were converted from previous systems. Consistency was a challenge.

As people in the system requested retirement, the list of problems increased and the backlog of cases that needed by processed increased.

# The Challenge

With data cleanup and data validation teams working at max effort, the client was pressing to discover and solve data problems as quickly as discovered but they were still having trouble completing those fixes in time for retirements. Many retirements had to wait for months until they were processed manually, and the backlog was increasingly causing more problems.

The client knew that some of their issues needed to be fixed before others depending on whether the participant was likely to retire soon. They needed to discover the priority of the issues that they should work on but it was difficult to do so with the tools they had.



# Solution

The project involved building a data observability platform to enable CalPERS to quickly and easily navigate the impact of 950+ criteria and validations created by their quality team on participants.

An ETL pipeline ran a series of validations and criteria for participants and created a data model that tagged each participant with those attributes (whether the participant passes that validation or whether that criteria applies to the participant). The solution then uses search discovery paradigm to expose the criteria and validations as tags to the data.

# Outcome

Stakeholder dashboards that trended the quality of data - with an ability to drill down to individual rows - were created so that CalPERS could target their efforts on:

- High-priority records (participants who were retiring soon)
- Low-lying fruit (participants that have one issue remaining)
- Highest return on investment (those fixes which will resolve the most participant records)

*“This helps me to see the big picture of where issues are happening, and then I can easily identify where the team can make improvements. We can finally describe ”*

