

# Financial Services – Big Data Ingestion & Transformation

## 1. The Context

At the request of the Fair Work Ombudsman, our client was tasked with reviewing the implementation of their payroll system for the previous ten years and demonstrating that they had met their Enterprise Agreement obligations.

- Complexity enhanced by the movement of different employees across the in-scope period, and the different entitlements available at different times.
- Re-examination of the Enterprise Bargaining agreement revealed that previous interpretations of that agreement were mistaken. Consequently, sources of truth had to be rebuilt from fact tables up.

## 2. Our Challenge

RUBIX. was tasked with the regular ingestion of large, complex data sets such as the employee timesheet, and employee payslip tables. The largest of these tables was updated each month and consisted of over 150 million rows

Despite the AWS Redshift cluster having significant compute and storage resources available, careful management of the data pipeline was required for congestion and load balancing purposes.

- Different obligations required a focus upon fields, requiring a strategic to creation and preparation of different views.
- The production data required regular transformation and cleansing on a massive scale – up to 15 Terabytes of data per day

## 3. How we Triumphed

Recognizing the complex and dynamic requirements of the environment which we were working in, the RUBIX. team created several reusable resources which were utilized by both RUBIX and the BAU client team. Some of these assets include:

- Standardization of ways of working and outputs to ensure interoperability between issues.
- The creation of a shared set of python libraries to reduce development time and ensure consistency across teams.
- Automated data quality checks and output tests to ensure that backward compatibility and clear audit trail for the Fair Work Ombudsman.

We have deep knowledge and expertise from being recognized as a **trusted partner and preferred supplier of Data & Analytics** for two of Australia's Big 4 Banks.