Abstract

During the transformation phase of near real time ETL there could be some technique applied so that we get better results in terms of speed and accuracy. Transformation phase concentrates on changing the transactional data into semantically suitable format for the data warehouse. We try to bring in some of the solution during transformation phase that could enhance the speed and accuracy of the phase like advanced query optimization techniques, designing a new workflow so that we could reschedule some of the task. E.g. some functions applied on two parallel flows could be applied only once if the flows are converging. Also we look into some of the solutions for stream data how we could merge stream data and stored data, the challenges like speed and memory utilization. We also explore solutions like event based transformation for selected items, and handling of metadata efficiently so that it could add valued to the transformation phase.

Index Terms

Computer Science

Embedded Systems

Keywords

ETC, CBR, MDB