The Automatic Design of Batch Processing Systems

by

Barry Dwyer, M.A., D.A.E., Grad.Dip.

A thesis submitted for the degree of
Doctor of Philosophy
in the Department of Computer Science
University of Adelaide

February 1999

Abstract

Batch processing is a means of improving the efficiency of transaction processing systems. Despite the maturity of this field, there is no rigorous theory that can assist in the design of batch systems. This thesis proposes such a theory, and shows that it is practical to use it to automate system design. This has important consequences; the main impediment to the wider use of batch systems is the high cost of their development and maintenance. The theory is developed twice: informally, in a way that can be used by a systems analyst, and formally, as a result of which a computer program has been developed to prove the feasibility of automated design.

Two important concepts are identified, which can aid in the decomposition of any system: 'separability', and 'independence'. Separability is the property that allows processes to be joined together by pipelines or similar topologies. Independence is the property that allows elements of a large set to be accessed and updated independently of one another. Traditional batch processing technology exploits independence when it uses sequential access in preference to random access. It is shown how the same property allows parallel access, resulting in speed gains limited only by the number of processors. This is a useful development that should assist in the design of very high throughput transaction processing systems.

Systems are specified procedurally by describing an ideal system, which generates output and updates its internal state immediately following each input event. The derived systems have the same external behaviour as the ideal system except that their outputs and internal states lag those of the ideal system arbitrarily. Indeed, their state variables may have different delays, and the systems as whole may never be in consistent state.

A 'state dependency graph' is derived from a static analysis of a specification. The reduced graph of its strongly-connected components defines a canonical process network from which all possible implementations of the system can be derived by composition. From these it is possible to choose the one that minimises any imposed cost function. Although, in general, choosing the optimum design proves to be an NP-complete problem, it is shown that heuristics can find it quickly in practical cases.

Declaration

This thesis contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution. To the best of my knowledge and belief, it contains no material previously published or written by any other person, except where due reference is made in the text.

I give consent for this copy of my thesis, when deposited in the University Library, to be available for loan and photocopying.

Barry Dwyer
February 1999

Acknowledgments

Although I have to admit that I have learned a lot in the process, completing this thesis has taken me an excruciatingly long time. I am therefore grateful for the faith and patience of my supervisor, Prof. C.J. Barter. I am even more grateful for the support of my wife, Linda, who has often had to forego the pleasure of my company while I worked on it. In addition, I would like to thank those of my friends with whom I have discussed my work. Even if they thought they weren't helping me, the act of explaining my ideas helped me clarify many difficult issues. Finally, I would like to thank the reviewers of my thesis for suggesting how it could be improved. I have done my best to follow their suggestions faithfully.