

# USING DATA REPLICATION TECHNIQUES TO MAINTAIN DATA CONSISTENCY IN SUPPLY CHAIN APPLICATIONS

George Feuerlicht

*Faculty of Information Technology,  
University of Technology, Sydney,  
P.O. Box 123 Broadway, Sydney, NSW 2007, Australia*

Richard White, Glen MacLarty

*Eagle Datamation International Pty. Ltd., 184 Bourke Road, NSW, 2015, Australia*

**Keywords:** Supply Chain Management, Data Consistency, Data Replication.

**Abstract:** Supply chain applications typically involve sharing of large amounts of information across a number of partner organizations. Examples of supply chain applications are common in manufacturing, healthcare and global trade. Such applications can involve a large number of partners in different locations with various levels of connectivity, ranging from high-speed Internet to unreliable and slow dial-up connections. Maintaining data consistency in supply chain applications is an essential requirement as the loss of data consistency results in poor quality and unreliable information. Current supply chain applications mostly rely on EDI (Electronic Data Interchange) or similar electronic messaging standards to ship data between partner organizations. In general, little attention is paid to ensuring that data remains consistent as it is propagated and updated along the supply chain. In this paper we first discuss the requirements for data consistency in supply chain applications and then briefly describe a data consistency management framework based on data replication. We use Freight Forwarding and Customs Brokerage (FFCB) applications examples to illustrate our discussion.

## 1 INTRODUCTION

Most supply chain applications are characterized by fragmentation caused by dissimilar IT (Information Technology) platforms used to support the operation of individual partner organizations. Typically, business partners use different ERP (Enterprise Resource Planning) and SCM (Supply Chain Management) systems internally, and communicate electronically using EDI (Electronic Data Interchange) or similar messaging standards.

Quality of information has recently become particularly important in international trade supply chain applications as a result of increased security enforced by customs agencies on international shipping. Recent changes to the customs requirements by the United States mandate that information relating to individual shipments destined for the United States must be supplied to the United States Customs 24 hours prior to the shipment being

loaded aboard a vessel or an aircraft. Additionally, the United States Customs may also require information about the logistics process from any part of the supply chain. Effective sharing of accurate information among freight forwarders is essential in order to provide a streamlined global logistics system that satisfies such security requirements. Platform heterogeneity, different semantics used by business partners, and intermittent connectivity make maintaining data consistency across a supply chain difficult in practice.

Traditionally, shipment information accompanies the physical shipment of goods in the form of paper documentation necessitating re-keying of data by the individual freight-forwarders. This paper-based process is now inadequate as the delays incurred by the delivery of documents to customs agencies, cooperating freight forwarders or other relevant parties is no longer acceptable. Most customs agencies provide EDI interfaces that allow freight forwarders or customs brokers to submit

customs documentation electronically. This allows customs agencies to automate processing of incoming data and has a number of important benefits including reduced delays associated with customs processing and improved accuracy of data by eliminating typographical errors associated with data re-entry. Australian Customs now process around 99.5% of all customs import clearances electronically (Woodward, 2003).

Electronic data transfer does not guarantee data consistency as data is often modified as it travels along the supply chain without ensuring that the changes are propagated to all relevant participants. Consider, for example data used in a Freight Forwarding and Customs Brokerage (FFCB) applications. FFCB export and import documents contain almost identical data records and much of the information required by each party in the FFCB supply chain is similar to information either created or modified by previous parties in the logistic process. Information relating to the shipment is passed from the freight forwarder to packing and unpacking companies, customs agencies and brokers, shipping lines, and airfreight providers. Information changes that take place as data is transmitted from party to party along the supply chain include consolidation of data, and updating of customs details and delivery dates. As information flows follow the movement of the goods, updates are not propagated downstream to all the participants in the supply chain. This makes it difficult to share accurate and up-to-date information that is essential for planning the logistics process for individual shipments and ensuring that the locally held information is consistent among all participants in the supply chain.

FFCB applications are frequently deployed in large enterprises with many small to medium sized geographically distributed branch offices. Some branch offices are only accessible using a dial up connection over unreliable telephone lines. As a result, locally installed applications operate independently as islands of information leading to unnecessary re-keying of data with a significant potential for loss of data consistency. The lack of consistent data image across the entire enterprise results in poor visibility of information for decision support applications. Applications deployed in such highly distributed environments and without continuous connectivity need to rely on local copies of data to ensure data availability, as reachability of remote sites cannot be always guaranteed.

In this paper we first briefly review related work dealing with supply chain integration and data replication techniques (section 2), and then describe a replication management framework suitable for weakly-connected distributed environments such as

encountered in FFCB applications (section 3). In conclusion (section 4) we discuss the advantage and limitations of the proposed approach to maintaining data integrity in supply chain applications.

## 2 RELATED WORK

Information sharing in the context of supply chain applications has been studied under the topic of Supply Chain Integration (SCI) and an extensive body of literature is available on this subject. For example, Ball et al. present integration architecture and discuss a number of information sharing methodologies including RosettaNet ([www.rosettanet.org](http://www.rosettanet.org)) (Ball et al., 2002). Other authors focus on modeling the operations for supply chain integration using a Virtual Factory Modeling Approach that captures material and information flows through the business processes (Jain et al., 2002). Process analysis was used to study interaction between enterprises within a given industry and applied to multi-tiered supply chains (Liu, et al., 2005). A comprehensive review of application of EDI in managing inter-organizational relationships has identified a shift from using EDI for dyadic relationships to implementing complex networks relationships (Elgarah, et al., 2005).

However, relatively little research so far has addressed the problem from a data integration perspective. From this point of view, the problem of sharing information and maintaining data consistency across a supply chain constitutes a special case of a more general problem of managing information in distributed environments. Most techniques for maintaining data integrity in distributed environments rely on some form of data replication. Replication techniques that are relevant to FFCB supply chain scenarios need to deal with intermittent connectivity and support update conflict resolution based on application semantics. We note here that data replication techniques are mainly relevant to situations where individual participants use standardized data structures to represent the information that is transmitted along the supply chain. This arises when organizations adopt industry standard schemas, or use the same supply chain applications that enforce identical data schemas for all participants. Addressing issues related to disparate data structures and semantics used by individual supply chain participants requires the consideration of schema transformation techniques and is outside the scope of this paper.

In the following sections we review replications techniques suitable for maintaining data consistency

in loosely coupled and weakly (intermittently) connected applications.

## 2.1 Data Replication in Weakly Connected Distributed Applications

Data replication has been studied extensively over the last two decades mainly in the context of distributed databases (Ceri and Pelagatti 1985), (Buretta, 1997), and distributed file systems (Sandberg, et al., 1985), (Satyanarayanan, et al., 1993), (Steen et al., 2002). Most commercial DBMS (Database Management Systems) products provide extensive support for data replication (Urbano, 2003), but commercial DBMSs are not designed to operate in weakly connected supply chain environments. The use commercially available systems requires that the same DBMS platform is adopted by all supply chain participants as the systems use proprietary protocols, and this is not likely to happen in practice. Most DBMS replication solutions assume continuous and reliable network connectivity, and tend to have only limited capability for resolving data conflicts using application semantics. These limitations make commercially available solutions unsuitable for highly distributed and weakly connected supply chain applications.

Data replication and associated methods for the management of data consistency in widely distributed and weakly connected network environments such as those encountered in mobile computing and in data grid applications have been extensively investigated (Hirsch, 2001), (Johnston, 2003), (Xiao, et al., 2003). Weak network connectivity arising from using dial up connections (or disconnection in mobile applications) implies that data changes need to be propagated asynchronously, i.e. when the connection becomes available. Particularly challenging are application scenarios that involve complex topologies of branch and sub-branch locations and highly variable levels of connectivity between individual sites that characterize FFCB supply chain applications. In such applications, connectivity may only be established intermittently on a pair-wise basis, for example between a head office and a local branch office, or between two partner systems within a supply chain (e.g. exporter and importer) but almost never at the same time across the entire system. The resulting network partitioning requires that the replication model supports operation of disconnected workgroups, i.e. groups of closely related sites. This type of replication support is not generally available

in commercial DBMS systems, which are designed to tolerate only occasional network failures, but do not operate effectively in partitioned environments with intermittent connectivity.

Extensions to the distributed database approach that address intermittent connectivity have been proposed in the literature (Pitoura et al., 1999) Pitoura et al. describe a system for managing data consistency among intermittently connected nodes based on distributed database techniques. Strong and weak operations can be specified and are selected for use according to the type of connectivity available. Disconnected nodes may only utilize weak operations to ensure data consistency.

Supply chain applications are more suited to replication models developed for mobile computing applications, which are characterized by long and frequent periods of disconnection between sites and by autonomous operation on local data during periods of disconnection.

## 2.2 Update Conflict Detection and Resolution using Application Semantics

To maintain high levels of data availability in individual locations along the supply chain, applications need to be able to read and write to locally stored data (i.e. a local replica), and cannot be restricted to a single updatable master copy potentially stored in a remote location. Multiple updateable replicas introduce the potential for update conflicts that arise in situations where multiple copies are updated within the same latency period (i.e. before re-synchronization takes place). For example, the exporter site may change the delivery address and the importer site the contact person's name in the same customer record. Update conflicts can be avoided in some application situations, for example using a suitable data partitioning design based on local data ownership, but these techniques are not generally applicable to complex application scenarios such as considered here. In situations where update conflicts are allowed to occur, reliable detection and resolution method is needed to ensure that data consistency is eventually regained across the entire distributed application. Applications using such data must be aware of the level of data consistency, i.e. they must be able to detect conflicts and must be involved in conflict resolution, as both conflict detection and resolution depend on application semantics (Terry et al., 1995). Application semantics (i.e. business rules) determine how a particular update conflict should be treaded and resolved. This can range from simple

time-stamp based conflict resolution strategies to complex business rules that determine how data consistency should be re-established given a particular combination of conflicting data values.

### 2.3 Replication Techniques in Mobile Computing

Recent interest in replication techniques has been largely motivated by the extensive use of mobile devices (Noble et al., 1998), (Madria, et al., 2001). As mobile devices are becoming increasingly more prevalent the ability to access data on such devices is gaining importance. Such techniques are highly relevant to the problem of maintaining data consistency in supply chain applications as mobile applications share many of the characteristics of loosely-coupled weakly connected supply chain applications. Several prototype systems have been proposed and implemented to investigate management of update conflicts in weakly connected mobile computing environments, including the Bayou system developed at the Computer Science Laboratory, Xerox Palo Alto Research Centre (Terry, et al., 1995). Similar to the requirements of the types of distributed applications considered here, mobile computing is characterized by long periods of disconnection, and the use asynchronous replication models with update conflict detection and resolution based on application semantics. The Bayou system, adopts a model in which clients can read and write to any replica without the need for explicit coordination with other replicas. Every site eventually receives updates from every other site, either directly or indirectly, through a chain of pair-wise interactions, re-establishing data consistency across the entire system. Clients can read data at all times, including data whose conflicts have not been fully resolved, and the application is made aware of the state of the replica (i.e. level of consistency of the data that is being read). Bayou prototype uses update dependency check and merge procedures providing a general mechanism for application-specific conflict detection and resolution. Data remains in tentative state until the conflicts potentially introduced by update operations are resolved. The Bayou prototype focuses on reliable conflict detection and automatic resolution using application-based semantics providing a mechanism for the replicas to move towards eventual consistency.

### 3 MAINTAINING DATA CONSISTENCY IN FFCB SUPPLY CHAIN APPLICATIONS

FFCB applications represent a specific sub-class of supply chain applications that are concerned with global trade and logistics, and have specialized requirements for data integration. As noted earlier, FFCB applications are characterized by simultaneous material and information flows, i.e. shipment of (material) physical goods is accompanied by the transmission of information in paper or electronic form. Information that needs to be managed includes relatively static data relating to individual supply chain participants (e.g. partner company details, etc). Partner company details contain information about the various parties involved in the supply chain such as the supplier and the buyer of the goods and parties involved in packaging, local transportation, warehousing or consolidation of the shipment within a larger shipment. This type of information is relatively stable and is unlikely to change during the lifetime of a shipment. Shipment details that contain information related to a given shipment such as customs clearance information, consolidation details and data related to estimated or actual delivery times of the shipment are typically subject to higher volatility making it more difficult to ensure that correct data is available to all parties.

A key characteristic of FFCB supply chain applications is that data is propagated in one direction in a pair-wise manner (e.g. from exporter to importer) using occasionally connected partner systems. Data can be updated in multiple locations along the supply chain impacting on multiple partner organizations. EDI and similar messaging approaches are essentially point-to-point solutions that do not facilitate the propagation of data changes to multiple locations within a supply chain. There is no mechanism for re-synchronization of common data to ensure data consistency. In order to avoid loss of data integrity data duplication must be controlled using replication techniques that re-synchronize data across the entire supply chain. Furthermore, the inability to maintain continuous connectivity across the supply chain results in network partitioning that makes it difficult to provide a consolidated view of the data resource for management reporting purposes. As a result of this fragmentation the level of integration within FFCB logistics supply chains tends to be very limited with

each partner organizations maintaining their own data in isolation.

### 3.1 FFCB Data Replication Scenario

As individual supply chain partner organizations (or their branches) maintain their information autonomously the most appropriate replication model for this situation involves multiple-master (symmetric) asynchronous replication. This replication model allows updates to multiple copies (replicas) of the same data records and requires the synchronization of common data across all participating sites. As the operation of the replication system is asynchronous, there is a potential for data conflicts that arise when the same record is updated in multiple locations within a single latency period. A given shipment record is typically managed concurrently by an exporter who is arranging export clearances and an importer who is arranging import clearances. If both partners update the details of the same shipment, (e.g. the export forwarder updates details relating to the outgoing shipment such as customs clearance reference numbers, and the import forwarder updates details such as import duties, quarantine details, and cartage details) the record becomes inconsistent.

### 3.2 Update Conflict Resolution

Several standard conflict resolution strategies are available including time-stamp resolution strategy that preserves the most recent update, priority-based resolution based on location of the update or user privileges, and minimum/maximum value resolution based on the value of the field being updated. More complex resolution strategies can be defined using combinations of standard strategies, or program logic based on application semantics.

### 3.3 Replication Management Architecture

In this section we briefly describe the proposed architecture of the replication management and consistency maintenance framework. As noted above (in section 2) we assume that all participating sites use identical database schemas and we do not consider structural and semantic data heterogeneity in our proposal. This situation arises for example when all participants use the same FFBC software package, or when the database schema conforms to an industry standard schema specification. While this may not be the prevailing practice at present,

there are indications that industry-wide standardization of data schemas in the form of XML schema standards is becoming increasingly more popular.

The proposed replication framework consists of four logical layers: 1) Business Rules Manager (BRM), 2) Update Conflicts Detector (UCD), 3) Replication API and database layer, and 4) the Transport Layer. The BRM layer provides support for the definition, storage and modification of application business rules that determine the consistency of the underlying database. The business rules are used to detect and resolve update conflicts and can be managed independently from the applications that use the rules. The UCD uses the BRM to identify data that is potentially inconsistent and is responsible for informing the application about the status of the data. The UCD initiates data re-synchronization at a suitable time (e.g. when connection between sites is re-established) and selects appropriate conflict resolution strategy based on the relevant business rules. The Replication API (Application Programming Interface) interacts with the underlying DBMS to re-synchronize replicated data via a reliable queue. Finally, the transport layer implements reliable queuing and performs the transformation of data into a suitable format for transmission. The transport layer is also responsible for discovery of currently connected partner sites, and initiating pending synchronization processes.

## 4 CONCLUSIONS AND FURTHER WORK

The requirements for supply chain integration are driven by the need for closer collaboration between individual business partners and will eventually result closely integrated extended enterprises. The need to maintain data consistency across the entire supply chain is emerging as an important challenge to developers of supply chain applications. We have argued that the maintenance of data consistency in supply chain applications requires data replication support using replications methods that can tolerate periods of network disconnection and re-establish consistency using data synchronization techniques suitable to weakly connected environments. We have briefly described a high-level replication management architecture that includes update conflict detection and resolution component based on application semantics (business rules). This work was initially conducted in the context of an Industry Link Seeding Project at University of Technology, Sydney (Feuerlicht, 2004) and is subject of ongoing

research. While the proposal outlines a high-level architectural framework for replication management in supply chain applications, a number of research issues need further investigation, including the selection of a suitable method for conflict detection and resolution that minimizes latency between consistent states, specification of the replication API and development of a method for the management of business rules.

## REFERENCES

- Ball, M. O. Meng Ma, Louiqa Raschid, Zhengying Zhao, 2002, Supply Chain Infrastructures: System Integration and Information Sharing. SIGMOD Record 31(1): 61-66
- Buretta, M. Data Replication, Tools and Techniques for Managing Distributed Information, Wiley Computer Publishing, ISBN 0-471-15754-6, 1997
- Ceri S. and Pelagatti G., Distributed Databases: Principles & Systems, McGraw Hill, 1985.
- Elgarah, W., Falaleeva, N., Saunders, C. C., Ilie, V., Shim, J. T., and Courtney, J. F. 2005. Data exchange in interorganizational relationships: review through multiple conceptual lenses. SIGMIS Database 36, 1 (Feb. 2005), pages 8-29. <http://doi.acm.org/10.1145/1047070.1047073>
- Feuerlicht, G. (2004) Maintaining Data Consistency in Complex Replication Environments, Industry Link Seeding Project, University of Technology, Sydney, Eagle Datamation International (EDI), 2003-2004.
- Hirsch, R., Coratella, A., Felder, M., & Rodriguez, E. (2001) A Framework for Analyzing Mobile Transaction Models. Journal of Database Management, Vol. 12, No. 3, pp. 36-47.
- Jain, S., Choong, N. F., and Lee, W. 2002. Manufacturing supply chain applications: Modeling computer assembly operations for supply chain integration. In Proceedings of the 34th Conference on Winter Simulation: Exploring New Frontiers (San Diego, California, December 08 - 11, 2002). Winter Simulation Conference. Winter Simulation Conference, 1165-1173. Year of Publication: 2002, ISBN:0-7803-7615-3
- Johnston, D. B. 2003, The Next Generation in Application Architecture: Occasionally Connected Computing [Online]. Available: <http://crm.ittoolbox.com/documents/document.asp?i=3260> [Accessed 23rd July 2004].
- Liu, P. and Hsieh, Y. , 2005. A study based on the value system for the interaction of the multi-tiered supply chain under the trend of e-business. In Proceedings of the 7th international Conference on Electronic Commerce (Xi'an, China, August 15 - 17, 2005). ICEC '05, vol. 113. ACM Press, New York, NY, 385-392. <http://doi.acm.org/10.1145/1089551.1089623>
- Madria, S. K., Bhargava, B. K. "A Transaction Model to Improve Data Availability in Mobile Computing", Distributed and Parallel Databases, 10(2): 127-160, 2001.
- Noble, Brian D. Mobile Data Access, PhD Thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, May 11, 1998, CMU-CS-98-118
- Pitoura, E. & Bhargava, B. 1999, 'Data Consistency in Intermittently Connected Distributed Systems', IEEE Transactions on Knowledge and Data Engineering, vol. 11, no. 6, pp. 896-915.
- Sandberg, R., Goldberg, D., Kleiman, S., Walsh, D., Lyon, B. Design and Implementation of the Sun Network Filesystem. Summer Usenix Conference Proceedings. 1985.
- Satyanarayanan, M., Kistler, J., Mummert, L., Ebling, M., Kumar, P., Lu, Q. Experience with disconnected operation in a mobile environment. In Proceedings of USENIX Symposium on Mobile Location-Independent Computing (Cambridge, Massachusetts, August 1993)
- Steen, V., Tanenbaum, S. Distributed Systems: Principles and Paradigms, Prentice Hall; January 15, 2002, ISBN: 0130888931
- Terry, D. B. M. M. Theimer, K. Petersen, A. J. Demers, M. J. Spreitzer, and C. H. Hauser, (1995) "Managing update conflicts in Bayou, a weakly connected replicated storage system," in Proceedings of the 15th Symposium on Operating Systems Principles, Copper Mountain Resort, Colorado, December 1995, ACM, number 22, pp. 172--183.
- Urbano, R. 2003, Oracle Database Advanced Replication, 10g Release 1 (10.1), Oracle Corporation, Redwood City, CA.
- Woodward, L. B., 2003, Annual Report 2002 - 03, Australian Customs Service, Canberra.
- Xiao Qin and Hong Jiang, Data Grid: Supporting Data-Intensive applications in Wide-Area Networks, Technical Report No. TR-03-05-01, May 2003, University of Nebraska-Lincoln, NE 68588-0115