



## Comparison of Differencing Technologies

Dinesh Sinha, Chitra Ramaswamy and Viraj Ajgaonkar

## TABLE OF CONTENTS

<b>1. Abstract .....</b>	<b>3</b>
<b>2. Glossary .....</b>	<b>4</b>
<b>3. Introduction .....</b>	<b>5</b>
<b>4. How it works?</b>	
<b>3.1 Block Level Differencing explained... ..</b>	<b>7</b>
<b>3.2 Byte Level Differencing explained... ..</b>	<b>8</b>
<b>3.3 Write Monitor Differencing and File Specific Monitoring explained... ..</b>	<b>9</b>
<b>4 Comparison .....</b>	<b>10</b>

---

© 1998 - 2005 Mobiliti, Inc., Edison New Jersey. All rights reserved.

Network/Unplugged, the Mobiliti logo, and any other images associated with the software are trademarks of Mobiliti, Inc., Edison, New Jersey.

Microsoft Windows, Windows 95, Windows 98, Windows NT, Windows 2000 and MS-DOS are registered trademarks of Microsoft Corporation in the United States and/or other countries. Other brands and product names are trademarks and registered trademarks of their respective holders. Patent pending.

## **1. Abstract**

File differencing can be one of the methods of Comparing two files and finding the difference between the two files and synchronizing the two files by only transferring the differences. Mobiliti Inc. uses five different differencing techniques which make incremental, real time backups and synchronization possible. Along with the differencing techniques Mobiliti products have iDESP (Intelligent Data Selection Process) technology which decides which differencing technique is best suited for a particular file.

This paper starts by defining some common terminologies. Secondly, it gives a brief description of what a Differencing Technology is. Then it explains in detail all the five Differencing Techniques and how they work. Lastly it compares all the Differencing technologies.

## 2. Glossary

**Source File:** The primary user file.

**Destination File:** A copy of the source file.

**Backup:** Backup is the process of making the destination file same as the source file by overwriting the destination file with the source file..

**Source Computer (Source):** Source computer is the computer on which the source file resides.

**Destination Computer (Destination):** Destination Computer is the computer on which the destination file resides.

**Synchronization:** Synchronization is the process of overwriting the older of source or destination file by the latest version.

**Source Network:** Source Network is the network in which source computer resides.

**Destination Network:** Destination Network is the network in which the destination computer resides.

**Link:** Link is the type of connection between source and destination computer.

**Source Process:** Source process is the process running on the source computer which backups or synchronizes the source file to the destination file.

**Destination Process:** Destination process is the process running on the destination computer which helps in backing up or synchronizing the source file and the destination file. The destination process is optional for many differencing techniques.

### **3. Introduction**

With any of the differencing technique, only the changes made to a file, not the entire file, are transferred across the network. This increases the speed of synchronization and file backup over slow network connections such as WAN, while dramatically reducing network bandwidth consumption. These differencing techniques reduce synchronization time by as much as 95-99%. The benefit to the enterprise is reduced synchronization and backup time, data communication costs, reduced network traffic, and increased user productivity.

The 5 different techniques of differencing are described below.

**Block Level Differencing:** In this method, the files being synchronized are broken down into blocks on the source computer (source) by source process and the destination computer (destination) preferably by destination process. A comparison is made of the blocks on both sides and the differences down to the block level are generated. These differences are then transferred during synchronization and backup. This technology is applicable for all file types but it needs a component to be installed on both sides (i.e. source process and destination process)

**Byte Level Differencing:** In this method, the files being synchronized are compared and differences down to the byte level are generated. This method produces much smaller differences, which are then transferred to the destination computer during synchronization and backup. However, it is necessary to maintain a reference file for the destination and the source to calculate the byte level difference. Also, it is best to have a process on both sides (i.e. source process and destination process).

**Write Monitor Differencing:** In this method, the changes or modifications made to the files by the applications modifying the files are logged. At the time of synchronization and backup, the "writes" that have been logged are condensed by removing any duplicate "writes", compressed and transferred across the network. This method is file format independent as the log contains the sequence of writes done on the source from the file system perspective. Example: This method works well with Outlook .pst files, and Access Database .mdb files.

## Comparison of Differencing Technologies



**File Specific Monitoring:** This method is very similar to write monitor differencing except for the fact that the writes log contains record in the actual file format.

Two things to consider in this method are that:

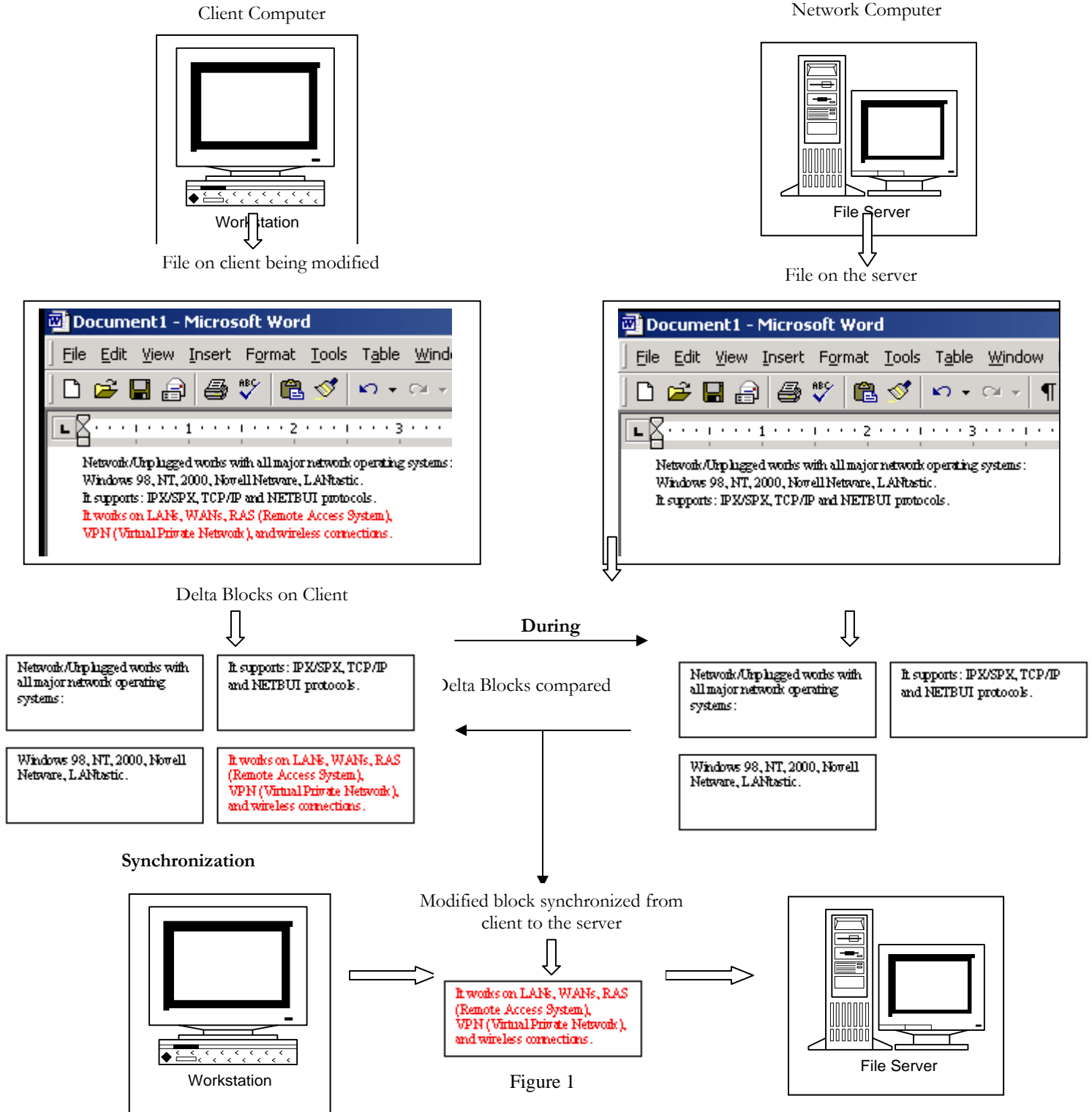
- 1) The application should provide a way to trap the writes. Besides, even the log will contain records in the actual format.
- 2) This method is not file format independent. Though this method is not file independent, but since it understands the file format, it can do a 2 way merging.

**Compression:** This is not a true differencing technique but is responsible for cutting down the band width. However, the advantage with this method is that it works well even when there is nothing in the destination (e.g. first time copy) and it can work together with any other differencing technique to enhance the performance even further.

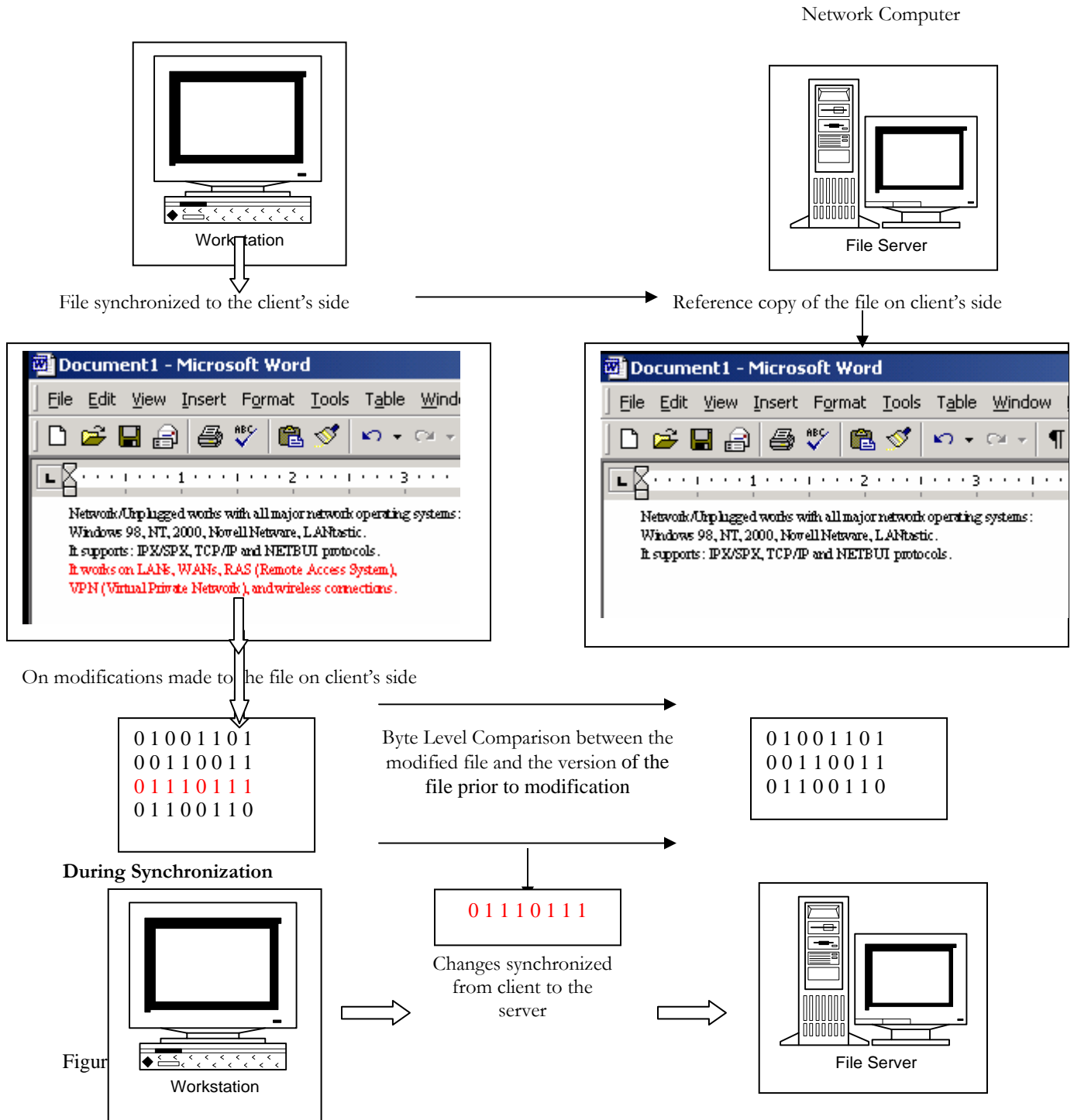
The following sections give a graphic illustration of the various delta technologies.

## 4. How it works?

### 4.1 Block Level Differencing explained...



## 4.2 Byte Level Differencing explained...



### 4.3 Write Monitor Differencing and File Specific Monitoring explained...

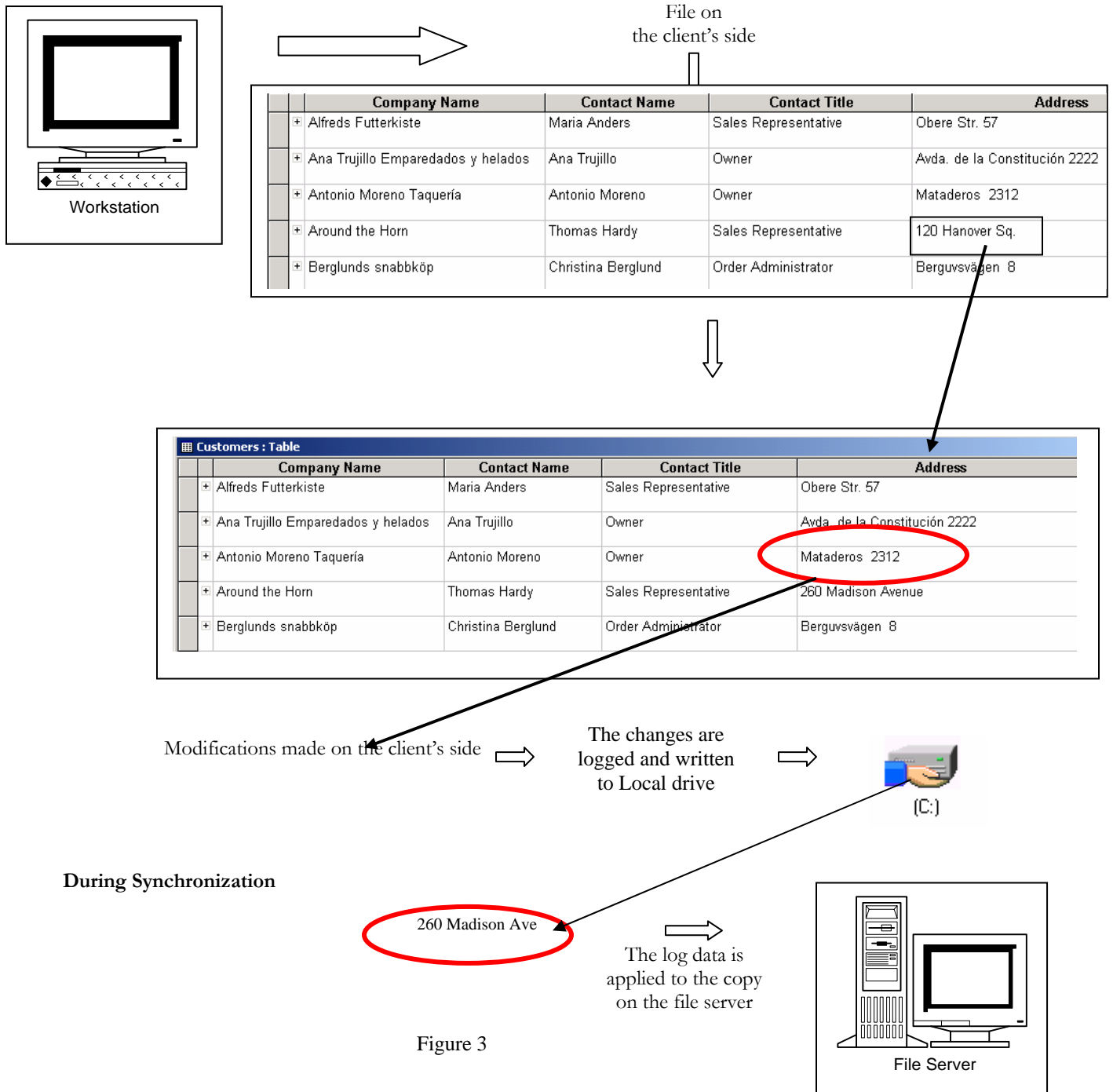


Figure 3

## 5 Comparison

### Comparison of Different iDESP Synchronization Technologies

The following table compares the various iDESP Synchronization Technologies.

	Block Level Differencing	Byte Level Differencing	Write Monitor Differencing	File Specific Differencing	Compression
<b>Computation Cost</b>	Less	More	Least	Least	Least
<b>Delta Size</b>	5%	2%	1-2%	1-2%	50%
<b>Data Reduction</b>	95%	98%	98-99%	98-99%	50%
<b>Process Required on Both sides</b>	Yes	Yes	No	No	Yes
<b>Additional Space</b>	No additional space required (an option of block level differencing can store check sums to further increase efficiency.)	Additional space required on client to store versions of files that are changed	Additional space required on client to store the log records	Additional space required on client to store the log records	No additional space required.
<b>Applicable over Slow connection</b>	Yes	Yes	Yes	Yes	Yes
<b>Applicable over LAN</b>	Yes (if Synchronization Server and File Server are the same)	Yes (if Synchronization Server and File Server are the same)	Yes	Yes	Yes
<b>Applicable File Types</b>	All Files	Typical office documents, like Word, PowerPoint	Outlook, Access	Limited to some types of applications e.g. outlook	All Files
<b>First time synchronization</b>	No	No	No	No	Yes
<b>2 way merging</b>	No	No	No	Yes	No