



Can you **really** recover your data?

A Thought Provoking Look at Backup & Recovery

Presented by
Peter Gale

MD, Comprehensive Solutions International

Various presentations - Sept. & Oct. 2005 (UK & Europe)



Who, What, When

- Peter Gale
 - 22 years IT Experience
 - 14 years Ingres Experience
 - 16 years as an Independent Consultant
 - Now MD of Comprehensive Solutions International
- Comprehensive Solutions
 - IT Consultancy with expertise in
 - Ingres
 - Disaster Recovery
 - Project Management
 - Compliance
 - Operations in the UK and the USA



Data – The life blood of your business

- Computer based data critical to almost ALL businesses.
- Loss of data and down time can cost big money.
- Disaster Recovery and Business Continuity Planning are the stuff of ‘Big Business’?
- Everyone has backup plans and processes (don’t they?)
- Not everyone has **recovery** plans and processes.
- Can you **REALLY** recover **ALL** your data?

- White paper at
www.comp-soln.co.uk/Backing_Up_Your_Business_WP.pdf
– Presentation
www.comp-soln.co.uk/Can_you_recover.pdf



What we will look at

- What is ‘**ALL**’ your data?
- What happens when you lose the data or access to it?
- Defining a Recovery Point Objective (RPO)
- Defining a Recovery Time Objective (RTO)
- Implementing RPO and RTO
- Ingres specifics

ComprehensiveSolutions®
Business Savvy. IT Smart™

Self-Assessment

- Hands up if you have done a recovery test of ALL business critical data in the last
 - 12 months?
 - 6 months?
 - 1 month?
- Why Not?
- Does that worry you?

www.Comp-Soln.co.uk

Copyright ©2005 Comprehensive Solutions International Ltd. All Rights Reserved

ComprehensiveSolutions®
Business Savvy. IT Smart™

What happens when data is lost?

- Airline
 - Lost bookings?
 - Critical flight safety information?
- Amazon.com
 - Lost orders
- NHS
 - Patient records
- Inland Revenue
 - Unpaid benefits
- Your Business
 - ?
- **The effects of lost data can be detrimental to your company and to your job security!**

www.Comp-Soln.co.uk

Copyright ©2005 Comprehensive Solutions International Ltd. All Rights Reserved



Your Data

- What does 1 hour or 1 day of down time mean to your business?
 - Lost revenue?
 - Compensation?
 - Impact on other business/people?
- Can you quantify it? That is an important first step.
- Can you recover ALL your data?
- Do you know what 'ALL your data' is?
- How will you validate that?
- Will 'ALL your data' be up to date when it is recovered?
 - Under all circumstances?
- How long will it take to recover 'ALL your data'?
 - What is the cost to your business of a long delay?



Validated Backups

- You think you have all the backups in place.
- Can you **prove** it?
- Do you know for certain, 100% of the time, that your recovery scheme works as designed?
 - Are the backups valid?
 - Do you have formal documented procedures for recovery?
 - What triggers execution of the plan and by what authority?
 - Is it detailed down to the commands to be typed and the expected output?
 - Are there checkboxes for completed steps & room for notes / time stamp?
 - Are troubleshooting techniques and procedures included?
 - How to validate
 - Do you have the expertise to do the recovery any time of the day or night?
- What is your Guaranteed Recovery Point?
- What is your Guaranteed Recovery Time?



Recovery Point Objective

- The maximum amount of transactional data loss after recovery.
- Or
- How far your data will be rolled back by the process of recovery.
 - Measured as a time period - usually minutes, sometimes hours.
 - RPO 0 = No transaction data loss



Recovery Time Objective

- The maximum amount time taken to recover the data and restore service.
- Usually in hours, sometimes in minutes or days.
- Could be dependent on a combination of events.
 - System Restoration (I.e., Disaster Recovery)
 - Computer system repair or Engineering fixes
 - Software installation & patching
 - Data restore
 - Process re-runs
 - Integrity checks
 - Notifying user community



The 'Dead Money' Argument

- Backup strategies cost money and the benefit is often difficult to quantify.
- Need to understand from the point of view of NOT having the required strategy (e.g., unplanned service outages & data loss).
- Achieving low RPO and RTO can cost significantly more money.
- Cost is justified by knowing the cost of ONLY achieving a certain RPO/RTO.
- For example
 - Strategy costing £5k gives you a 60 minute RPO and a 2 hour RTO.
 - Possible revenue loss £100k
 - Strategy costing £10k per year gives a 10 minute RPO and 30 minute RTO. Possible revenue loss of £5k
 - Use the £10k backups once in 10 years and it pays for itself.



Defining your RPO

'I want an RPO of 10 minutes, but tell me what it would take for an RPO of zero'

- RPO zero is everyone's ideal, but is it realistic?
- This client knew (from bitter experience) that there can be a large price tag on that.
- DBMSs give RPO 0 recovery for loss of the database itself
 - Most single item failures are covered.
- INGRES: Impact of losing JNL, Log file, last checkpoint etc.
- Key to achieving an RPO lies in the 'backups of the backups.'
- Non-database files need to be considered
 - Source code
 - Interface files
 - Documents
 - Customized configuration files, terminal map files, etc.



What affects your required RPO

- The raw cost of data loss
 - What is the business impact of the loss in your environment?
 - For example, in the healthcare industry it could mean scrapping a very expensive medical device.
- Re-processing/re-entry of data
 - Will you know what needs to be re-keyed?
 - Chance for lost orders and duplicate orders
- Indirect costs
 - Loss of confidence by your customers
 - Potential loss of future business
 - Compensation
 - Impact on other businesses / people
 - Impact on things like financial reporting. Is the lost data *material* to operations affecting financial reporting?



Calculating the RPO Cost

- Develop an understanding of the costs
 - Average value of a transaction.
 - Transaction volumes over time and by time of day.
 - Is there a manual data capture (hardcopy backup) system?
 - Cost of manual recovering data.
 - Impact of manual recovery. Will you ever catch up?
 - Transaction value depreciation
 - Does this happen?
 - Does the value ever become 0 or even negative?
 - Other costs specific to your environment or industry?
- Service level agreements (SLAs), customer charters, contractual commitments and obligations.
 - These costs are above and beyond the cost of data loss and downtime!



RPO Example

- Average Tx value - £125
- 500 per hour
- Orders must be processed in 2 hours or cancelled.
- 25 agents. 5 minutes per order

Tx Loss Hours	Txs to re-enter	Hours to re-enter Txs	Txs that will be cancelled	Value of lost orders
00:10	83.3	1.7	0	£ -
00:20	166.7	3.3	67	£ 8,420.14
00:30	250.0	5.0	151	£ 18,880.21
00:40	333.3	6.7	235	£ 29,340.28
00:50	416.7	8.3	318	£ 39,800.35
01:00	500.0	10.0	402	£ 50,260.42
01:10	583.3	11.7	486	£ 60,720.49



Other RPO factors

- Enhanced Business Processes
 - Extra staff to clear backlog
 - What is the cost?
 - Process orders manually and re-enter later
 - Re-run interface processing
 - Multi-stage recovery
- Limit to what can be done through repeat work and increased resources.
- Basic recovery capability is paramount!



Defining your RTO

- RTO is the acceptable 'down time' until complete recovery.
- Recovery often involves many stages based on various dependencies.
 - Hardware recovery.
 - Data/software recovery.
 - People Recovery.
- Warm/Hot standby solutions provide short recovery times but at a high price.
 - Choice is based on the RTO
- Need to determine what can be achieved with your existing strategy.
- Is that acceptable to your business? Document this for future reference either way.



Defining your RTO

- Two things are established from defining the RTO and associated backup strategy
 - The cost to the business of the down time.
 - The recovery time itself.
- Focus is on the cost of downtime and lost business
 - Transaction value & volumes.
 - Is there a manual system that can be temporarily implemented?
 - What is the impact on volume?
 - What is the impact on timeliness and service to your customers?
 - What is the cost of entering those transactions later?
 - Additional costs.
 - Regulatory costs.
 - Contractual costs.
 - Customer impact.
 - What is the business impact of delays in your service for your customers?

ComprehensiveSolutions®
Business Savvy, IT Smart™

RTO Example

- Same order system
- Manual transaction handling during down time
- Transaction entered in spare time when system is back up

Down Time	Txs to re-enter	Hours to re-enter Txs	Txs that will be cancelled	Value of lost orders
00:10	83.3	1.7	0	£ -
00:20	166.7	3.3	67	£ 8,420.14
00:40	333.3	6.7	235	£ 29,340.28
01:00	500.0	10.0	402	£ 50,260.42
01:30	750.0	15.0	653	£ 81,640.63
02:00	1000.0	20.0	904	£ 113,020.83

- RTO will likely be bigger than RPO
- Much higher cost implications!

www.Comp-Soln.co.uk

Copyright ©2005 Comprehensive Solutions International Ltd. All Rights Reserved

ComprehensiveSolutions®
Business Savvy, IT Smart™

RTO/RPO Combined

RPO/RTO Value Matrix		Tx Loss Hours			
		00:10	00:20	00:30	00:40
Down Time	00:10	£ -	£ 8,420.14	£ 18,880.21	£ 29,340.28
	00:20	£ 8,420.14	£ 16,840.28	£ 27,300.35	£ 37,760.42
	00:40	£ 29,340.28	£ 37,760.42	£ 48,220.49	£ 58,680.56
	01:00	£ 50,260.42	£ 58,680.56	£ 69,140.62	£ 79,600.69
	01:30	£ 81,640.63	£ 90,060.76	£ 100,520.83	£ 110,980.90
	02:00	£ 113,020.83	£ 121,440.97	£ 131,901.04	£ 142,361.11

www.Comp-Soln.co.uk

Copyright ©2005 Comprehensive Solutions International Ltd. All Rights Reserved



The Backup Strategy

- Three way balancing act
 1. Cost of an ideal solution.
 2. Cost of not having an adequate solution.
 3. Likelihood of a failure.
- Typical Strategy
 - Hardware redundancy
 - Good practice but not 100% guaranteed.
 - Cannot protect against non-hardware errors.
 - Database backup – Daily.
 - Backup of database backup – Daily.
 - File System backups – Full (weekly, monthly), Incremental (daily, weekly).
 - Usually viewed as adequate.
- Day time backups are rare, but should they be?



Identifying ALL the Data

- Vital to identify ALL the data and the required backup frequencies.
 - Where is my data?
 - How do I recover it in the event of a media failure?
 - Can the data be reconstructed from other data sources?
- Start a 'Data Inventory'
 - Simple list of data repositories (ALL of them).
 - Database(s).
 - Interface files.
 - Application source code and executables.
 - User file systems.
 - Log files (auditing/compliance?)
 - More...



Recovery Dependencies

- Extend Data Inventory to include items needed for recovery
 - Database.
 - Directory Structure (Location paths).
 - Database Backup (Checkpoint & Dump files).
 - Transaction data (Journal files).
 - Configuration data (Database Config file).
 - Potentially much, much more...
 - Application Source code
 - File system backup?
 - Changes since last backup?
- Each dependency becomes a data item in the inventory
 - Not obvious 'data' items.
 - Do these things exist? (e.g. record of source code changes)
 - Are they backed up and how frequently?



What if..?

- With the basic inventory complete we can begin the 'What if?' questions.
- 'What if I cannot access the current transaction data?'
 - Additional backup requirement identified (Ingres journals).
- 'What if I cannot recover my Database from the last tape Backup?'
 - Identifies the requirement to keep multiple tape backups.
- 'What if the application source code is lost?'
 - More frequent backups?
- 'What if the scope of the failure is more severe?'
 - Is your backup & recovery strategy integrated with your Disaster Recovery / Business Continuity strategy?

Data Inventory

- Extend Data inventory to describe impact of backup frequencies/types etc on recovery.
 - More frequent may reduce recovery time and data loss.
- Testing required to ascertain recovery times
- Iterate until required levels of backup are obtained
 - Typically 2 layers for everything (redundancy is important)
 - Example in the white paper (www.comp-soln.co.uk/whitepapers)

Data Inventory										
Data Item	Recovery Dependency	BU	RPO	RTO	RPO Backup Frequency	RPO Notes	RTO Backup Frequency	RTO Notes	Backup Method	Notes
mydatabase		Y	10	30	N/A	Any valid ckpt can be used if all journals are available	Daily	15 mins to recover as of 27/07/05	Ckpt.ksh	
Data Locations	Y									
Checkpoint Files	Y									
Dump Files	Y									
Journal Files	Y									
Database CNF	Y									
Customized ckimpt.def	Y									
Data Location		Y	0	1	RT		RT		Documentation	
None										
Checkpoint Files		Y	N/A	15	N/A	Static Files	Daily	Immediately after Checkpoint or Dual Checkpoint	SysBackup.ksh or Ckpt.ksh	
Checkpoint locations	Y									
Checkpoint backup										
Dump Files		Y	N/A	15	N/A	Static Files	Daily	Immediately after Checkpoint	SysBackup.ksh or Ckpt.ksh	
Dump Location	Y									
Dump file backup										
Journal Files		Y	10	15	9		As per RPO		JournalArchive.ksh	
Journal Location	Y									
Journal file backup	O									
Database CNF		Y	10	1	9		As per RPO		Automatic	
Root Data Location										
CNF Backup (DMP Location)										
CNF Backup (DMP Location)		Y	10	1	9	Backup when changed			TBA	
CNF Backup 2										



Achieving the RTO

- Factors affecting RTO
 - **Size:** More data takes longer to backup and recover.
 - **Physical access to backups:** May be offsite.
 - **Frequency of backups:** Reduce the amount of transaction data to re-apply and/or processes to re-run.
 - **Validation:** Integrity checks.
- Exploit existing backup/recovery software
 - Faster hardware (backup to disc, faster tapes, etc.)
 - Parallel processing.
 - Configuration.
- Invest in new technology
 - Disc mirror splitting (e.g., EMC's BCV system).
- Allow for the worst case scenarios
 - Failure just before or during the backups.
 - Out of hours failure.



Achieving the RPO

- RPO zero
 - Backup all changes when they occur
- RPO n
 - Backup changes more frequently than n .
- Capture 'change' information wherever it is needed
 - Transaction data.
 - Code changes.
 - Interface files.
 - Anything else important to your environment.
- Monitoring processes required
 - Backup more frequently than the RPO.
 - Or when changes occur. (complex)
 - Ensure that they complete successfully within the desired time window.



Ingres specifics

- Database
 - Checkpoint Frequency does not affect RPO.
 - Backup frequently to reduce Journal recovery time.
 - Utilise high speed devices and/or parallel checkpoints.
- Checkpoints
 - Must have more than 1 valid checkpoint and all its journals.
- Dump Files
 - Backup with the checkpoint.
- Journal Files
 - Backup at RPO-1 frequency.
- Database Config File (in the dump directory)
 - Backup at RPO-1 frequency.
- Transaction Log File
 - Regular consistency points.



Ingres specifics

- RPO Factors
 - Archiver Interval must be less than RPO
 - cp_interval may not be good enough with large log file. Use cp_timer
 - r3. Possibly reduce archive_refresh and/or cp_interval_mb
 - Possibly force consistency points (dm1305) and Archive cycles (dm1314)
 - Journals must be copied when consistent, i.e., Archiver is inactive
 - Monitor iiacp.log
 - Check journal size before and after copy
 - Or check for Archive Cycles during the copy
 - Consider LOCAL (fast) and REMOTE (slow) archives
 - CCSI Tools: JournalArchive.ksh
 - Monitor journal archive process to ensure desired frequency.
 - Config file must be backed up (from Dump location) every time it is backed up by Ingres.



Ingres specifics

- RTO Factors
 - Reduce recovery time using hardware.
 - Increase the number of locations.
 - Disc bandwidth may allow multiple locations on the same logical file system. (Use different Paths for each location)
 - /ingdata1 -> ingdata1
 - /ingdata1/loc2 -> ingdata1_2
 - Unlikely to affect day-to-day performance.
 - Daytime Checkpoints
 - Ensure application is not doing any DDL (Create, modify etc.)



Backup Validation

- All of the above is useless if you don't regularly prove you can recover as required
- Avoid the 'False Sense of Security Syndrome'
 - *'We have all the necessary backups in place, we use trusted and reliable software, so we are OK'.*
- Schedule regular recovery tests
 - Changes in size, transaction volume, many factors might affect recovery.
 - Gather timing metrics and track, looking for trends.
- Test recovery after any changes
 - Configuration.
 - New application release.
- Validate the recovery
 - Integrity checks.
 - Completeness.
- Look for differences and/or anything unusual - that is generally bad.



In Summary

- Backup Plans are of limited to no use if they don't match the recovery requirements!
- Backups must be multi-level to ensure reliability.
 - Backups of the backups
 - Backups of the changes
- The cost implications of your RPO and RTO should be known
- Your likely recovery times should be known
- Compile a 'Data Inventory'
- Validate your solution regularly
 - Even if you stick with the one you have now.
- Make certain that everything is documented at a very detailed level. Remember, people are points of failure as well!



Q & A

