Transaction Processing System

Characteristics of Transaction Processing Systems

TRANSACTION

- **Series of events** (a request, an acknowledgement, an action and an outcome) **important to organisation**
- Either **complete success** or **complete failure** (leading to rollback)
- Ensure events occur successfully – events send response to **show success or failure** to determine whether to **commit or rollback**. The events then **send acknowledgement** of performing request

COMPONENTS OF A TRANSACTION PROCESSING SYSTEM

- ***Purpose*** – **info system** that **collects, stores, modifies & retrieves transactions** (info process) of organisations
    - Must have **rapid response, reliability, inflexibility, controlled processing** (e.g. TPM)
- ***Data (aka master file)*** – Stored in **DBs** & transformed into info through **info processes**. Possible **integrity issue** resolved by transaction log. **Additional data** used to record details of each transaction
    - **Older systems** – Details of transaction recoded in **transaction file**. **App controlled transaction file and master file**. Changes to the master file made when transactions committed.
    - **Newer systems** – Same but **DBMS** or **TPM** used instead of app
    - **Transaction log** – used to **commit** or **rollback**
        i. Added records or changes recorded in log. When committed, **records in log replace/add to master file**. If rolling back, records in log aren't written.
        ii. **Record original data in log & change master file**. If committed, nothing. If rolling back, original record pasted over master file. Also, entry made in log to specify to delete record.
    - Most DBMS **record before & after ver. of data** in logs to allow
        - **Backup** with use of **logs to commit** until present &/or **rollback** incomplete transactions
        - **Audits –** shows when, what & who performed each transaction
- ***Information technology***
    - **Hardware** – **Server machines** (& redundant components), **storage & communication devices, transmission media, backup devices**, **client workstations** (client apps), **collection devices**
    - **Software** – **DBMS software** (& log), **client apps** (interface for participants in client-server network), **proprietary software apps, transaction processing monitors**
        - **TPM** – Software apps to **coordinate processing of large TPSs (**may have many DB servers)
            - **Connect** to systems operated by other org.
            - **Integrity** of transactions using different servers/systems, **committing, rolling back**
            - **Balance load** of transactions sent to each server
- ***Processes*** – Mainly Storage & retrieval, but also collecting, analysing, processing
- ***Participants*** – People, who operate TPS, enter data to initiates transactions, in enviro for online real time.

BATCH TRANSACTION PROCESSING

- **Collection & storage** of data for processing at a **scheduled time** (system is quiet) or when **sufficient data**
- **Separate from TP operations &user interactions**
- Performed in **parallel with other processes**, thus **ACID important** (esp. isolation)
- Transactions stored (usually on **magnetic tape/ paper**) until convenient or economical to process them
- **Advantages** – run on **regular schedule** with **no operator** need, **cheaper** than real time (esp. I.T. needed), **no wait time** as data usually not needed by other processes
- **Disadvantages** – If error occurs, whole system halted & restarted, exclusive access to data & no sharing,
-

REAL TIME TRANSACTION PROCESSING/ONLINE TRANSACTION PROCESSING (OLAP)

- **Complete transactions immediately after they've been initiated** – feedback if needed
- Can't control **transmission speeds**, only **data access & processing speeds** by hardware & software
- **Online users** – **employees** of org., **customers** entering details via web or other networks
- **Advantages** – reflects current situation, no time lapse
- **Disadvantages** – Expensive (compared to batch), requires human operator

COMPARE AND CONTRAST BATCH AND REAL TIME TRANSACTION PROCESSING

Real Time Processing	**Batch Processing**
Collects, stores and **modifies records** of transaction	
Has **data integrity procedures** (validation, verification, referential integrity and ACID properties)	
Transactions are **handled individually**	Transactions are **handled as a group**
Processing and **output is immediate**	Processing and **output is delayed**
System failure critical as prevents processing taking place	**System failure less critical** as possible time to restore system before next batch
If event fails, opportunity to **correct data immediately**.	Errors stored on an **error file** & corrected at a **later time.**
Fewer errors as **data validation** present	**Data validation** afterwards

DATA VALIDATION IN TRANSACTION PROCESSING

- ***Data Integrity*** – A measure of how **relevant, correct** and **accurately** data reflects its source
  - **Data validation** – At **time of collection**, data is **reasonable** or **correct** in terms of enterprise's business rules. Performed by **client app**. 1 undetected data error can affect numerous transactions
  - **Data verification** – Data collected **matches & continues to match source**. Simplified if **unique identifier** available across all systems, but raises **privacy concerns**.
  - **Referential integrity** – All **FK match PK in linked table** otherwise **orphaned records**. Data validation & verification issues can affect referential integrity.
  - **Resource manager** – keeps track of all data in transaction & allows terminal access for different data
  - **ACID Properties** – In real time TP, if **ACID test is passed**, then **data integrity is achieved**
    - Atomicity – Either **completely successful** (commit, permanent data changes) or **none at all** (roll back to original state)
    - Consistency – Transaction uses data from **1 consistent state** & **after, data left in a consistent state.** All data accounted & **each step of transaction carried in same way** so data is correct.
    - Isolation – Process data **without interfering** with or **influenced by** other processes currently executing. As data is in **inconsistent state** during processing, **record locking, logs** & **"two-phase commit"** needed, or erroneous data may be processed.
      - **Record lock** – **Prevents other operations from accessing the data**. Other processes are aware of record lock & will wait until released.
      - **Two-phase commit** – 1st phase **recorded in log** & **locks record**. 2nd **alters actual data** permanently & **releases lock**.
    - Durability – Ensures **committed transactions permanent**. In real systems, requires all **changes written to permanent or secondary storage** before transaction is truly committed.

HISTORICAL SIGNIFICANCE OF TRANSACTION PROCESSING AS THE FIRST TYPE OF INFORMATION SYSTEMS

- ***1950s*** – **clerks for manual processes** in TPS **replaced by computers**
- ***1980s*** – **TP apps** developed (often in Cobol) for each individual organisation was **replaced by DBMS**
- ***Today*** – most TPS are based on **1+ RDBMS with TPM & client apps** to meet organisation's specific needs
- ***Vital to the development*** – Cobol (language) → CICS (TPM) → SQL → RDBMS (oracle) → Microsoft

SUITABILITY FOR AUTOMATION BASED ON CLERKS

- **Strict sequence of events** – each event acknowledged to start next ones & roll back if failure of any events
- **Large organisations** – common for 1 clerk to perform one event of all transaction, then pass to next clerk
- **Strict sequence and rules** of such transactions make them well suited to automation

Types of Transaction Processing Systems

WEB-BASED – Reservation system, internet banking

NON-WEB-BASED – Point of Sale (POS) systems, Library loans system

ON-LINE REAL TIME

BATCH – Payroll, cheque clearance, bill generation

SYSTEMS THAT APPEAR REAL TIME, RESPONDING AS TRANSACTIONS OCCUR, BUT REALLY BATCH PROCESSING, SUCH AS CREDIT CARD TRANSACTIONS – Customer, merchant, card issuer, acquirer

Storing and Retrieving in Transaction Processing Systems

STORAGE OF DIGITAL DATA IN DATABASES AND FILES

- TPS requires efficient method of **storage & retrieval of data**
- Data normally stored in **DB or data warehouse** & needs well designed **backup & recovery** procedures.

RETRIEVAL OF STORED DATA FOR FURTHER TRANSACTION PROCESSING SUCH AS PRINTING INVOICES

SYSTEMS TO STORE PAPER RECORDS OF TRANSACTIONS – document **imaging** (image file, searching/editing unavailable, save storage space), **OCR** (image file processed to have characters recognised saved in word doc file)

DATA BACKUP AND RECOVERY

- **Snapshot copy** of data at point in time & allow system to **recover** back to state it was backed up.
- Commonly **hard disk failure** (esp. read/write heads) **– <1 year** (manufacturing fault) & **>6 years**
- Recovery is valuable – allow recovering from broadest range of possible problems.
- Frequency of backup & length data is kept depends on **value & nature of data**
- ***Grandfather, father & son*** – Most commonly used rotation scheme involving **3 generations**
  - **Daily** (son, full/partial), **weekly** (father, full, onsite), **monthly** (grandfather, full, offsite)
  - Usually final yearly backup archived permanently.
  - Either **replace tape** regularly or **promote son to father**, etc.
  - **Detail of procedure** determined by needs of organisations & can extend to **great grandfather**
- ***Off-site storage*** – storing backup **separate from main system**. E.g. magnetic disks at employee's house.
- ***Secure on-site storage*** – Backup location **close to system** – **faster access** but **riskier** & need extra precaution.
- ***Full and partial backups***
  - **Full backup** – Complete copy (**OS, program files, configuration settings, data), data** for TPS regularly. Impractical to perform regularly as takes **long time & large amounts of storage**. **Archive bit to false**.
  - **Incremental Backup** – Backup files with **true archive bit. Faster** as less storage. Before incremental, full backup needed. Fast saving time countered by **time needed to recover files**. Made **at least daily**.
  - **Differential Backup** – Backup **all files changed since last full backup**. **Doesn't alter archive bit** after backup, this size of **backup grows** until next full.
- ***Recovery testing*** – Testing backup by **restoring files into temporary directory** to ensure backup works
  - **Backward recovery** – undoes unwanted changes
  - **Forward recovery** – Backup to recover as recently & logs to recreate loss data up until system failure

- ***Suitable media*** – backup system **compress data** prior writing to **double capacity**
    - **Magnetic tape** – **Light, portable** & **no complex electronics**, **long term & off-site**. **Robotic system** to automate tape libraries. Improves **fault tolerance** as drivers replaced without affecting system. Includes **helical & linear**
    - **Hard disks** – **Externally connect** via **USB/firewire ports** or **Ethernet networks**. Very **expensive**, with weight & **mechanical complexity** significant when used for **off-site storage**
    - **Optical media** – **multiple DVD-RWs for full** (with data compression), **1 for partial**. Not viable for backup to tape in terms of capacity, quantity & moving disks in & out of drives.
        - **Recordable WORM** – **archive critical data** in unalterable form for long periods of time
    - **Online** – **Automated backup** process for individuals & small businesses by managing **secure storage at secure remote site. Large organisations** maintain complete operational copies of entire system at remote locations via online communication lines to remote site.
- ***Specialised backup software*** –Configured to **enforce backup procedure** (+ verification). Backup of **selected files** (based on date/size/type), **optimal compression, encrypt** backup, **schedule** backup.
- ***Transaction logs*** – **Historical details** of transactions made. **Restore TPS** back to **consistent state** at anytime
    - In recovery, **recommit** completed transaction or incomplete transactions **rolled back**
- ***Documenting backup and recovery procedures*** – Various **backup media** used to **maintain many backup copies made at different times** so system's data can recover to different past states. Personnel responsible must document **when, what & how** of backup & **storage media** used, **verification & storage location**.
- ***Mirroring*** – Writing data to two drives simultaneously (**100% redundancy**) & allows for fast recovery (**hot swapping**). **Fault tolerance**, but doesn't protect from total system failure
- ***Rollback*** – Log used by system to **automatically rollback incomplete transactions**, returning data to most recent consistent state. Generally action **aborted** after certain period of time.
    - **Server (initiating transaction) crashes** – **message to other involved systems to abort**
    - **Server (contributing to transaction)** – **inform initiating system** who rolls back complete transaction
    - **Roll forward** – Recover (consistent data) & recommit transactions made since last update using log

UPDATING IN BATCH SYSTEMS

- ***Historical significance*** –**First type of TP** via punch cards manually punched & processed sequentially by operators. **Real time not** possible as hardware only performs **1 task at a time** & **output stored sequentially**. Made **max use & efficiency of limited resources**.
- ***Limitations of batch processing*** – Performed at **evenings or weekends** when OLAP requirements lowest – **data not current** (master file updated periodically). Time consuming due to **sequential storage**. **User interaction limited** & error log generated after processing.
- ***Technology required*** – large **storage** capacity (sequential – tape), simple **software & user interface**
- ***Steps in a batch update*** – Collected transaction data added to **transaction file**. When **convenient/ sufficient amount,** file submitted for **processing** (similar to RT) & successful transactions **written on master file**
- ***Suitable applications*** – Paper-based collection, off-line systems, bulk generation

UPDATING IN ON-LINE REAL TIME SYSTEMS

- ***Relevance and impact*** – **immediate update** of data (current), **record locking** required, immediate **feedback** if error, **multiple users** simultaneously process transactions
- ***Technology required*** – **Fast communication** links, fast **processor**, **direct access** storage devices, **client-server system**, **user friendly** interface, **input hardware** (barcode scanner, ATMs, EGTPOS terminals, PCs), **software** to enable simultaneous transactions, real time **validation**
- ***Steps in on-line real time processing*** – Each transaction processed as it is initiated. **Add transaction to transaction file**. Read key field from file record. **Locate same record** in master file & **update**
- ***Suitable applications*** **–high demand for TP** as hardware & software expensive, **on-line** & **web-based** apps

Other Information Processes in Transaction Processing Systems

- ***Collecting in transaction processing – Hardware***
    - Automatic Teller Machines (ATMs) – **Card reader** reads account info on **magnetic card** (identify user) via **magnetic stripe readers** (swiped or inserted (less error & increase security). **Keypad & touchscreen** to enter data, **screen & printer** for output.
        - ATM **identifies user** & **financial institution** via card. If user has account with owner of ATM, **amount debited**. If not, **amount moved** into account of financial institution.
    - Barcode readers – Decoder (no dedicated interface software) to organise data into character representation, then matched with index ID no. in DB.
        - **LED & laser (star)** – **light reflected** off barcode detected by a **photocell**, which state changes based on **intensity level** & affects **current flowing**. Current changed into **binary via ADC**.
        - **CCD technologies** – **Row of LEDs & photocells** used. Light reflected off image to **mirror,** into **lens** to **focus image onto CCD**. Then same as LED & laser.
    - Radio Frequency Identification (RFID) tags – RFID tags (96b) **uniquely identify resources**. **RFID readers transmit frequency** specific to antenna of **passive RFID tags** to power it. Tag responds by **transmitting stored ID** back to reader. Operates at **10 – 100m**. ID used in conjunction with DB, to check movement of goods.
- ***Collection from forms*** – (user interface) **collect data required for transactions**: **paper-based** (indirect users manually complete, data entered, batch processed) or **web-based** (removes data entry operator)
    - **Web-based** – screens part of **front-end client apps** that connect via **LAN to backend DBMS** or can be web-based clients where **data travels over internet** , then via **web server** to **storage in DB**
    - **General form design principles** – know who **users** are; **nature of data items** collected; consistency; **readability** (fonts, capitalisation; graphics & colours used sparingly); **white spacing** for grouping; preferably **left justified** (to lead users in desires input sequence)
    - **Principles for design of paper forms** – Paper form & data entry screen structured to **assist data entry** & **manual completion**; clear **instructions & positioning** as unable to react to user input; **design** (colour, texture, fonts &paper itself) for majority users; **appropriate space** for answers.
- ***Screen design for on-line data collection*** – show **available functions** (dull not hidden); **feedback**; **undo** feature, warning messages; adhere to **OS standards** for user interface design
- ***Web forms for transaction processing (real time and batch)*** – **validate data in DL page** if possible (reduce data transferred. Feedback); **design** so **transmission of data** needed for **validation** occurs prior next form ("submit"); **software tech** works properly; **security** of personal info (encryption, SSL)
- ***Analysing data, in which output from transaction processing is input to different types of info systems***
    - **DSS** – Provides **solutions & consequences** to managers to assist **decision making** (e.g. OLAP)
        - Improve future performance via predictions based on historical data (**data warehouse**)
        - **Data mining, querying data marts** (improve efficiency of info extraction) are DS technique
    - **Management information systems** – **Data** in TPS **into info** to **assist managing business operations**
        - **Statistical analysis of existing data** – sales reports, profit & loss statements, sales trends
        - **Small** (info generated by manager, MIS in TPS) or **large** (1+ dep. For MIS, separate system(s))
        - Need **computer skills** & **grasp of business processes** to make data into relevant info
    - **Data warehousing systems (for data mining)** – ***Pg. 7 IS&DBs***. Advantages:
        - **Old data purged from system** – improve performance as less data to read when processing
        - **Centralise & store data** in **similar format** even if changes – simplifies analysing in a place
        - **Snapshot of real data**, so analysis proceed efficiently – no **record locks**, **ACID, data integrity** issues, effect on performance of operating system
    - **Enterprise systems** – Any system that performs **processes central to overall operation** of enterprise
        - **Hardware, software apps & data** critical to its functioning

Issues related to Transaction Processing Systems

CHANGING NATURE OF WORK AND THE EFFECT ON PARTICIPANTS, INCLUDING – Early 1970s, thought new tech meant **reduction in hours**, but instead introduced **new industries & employment, longer, stressful hours**

- ***The automation of jobs once performed by clerks*** – Automation is use of **IT to do tasks once performed by humans.** Either be **retrained** or **taken over** by younger, more skilled workers. E.g. less retail assistant or shop managers due to businesses web-based.
- ***Shifting of workload from clerks to members of public*** – customers enter data for **online banking**, retailers enter data into **EFTPOS terminal**
- ***Assess the impact on participants involved in TP*** – Either **positive (right support given)** or **negative**. New **career opportunities & applications** developed due to computer-based systems.

THE NEED FOR ALTERNATIVE PROCEDURES TO DEAL WITH TRANSACTIONS WHEN TPS IS NOT AVAILABLE

- Large centralised systems maintain **backup power generators** & **redundant communication lines**
- Staff trained, non-computer procedures trialled & tested at regular intervals to ensure operation to continue
- **Paper-based** forms to record transactions (e.g. EFTPOS)

BIAS IN DATA COLLECTION – false emphasis or representation, leading to inaccurate info

- ***When establishing the system and deciding what data to collect*** – Bias from **incomplete data** collected in aim to simplify system. Internet users not representative as **higher tech skills, incomes & younger**
- ***When collecting data*** – Identifying **suitable source of data** (cheapest/easiest data source than best choice)

THE IMPORTANCE OF DATA IN TRANSACTION PROCESSING, INCLUDING

- ***Data security*** – passwords, backup, physical barriers, anti-virus, firewalls, data encryption, audit trails
- ***Data integrity*** – data validation & verification, referential integrity, ACID, minimising data redundancy
- ***Data quality*** – How **reliable & effective data is to organisation**. Issues when survey not completed truthfully or trying to combine data from different systems. Effectiveness of info in data mining & OLAP reduced
    - **Data Quality Assurance** (DQA) standardises definition of data & includes processes that "cleanse" existing data so it meets these data quality standards

CONTROL IN TP AND THE IMPLICATIONS IT HAS FOR PARTICIPANTS IN THE SYSTEM

- **Monitoring employee's work routine** via recording which participants entered which data, how long workers away from desk. **Management has right** to monitor, but **what level is reasonable**? Excessive rules may lead to **conflict & loss of productivity.**
- **Current management theory** – higher levels of productivity achieved when participants motivated by being given responsibility for tasks (less likely to deviate)
- Repetitive tasks→motivation loss→ non-work activities→authoritative control→reduce motivation→ downward trend

CURRENT AND EMERGING TRENDS IN TRANSACTION PROCESSING

- ***Data warehousing and data mining – see IS&DBs Pg. 7***
- ***Online Analytical Processing (OLAP) and Online Transaction Processing (OLTP) – see IS&DBs Pg. 7***