Applying SAN to OLTP Workloads , Web Transactional Workloads , Data Warehouse Workloads
Case Study : The import auto Industry

Name	USN
Pratiksha Hegde	2GI19CS097
Priyanka R Retrekar	2GI19CS104
Sakshi Bagewadi	2GI19CS128
Nikhil Channavar	2GI19CS082

Applying SAN to OLTP Workloads

- OLTP Workloads : Multiple users making a set of consistent transactions.
- These transactions may be simple/complex but consistently range according to 80/20 rule.
- Taking an example of banking application:
    - Most of the transactions are deposit transactions.
    - the simple transactions I/O content is approximately 80 bytes of customer and account information.
    - each transaction generates an I/O operation to complete the transaction process.
- The I/O workload is characterized by many small transactions where each will require an I/O operation.
- SANs are excellent choices for OLTP workloads because they can provide more data paths than any other storage model.
- SANs also provide the additional benefit of data partitioning flexibility.
- SANs offer an environment that enhances a configurations ability.

The data organisational model

- Deals with the database type and its characteristics, as well as the byte transfer requirements.
- RDBMS provides OLTP workloads with a more defined set of processing metrics that enhances your ability to estimate I/O behavior and utilization.
- The use of the relational database has become so accepted and widespread that its macro behavior is very predictable.
- OLTP workloads using an RDBMS can be supported through RAID arrays using level 5 configurations.
- This provides both redundancy and fault tolerance needs.
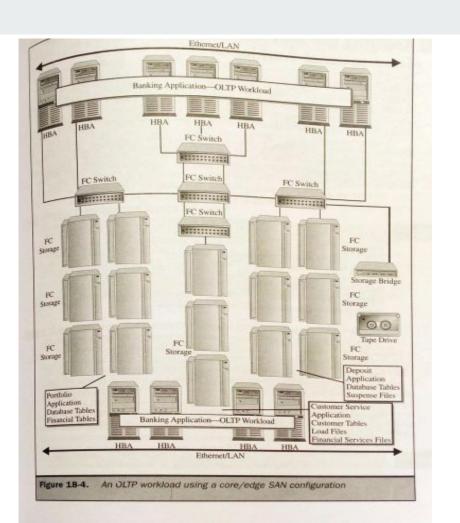
User Access

- User traffic plays a significant role in defining the number of data paths necessary for the OLTP transactions to be processed within the service level.
- We need three types of information from the end users : the number of transactions, the time period within which the transactions will be executed, and the expected service level.
- The deposit transaction used in our example banking application comes in three forms: simple, authorized, and complex.
- The configuration must support a high availability uptime, usually expressed in terms of the percentage of downtime for a period.
- 99 percent uptime requires the configuration to only be down 1 percent of the time.

Data Paths

- By comparing the data organizational model with concurrent factor , we can begin to formulate the number of data paths necessary to meet workload service levels.
- The concurrent factor provides us with a minimal and logical set of paths to sustain our service level, given the probability of all tellers executing deposit transactions at the same time.
- This estimate provides a more accurate picture of the amount of resources needed to sustain the service level in real time.
- The probability of all tellers executing a deposit transaction is actually quite high during peak hours, and could be calculated at 90 percent. Therefore, for each time period, 90 percent of the tellers would be executing a deposit transaction.
- mix of simple, authorized, and complex deposit transactions to be 80, 15, and 5 percent, respectively.

The Design and Configuration of OLTP-Based Workloads

- This configuration is comprised of four FC switches, 15 disk arrays, intersystem -link ports, and an integrated FC-SCSI bridge into a tape library.
- It assumes a relational database that is capable of partitioning among the storage arrays and leveraging a RAID 5 level protection scheme within each array.

Figure 18-4. An OLTP workload using a core/edge SAN configuration

Applying SAN to Web Transactional Workloads

Applying SAN to Web Transactional Workloads

- Workloads generated from Internet or intranet sources can be very complex.
- Based upon the processing architectures of web server software, which enable interactive web services, many of the workloads will be almost entirely transactional in nature.
- However, within this transactional model, another type of transaction type will emerge – the messaging transaction.

Messaging Workloads

- Though messaging transactions are processed in real time, they will have a different set of processing characteristics than typical OLTP transactions.
- Messaging provides an asynchronous form of processing where the client application who has submitted the messaging transaction relinquishes any synchronous connection with the corresponding application.
- The most familiar of these is e-mail, although many other Internet transactions are message-based.

The Data Organizational Model

- WIT workloads simply use the OS file system as the data organizational method.
- This provides a straightforward view when determining byte transfer rates for WIT I/Os given the format of web file structures.
- However, this must be augmented when considering the amount and type of delayed I/Os generated from messaging transactions.

User Access

- Defining user traffic is the most challenging activity in estimating WIT workloads.
- In web-based processing environments, the classic time periods do not apply.
- This is especially true in retail-oriented web sites that rely on customer interaction for sales and which are open 24/7.

Data paths

The number of data paths, as with any workload, is key to the performance of WIT(Web Internet transactional) workloads. It's important to keep in mind that the probability of multiple data paths may have little or no effect if client requests are queued at the file system.
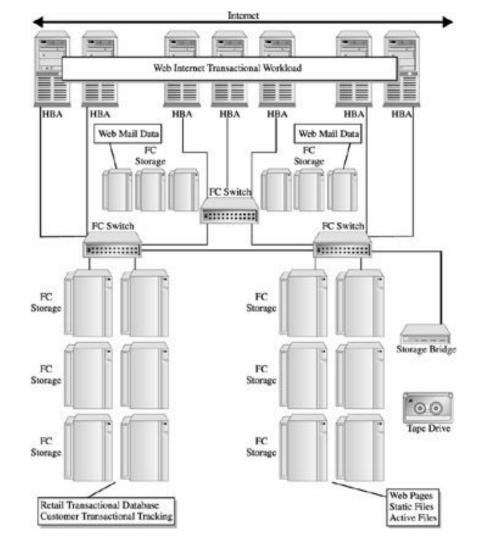
Consequently, there may be a need to consider the following:

- The ratio of channels per server (for example, the number of HBAs per server)
- The need for SMP-enabled servers (meaning more concurrent processing)
- Ensure workload balancing can be used in some form (for instance, in a workload balance software package, processing affinity options, or process prioritization)

The Design and Configuration of Web-Based Workloads

Three FC Switches, **18** disk Arrays, intersystem -link ports, and an integrated FC-SCSI bridge.

This supports both transactional through the meshed configuration.

Applying the SAN to Data Warehouse Workloads

- Data warehouses provide strategic and tactical business information used to direct business planning and forecasting analysis.
- Data warehouses are based upon corporate information and are used to drive corporate business decisions.
- Many data warehouse configurations are moving into the multiple terabyte range and, consequently, provide the foundation for why data warehouses and data marts are good candidates for SANs.
- Data warehouses are built around relational databases; therefore, they are under the control of the users database system of choice.
- DW transactions are serviced on a synchronous and asynchronous basis.
- SANs provide the most comprehensive solution for data warehousing since the MPP database machines. SANs, through their increased addressing capacity, provide an extensible and scalable solution to configuring a very large database.

- **Data Organizational Model**

  The use of a relational database system as the data organizational model for the data warehouse is guaranteed . Working closely with a database administrator greatly enhances the SAN design and operational factors.Database size determines the number of storage devices necessary for pure capacity. DW sources are operational databases that handle a company's day-to-day operation.
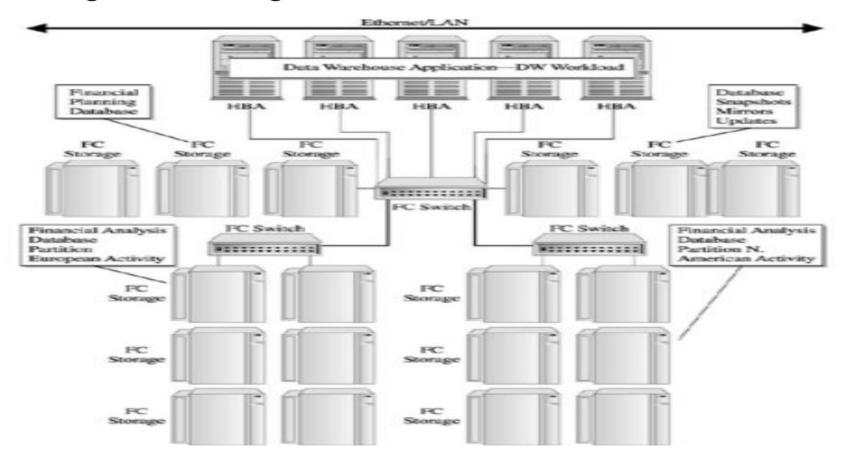
- **User Access**

  The one break you can expect from DW configurations is that the user community supported is generally limited. Although user traffic may be limited, keep in mind that the I/O content of answer sets can be very large.

- **Data Paths**

  The necessity to physically partition the database across storage arrays requires more data paths. Although the number of data paths needed is dependent on the service level, the ability to access multiple terabytes of data using complex queries in a reasonable time period will drive the number of data paths required by a SAN.

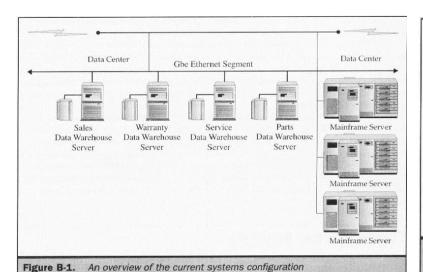# The Design and Configuration of Data Warehouse Workloads

Case Study : The import auto Industry.

- The import auto industry provides distribution services for automobiles **manufactured** in both Europe and Asia.
- The activities that surround the distribution of products within the U.S have provided a rich **datacentric** set of activities that has prompted **leading-edge data** centers in the US.
- Because the automakers` **IT organizations'** data centers have had to address data-processing activities on an international level.
- Key among these activities is the sale of parts **products**, **service contracts**, and the **increasing amount** of technical support provided to dealerships.
    - **For example**, each time a warranty claim is presented and processed within a dealership, the manufacturer pays for the service, parts, and support.
- This case study discusses the **activities** surrounding the analysis, justification, and implementation of a storage area network (SAN)
- However, the major applications supporting these areas **lagged** the rest of IT, with largely centralized systems without the necessary distribution of data from the mainframes to **enhance** or synchronize.
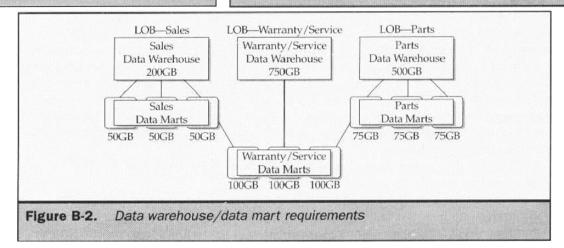
The Situation Analysis.

- The current application model consists of all **critical** application systems being run on the mainframes..
- These are the **current data** warehouses for sales, warranty, service, and parts.
- .They are currently loaded from the **operational databases** on the mainframe during a nightly batch cycle.
- The new application requirements are to **restructure** each data warehouse and **distribute** its data into a smaller data mart supporting each of the individual lines of business (LOB).
- The demands require that the warranty and service data warehouses be **consolidated** into a larger integrated data warehouse, along with a smaller but **integrated** data mart for each LOB.
- In addition, the larger, but longer term requirements are to build a **corporate** data warehouse to provide key executives with real-time analysis on sales and issues affecting sales.
- Figure B-1 depicts the current systems configuration overview and data capacities as provided by the storage administrator.
- The challenge is to develop an infrastructure that will house the information and **provide online access** to the predefined analysis that will be run against the data.

Figure B-1. An overview of the current systems configuration

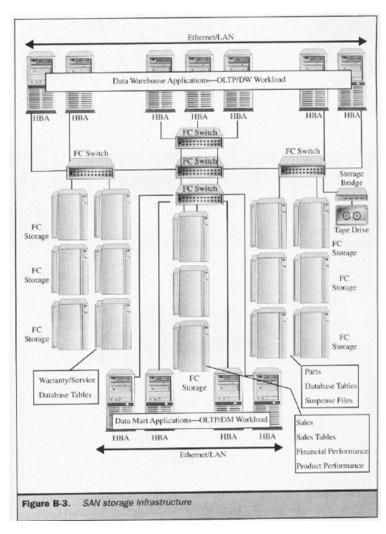| Processing Req.	Sales DW/DM	War/Service DW/DM	Parts DW/DM
Age of Data < 24 hours	✓	✓	
Age of Data > 24 hours			✓
Batch Loads	Mon–Fri: nightly	Mon–Fri: nightly	Mon–Sat: nightly
Response Time	< 3 seconds	< 10 seconds	< 10 seconds
Availability	Mon–Fri: 6 A.M.–7 P.M. Sat–Sun: 8 A.M.–5 P.M.	Mon–Sat: 24x7 Sun: as available	Mon–Sat: 6–7 P.M. Sun: as available

Table B-1. Processing and Service-Level Requirements for Data Warehouse and Data Marts

Figure B-2. Data warehouse/data mart requirements

The SAN Solution.

- The final configuration provides the **increased** storage **capacity** and enhanced **performance** that was expected of the solution.
- The data warehouse/data mart project was accomplished on time and has proved to be **responsive** to the **I/O workload**.
- Because the **estimates** were **accurate** and allowances were **built** into the configuration for expansion, the storage capacity and access performance is ahead of the I/O workload estimates and business requirements, making both of these execute within service levels.
- The current configurations are moving into the **planning** for the next highly **visibly** project, the corporate data warehouse.
- The configuration shown in Figure B-3 illustrates a **summarization** of the data warehouse/data mart application systems.
- However, it`s important to point out the results that the SAN solution provided this company with the ability to manage its business more effectively through the **use** of **multiple technologies**.
- More importantly, the applications have been used to **reduce** the **warranty** receivables significantly, and that results in an important improvement to the company bottom line.

SAN Storage Infrastructure.

Figure B-3. *SAN storage infrastructure*

Thank you