Paper 129

The SAS® Data Warehouse and Data Integration in Sasol Synthetic Fuels

Piet Venter – SASOL Information and Logistics Division - Johannesburg South Africa

ABSTRACT

Any Company experiencing rapid growth without a perfectly mastered IT systems development roadmap sooner or later suffers from various degrees of data integration problems. This occurs throughout the different systems lifecycles and worsens as freestanding systems excels and matures in their isolated domains of functional application.

This paper highlights the historical systems development situation at Sasol Synthetic Fuels (SSF), the South African based fuels and chemicals from coal producers and discusses ways in which integration can and was addressed. A three level solution of physical integration, middelware collaboration and data warehousing are suggested. Examples of the application of Enterprise Resource Systems as a common source system platform for integration at database level, an in-between-systems layer of definition and communication as well as the data-warehousing paradigm as supported by the SAS® System are shown.

INTRODUCTION

When an organization suffers from uncontrollable growth in size and complexity of information systems, duplication and isolation of data elements seems inevitable. Although the systems are good for what they were designed for, a need always arises to combine their contents to enable operational, tactical or strategic decision making. Any attempt at integration is flawed by inconsistency in definition, time sequencing and other symptoms of lack of coherence and standards. The phenomenal growth of Sasol and the resultant demand for bigger and more complex systems contributed to these types of systems in SSF.

In order to rectify, an attempt was made towards data integration at three levels. Firstly, on a physical platform by combining data into an integrated source environment as is done with popular Enterprise Resource Planning (ERP) systems today. Alternatively, by inserting a communication highway between systems to enable data sharing, or using the principles of data warehousing of controlled separation of data from source systems into a new information retrieval base.

SASOL TODAY

Sasol grew from the pioneering beginnings in the 1950's when the S.A. Coal, Oil & Gas Corporation was founded. This led to the announcement and development of two new synfuels plants (which became SSF) in the mid and late seventies, contributing largely towards Sasol becoming the world leader in the benefication of coal to fuels and chemicals. Sasol is proudly South African and is committed to the development of South Africa, globalising and believes in building mutual beneficial relationships with all their stakeholders, locally and internationally.

Sasol has 25 000 employees, 15 500 share (stock) holders and being amongst the top 10 companies in South Africa, a market capitalization of R12 billion. (~$2 trillion)[1]. It supplies 38% of South Africa's liquid fuel needs through Synfuels (28%) and Natref oil refinery (10%). It also produces the majority of feedstocks used by the chemical and plastic industry in South Africa and produces more than 120 products.

SASOL SYNTHETIC FUELS

SSF believes in being "Winners with fuels and Chemicals" and combines coal, water (steam) and oxygen to produce oil, chemicals and gas through a unique Sasol Synthol process. To feed this process, coal are consumed at an average of 47 000 tons per day, steam generated at 4150 KPa and 435° C and Oxygen fed from the largest air extraction plants in the world. In order to support the data processing and information needs of such a vast operation, numerous systems evolved and matured in a complex data processing and information delivery environment.

.

THOSE EARLY SYSTEMS

The dedicated mainframe systems era of around 1978 to 1994 produced highly complex transactional-based system to satisfy the functional business requirements. In their very nature these systems were grouped together as being Financial, Commercial, Administrative, Human Resource Management etc. With the coming of the PC and the server environments, systems continue to exist and expand in functional silos. Being of such vast magnitude and complexity, even within a functional area, data definition and attributes would diversify and fragment. Desperate measures were now needed to enable the delivery of timely and accurate combined management information as effective and efficient as possible.

A CHANGE OF PLAN

New pressures on the business to expand same and new products into local and global markets changed the marketing needs. The change in the macro environment also resulted in new economic demands e.g. the announcement by the Government in 1995 of the phasing out of tariff protection received on the shortfall of income as calculated against the international oil price. Hereby a level of protection calculated at $17 per barrel would have resulted in a loss for SSF of $169,3 million by July 1997. By 1995 the Information Management expenditure topped R100 million per annum, mainly directed towards systems maintenance. Systems were exceptionally fragile, data integrity was bad and the flow of information was painfully slow.

A total strategic rethink and resultant transformation process elevated the need for proper strategic planning and governance of the information management function to corporate level. The need for new, integrated systems and information for decision making became critical, and SSF embarked on a total revamp of the Information Technology processing and delivery infrastructure.

A THREE PRONGED ATTACK

The need for a new Enterprise Resource Planning System (ERP) to combine the fragmented systems at the lowest level was needed. This would solve a large part of the integration dilemma, which were at that point in time being addressed by very innovative but costly and inflexible interfaces between systems. It was also realized that an ERP was no silver bullet to overcome all of the problems. Some systems should exist as they are and only the capability to interact and share data should be addressed. This brought information exchange and collaboration technologies to the table.

Up to about 1997 various custom-build Executive Information Systems (EIS) was attempted. This was accomplished by extracting data from various systems and consolidating it onto a server or even PC based delivery platform with Open Database Connectivity capability. Very innovative user interfaces in some high level language was build to interact with the data. Although extremely popular, these systems lacked the robustness and flexibility of a true industry standard EIS. This introduced the data warehousing paradigm of data extraction, transformation and forwarding of data onto a new platform to form the bases of decision information delivery for the future.

THE SSF ERP

Systems of different architecture and platforms were used to serve the areas of financial, human resource planning, commercial, administrative, while a product called MIMS® was used to serve mainly the maintenance needs. The latter also offered the total systems solution and was considered the logical and most cost effective choice to use as ERP. The implementation of MIMS® solved most of the pressing integration needs and currently serves as a good source of data for mainly the areas of financial, maintenance operation, commercial (procurement, contracting and stores management) etc.

Other advantages are:

- data could now be accessed from a common base of definition;
- higher degrees of data integrity can be achieved because elements are now only recorded once;
- links between functional areas are easier;
- there is more of a joint focus and setting of collective goals and
- more information transparency.

THE INFORMATION BUS

When data from different functional areas was combined, it was found that people tend not to speak to each other. Production would be interested in the output and efficiency of the facility, while the Financial Director would concentrate on profits and return on investment. In this process, systems tend to be so specialized and focussed that combining them physically would not make sense. A case in point would be the SSF Production Information (PI) and General Ledger (GL) systems. The calculation of for example unit cost, (production volume in relation to financial contribution and rate) from these two systems are of extreme importance to everybody, although data extracted are from two systems that evolved completely separate from each other.

It was also found that data integrity *per se* could be stepped up dramatically. It now makes sense to create a new layer of definition and consolidation between systems were certain validation and correction of data could be done prior to passing it on to the next system. It is also cost effective and technically viable in some cases to leave the source systems as they are and do the needed manipulation of data on some other physical or logical layer.

To enable this, a layer of definition and communication is inserted. This acts as the transaction messaging and data-passing vehicle between systems. It therefor enables the mechanisms of publish and subscribe (system A transmits all and B enrolls in what is needed), request and reply (A request a response, system B reacts) and broadcast and receive (A includes B in it's distribution list). SSF is now using this technology in the areas of data validation, integrity and change control and will expand the use in future.

The advantages are that systems can stay intact and integration can be more directed, or more uncontrolled, if preferred. It also suits the today's WEB-enabled paradigm of interaction. There are however a few drawbacks. This technology is very complex and still maturing in the market. Information security and the question of ownership and change control on common data definition and interfacing should be carefully managed.

THE DATA WAREHOUSE

The EIS systems that were build highlighted some problems. There should be a master plan to coordinate the data extraction efforts so as not to burden the source systems unnecessarily. Conflicting definition of data and results defeats the purpose of "one version of the truth" for all the different users. The platforms chosen to store and manipulated the extracted results were also not expandable enough and quickly ran out of steam. The programming effort required building and maintaining these systems were also not justifiable. One of the largest problems was security as users insisted on the exclusivity of "their" data.

A thorough market research showed that an end-to-end data warehousing solution was preferable and the SAS® System was chosen. The areas targeted for the first applications were Production Reporting, Unit Costing and Financial Reporting. Data for these different applications were extracted from the Process (PI) and ERP (MIMS) systems and fed into a new relationally defined data structure (called the operational data store, or ODS). Here data can be cleansed, manipulated, consolidated and aggregated into new usable forms. This platform also serves as source for the multidimensional data structures required for OLAP (Online Analytical Programming) efforts.

The following objectives for the Warehouse were set:

- Data should be consistent in definition
- Data should be accurate and reliable
- It should be quick and easy to access the data and
- Data manipulation should be flexible and expandable

In short – the objective is to get data *out,* therefore the data warehouse should be separated from the operational environment. Data should be kept in such a way that a sustainable competitive advantage is created. In this process the underlying problem of non-data integration can also be solved. The question would be if a data warehouse were the right way to accomplish this. Satisfying strategic and tactical information needs by using a simple query and reporting tool onto the source systems does not suffice. The data structures and access keying of those systems are built for transaction processing throughput. Sophisticated manipulation, extensive querying and data mining would kill the performance of such operational systems.

CONCLUSION

By using the three methods of an Enterprise Resource System, Information Messaging and Data Warehousing, data integration in SSF could be started. The road ahead is still long and winding but as lessons are learned and new technologies, tools skilled resources become available and users become more willing and knowledgeable, more inroads will be made. It is foreseen that by making full use of all the functional capabilities of the ERP data integrity will be corrected at source level. By exploiting the Information Messaging further, more data definition integration can be achieved. If the warehousing implementation effort is sustained, the SAS® System will in this way contribute significantly in the future in the fight for robust, integrated and trusted data.

By completing the warehouse definition and expanding the implementation of the SAS® Data Warehousing solution, it should complete the quest for perfect data.

REFERENCES

[1] 10^{12} One trillion us = one billion UK, Europe and SA. An exchange rate of R6 per $ assumed (Hornby. A. S., O*xford advanced learner's dictionary of current English.,* ® Oxford University Press 1974, p1018)

ACKNOWLEDGMENTS

The author would like to thank the members and users of the various systems departments in SASOL and especially the MIMS team for their inputs.

The management team of Sasol Synthetic Fuels gracefully allows the publication of this paper.

The SEUGI 16 users group was also confident in the content and presentation of this paper and their sponsorship to SUGI 24 is a great honor.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Sasol Information and Logistics Division
12 Baker Street,Rosebank 2196
P O Box 5486,Johannesburg 2000
Republic of South-Africa
Telephone +27 11 441 3198
Mobile +27 82 450 1985
Home +27 17 631 3431
Fax +27 11 442 6699
Internet piet.venter@sasol.com