Data Marts for the Semiconductor Industry

Sern-Im Lee, In-Bo Sim & Todd C. Williams, *COMPAQ Computer Corporation, Camas, WA, USA*

Abstract

Circuit designs and IC processing techniques become more complex every day. As a result the scope of manufacturing data required for analysis becomes larger and more critical. In order to handle the wealth of information, from a plethora of sources, many data warehousing concepts must be utilised. These include both data warehouses and data marts for all aspects of the manufacturing process. Sources of data for the circuit owner, quite often a Fabless entity, must include the wafer manufacturer, the IC manufacturer and the assembly shop. Currently all this information must also be fully Web-enabled. This allows the entire supply chain complete and confidential access to the information.

The concepts and requirements of a data mart are discussed in this article as well as some tools that assist in the data analysis process.

What is a Data Mart?

Careers have been built on the theoretical definitions of data warehouses and data marts. Those definitions are in hot debate in the decision support and business intelligence community. For the purposes of this discussion a data warehouse is an aggregation of data marts.

A data mart is a logical subset of the complete data warehouse, a complete pie-wedge of the overall data warehouse pie. A data warehouse is made up of the union of all its data marts. "Data marting" is the process of bringing together disparate data from multiple data sources from across departments for decision-support purposes.

Misconceptions of a Data Mart:

- It is not a "system" or a "product" that is driven by the information systems (IS) department.
- It is not a product where one shape fits all.
- It is not a system that gets created and installed and never touched or modified again.

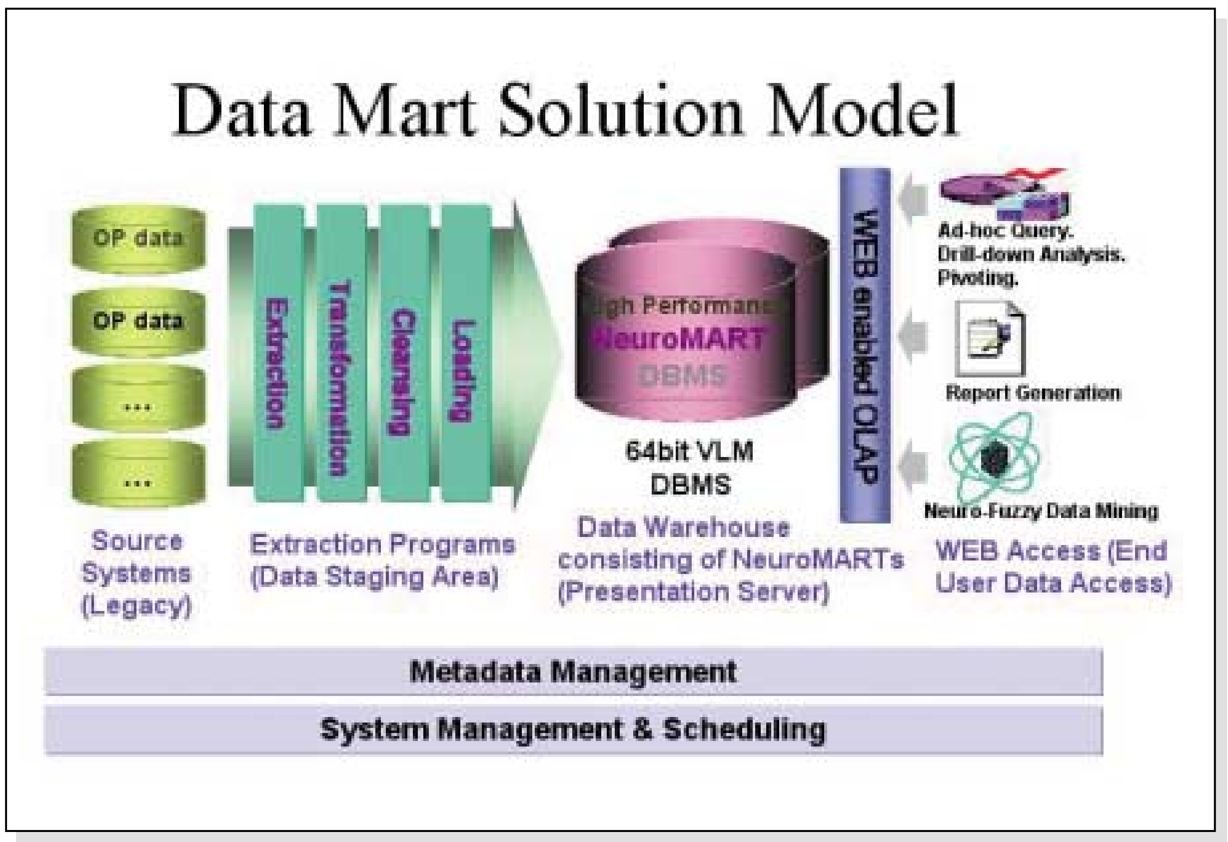


Figure 1
The data marts for semiconductor solution

What a Data Mart Is:

- It is an ongoing and continuously evolving "process".
- It is user driven and is continuously modified to fit the most current analysis needs of the users.
- It is a central component in the corporate decision support system (DSS) acting as a data gatherer.

Data marts are a physical entity, while data warehousing is the concept. Data warehousing is the act of binding data marts together under a common architecture to give its users a uniform access to enterprise data.

To ensure that the complete data warehouse will be robust and resilient in the face of continuously evolving requirements, data marts must follow a proper architecture that guides the design of the data marts.

The functional goals of data marts and data warehousing are:

- A query-set based database made for data analysis and decision support
- A specialised system to bring together the data needed and provide the right information at the right time.
- The basis for online analytical processing (OLAP), data mining, data warehousing and executive information systems.

Data warehousing, to its users, is a means of discovery. Discovery that is convenient, fast and easy because the data needed for a particular question is nicely organised together within the data warehouse.

USERS OF A DATA MART

The users of a data mart encompass a large group of people with varying degrees of skills and needs. From an organisational perspective, the users may be engineers, supervisors, managers or executives. From a supply chain perspective, the users could be customers, suppliers or interdepartmental clients. All of the users require different security levels to ensure proper access to only the data they should see. Every user must have the ability to get information they need without interruption and the administrators of the system must have tools to provide that data.

DATA MINING TOOL

A data-mining tool is a sophisticated part of a data warehouse with analytic capabilities that transform and digest the output from the data warehouse. Two applications of mining models are forecasting systems that try to predict the future outcomes and yield management models that cluster and classify wafer defect patterns and correlate the findings with equipment to find the responsible equipment and recipe.

DATA MART FEATURES

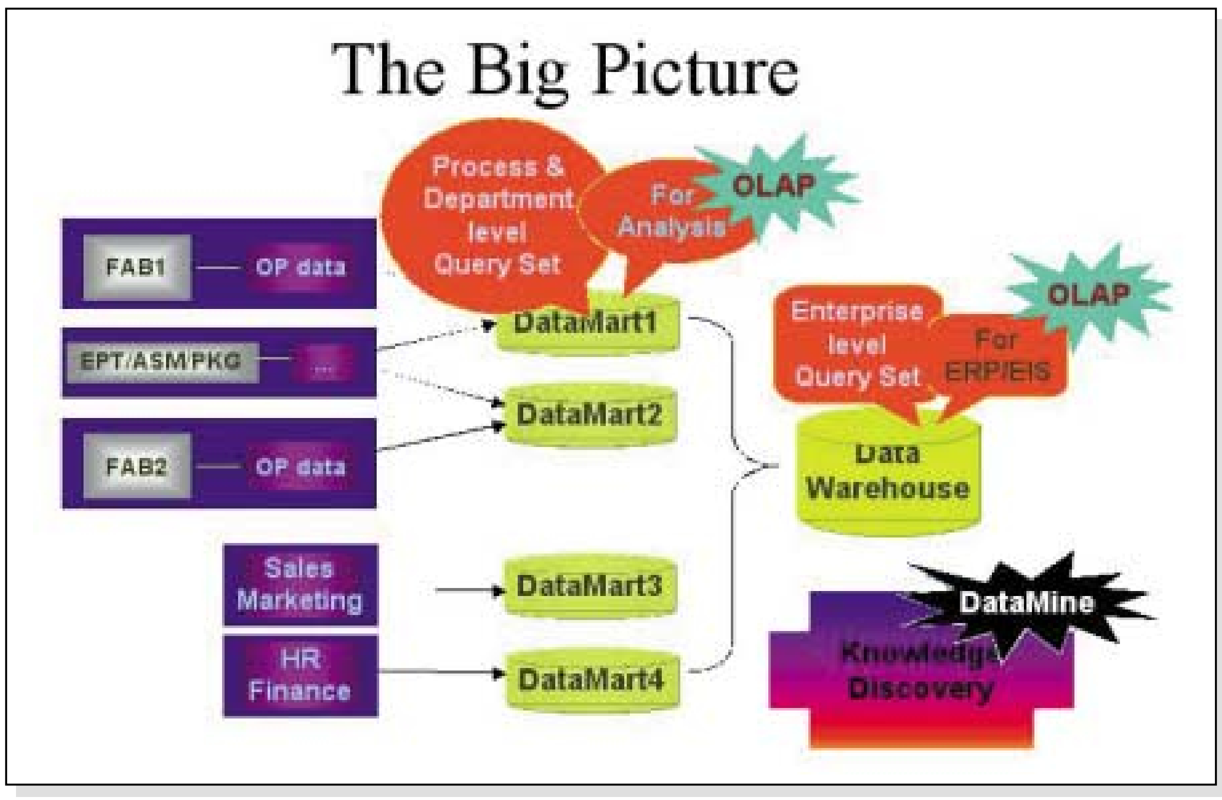
Data marts for semiconductor are a system integration solution applied to the semiconductor industry. Various systems inside the supply chain must be exposed and integrated with other systems in the supply chain to create one virtual system.

DATA WAREHOUSING TECHNOLOGY

Having data warehousing technology in data marts for semiconductor has many benefits to engineers, because data marts for semiconductor are about a new way of thinking about data.

An engineer asks ‘what is the current yield for process ABC for equipment = ‘1,2,3’?’ Questions like these get answers for that process, that equipment, on that one day.

Figure 2
How it fits together



A yield manager or a QC engineer takes a different perspective and considers whether that particular process ABC is profitable, should every lot go through that process or should the process order be changed from equipment 1,2,3 to 1,3,2?

With data marts, an effective analyst can step back from the details of today’s operations and take a broader view of the semiconductor business. Because the analyst only has to wait a few minutes for answers, the train of thought will continue to invite more questions and arrive at a decision based on facts.

For the company investing in a data mart, the initial return comes from having existing processes automated, putting reports online and giving engineers a clean source of data. The biggest return comes from storing company knowledge and having improved access to data that can spur innovation and creativity, which comes from new ways of looking at and analysing data.

DATA ANALYSIS SCOPE

This is where data marts for semiconductor differ from other MES, EDA and CIM systems, the data scope. Data marts handle correlated overall analysis for Fab, parametric testing, package, and assembly line data. This is cross-lines, cross-Fab data handling. Engineers can do recursive inductive analysis for specific problems – abnormal lots and equipment.

REPORT GENERATION

Using the OLAP tool, ad hoc reports should be easy to create. Any authorised user should be able to create unique reports, through a GUI using a drag-and-drop format. There should be no need for the IT department to change the code in either GUIs or the data marts for ad hoc query purposes. Engineers and end-users have the control of what data to see, how they want to see it, and when they want to see it, now and in the future.

Report distribution should be automatic via the Web, e-mail or a repository located inside the data mart database. A paperless office environment is possible as users only need to have the right security to gain access to their Web server home page and view reports. With proper security, reports can even be viewed from outside the company’s firewalls.

SCHEDULED BATCH

Once reports have been created, their distribution should be automatic. The OLAP tool should handle the scheduled batch processing of reports. Users should be able to schedule reports for daily batch runs and then have them published to the Web for viewing.

LOT HISTORY TRACE

Lot history trace allows tracing a lot and its family members from ingot to its finished chip state. Users should be able to specify the lot they want to trace, which process and recipe to trace and they would get that lot’s data and its family’s data in return.

Lot history trace can cover lot state changes such as lot split, lot merge, lot change, inter-route lot merge, bonus and others. For instance, when a lot A splits into A1 and A2, they are considered to be family members of lot A, so the lot history trace will list A as the parent and A1 and A2 as children. It can also give detailed data on the lot at the time of the state changes (i.e. wafer id, start time, operator who was responsible, equipment that did the recipe, wafer sort tester values, etc.). The tracing data is available from Fab all the way to packaging.

WEB-ENABLED INFORMATION SHARING

The solution should be able to replicate a user's actions to create reports and auto-publish them to the world-wide Web. Users should have the capacity to enter their username and password, and use their favourite Web browser to log on to their Web accounts and view their daily, hourly or weekly reports.

STABILITY

A data mart for semiconductor should be a three-tier client/server architecture consisting of a data mart server (database), a batch schedule server and/or a Web server, and a client (OLAP) for end users.

The data mart server should, at the minimum, be a two-node cluster for high availability (amount of time the system is available to the end users) in case of hardware failures.

For database failures, data marts should have fail over functionality designed into the data mart database. The underlying database should allow for queries to fail over and reconnect to the other node so that they can continue to run.

ROBUSTNESS

A data mart's robustness (the ability of software architecture to meet current demands as well as to grow easily to meet greater needs in the future) comes from its design philosophy. Each component should be modular, so that changes are made to just one part that needs to be changed and nowhere else, and the distribution of changed parts to users should be automatic.

When new tables or fields are added to the data mart the only change required to use that new data should be to adjust the OLAP's view of the database, and save it to the OLAP repository inside the data mart. Distribution to users is then simple, users will automatically get the latest copy of the OLAP 'universe' (a business view of the underlying data) when they restart the OLAP GUI. New reports using the new table data should be made with no changes in code to any GUI, all this taking minutes, not days.

SECURITY

The data contained in a data mart is not just the raw manufacturing data of the company, it also houses the knowledge base of the analysis know-how. In addition, this information should be exposed, not only inside the company but also outside the company. The users could be customers; they, in turn, may be each other's competitors. Therefore, data marts need a tight security mechanism. Access must be restricted through stringent, yet easy to maintain, access control methods.

Data marts must offer individualised access privileges, where different user groups can have different 'privileges' in viewing data and accessing reports. For example, a trainee engineer may need to be restricted to view only the data related to his set of equipment, while a senior engineer may need to see any data in a Fab and a QC engineer may need to see any yield data in all the Fabs. These access rights should not be set on the data mart database itself, but done from the OLAP's supervisor GUI, which should be very intuitive and easy to use.

THE TOOLS FOR ANALYSIS

The OLAP tool

An on-line analytical processing (OLAP) tool must act as a front-end to the data warehouse, the presentation layer for end-users. It provides a semantic layer, an intuitive business view of the data. It should hide the nuts and bolts of the data warehouse so that its users need not understand the technical details. Ad hoc query (users make their own queries by directly manipulating relational tables and their joins, using a GUI), drill and slice and dice are some of the characteristics of the OLAP tool.

A well-designed OLAP presentation layer provides integrated query, reporting, and analysis in one GUI and makes the data warehouse easy and fun to use. The basic functions of the OLAP layer should include the following high level capabilities:

- An integrated query, reporting and analysis solution. It must accommodate users that need the power of an enterprise reporting tool or prefer the ease of an executive information system (EIS).
- An enterprise-wide reporting and broadcast component. All businesses rely on sharing a consistent view of corporate data across the enterprise. With a broadcast agent, users can schedule, publish, and distribute reports autonomously in a controlled environment, wherever their location.
- Configuration and support tools to streamline the set-up of the OLAP tool. These tools should be graphical design tools that include routines for automatic design creation.
- Tools to provide administrators with an environment to manage and administer the mass deployment of the solution throughout the enterprise. It should include tools that provide security and user management on a minimal number of screens.
- Since there are many common sources of data in an enterprise, the OLAP tools should come with a multitude of data source loaders. At a minimum, support should include access to: Oracle, Informix, Sybase, Microsoft SQL Server, DB2, CA-Ingres, Rdb and others, as well as support for OLAP servers including Arbor Essbase, Oracle Express, Informix Metacube, and Microsoft OLE DB for OLAP.

WEB ACCESS MODULE

The data mart must be fully Web-enabled. Users must be able to Web-publish and do query/analyse/report on the Web. The Web access module must be fully integrated with the data mart, batch scheduler as well as the OLAP tool. Web access must be available both over the Internet and the Intranet.

Important and frequently viewed reports would be published to the Web and refreshed with the latest data every morning. A "paperless office environment" can then be achieved and users should only need their username and password to view the reports with their Web browser.

The integrated query, reporting, and analysis solution should also be available on the Web. This allows thin-clients to do the same kind of analysis that heavy-clients with OLAP installed can do. This gives all users access to the data, both over an intranet for corporate decision support, and over an extranet for customers, partners, and suppliers. This access to "e-business intelligence" drives huge opportunities for companies to increase profit, reduce cost, and forge new customer relationships.

HOW DATA MARTS ARE USED

Using data warehousing techniques to do engineering data analysis has many benefits to the semiconductor industry. Following are two case studies that are actual experiences based on the development of Compaq's Neuro-MART for Semiconductor.

CASE STUDY 1:

An analyst wants to know "where are the group of lots that passed Fab-recipe5 on January 1, 1999 today, March 6 1999". With the time difference of more than 2 months, some of these lots would be in the assembly process, some are still in the packaging process, some have gone through rework processes (repeatedly) and some may have been scrapped.

What are the steps needed to perform that analysis in the pre-data mart environment? The analyst first connects to and searches the Fab database to see which lots passed Fab-recipe5 on January 1, 1999. Then, with the list of lots found, the user would search each of the parametric testing, assembly and packaging databases to see if these lots are in there. Then once all the data was found it needed to be merged into an Excel file, do some cut and pasting of the data and run Excel macros to make the final report which lists "the current lot status of lots that passed Fab-recipe5 on January 1, 1999".

In this case, the process took two full days of a senior engineer's time to make that final report. The main problem was with the actual data-inconsistency in information and difficulty in correlating the information due to multiple data sources. Data generated from multiple MES and CIM systems used proprietary data formats, so after compiling the data for analysis there was redundancy and gaps along with multiple names for the same data.

In addition, logical linkage between the different data sources did not exist. Data was located in separate databases where the actual production lines were Fab data with the Fab line, packaging data with the packaging line, etc. This made lot history tracing from ingot to chip impossible.

Due to this, when a customer complaint was filed on a single chip that was sold three months ago, it was very difficult to point to the source of the problem. Information for when processing occurred, which recipe, process, equipment and department were spread out over "islands of data". The link between them did not exist. Hunches would need to be made and searches of the "islands" made for the root cause of the problem. Even when a reasonable cause was found, convincing the department involved that it was one of their processes that was causing the overall low yield was difficult, as they always questioned the data correctness and reliability.

The solution to the problem was to create data marts for engineering data analysis (EDA) purposes. These EDA data marts were built for each line and then integrated to create one virtual Fab a seamless information sharing process across plants and CIM systems geographically separated and using different system vendors. Lot history tracing from Fab, parametric testing, assembly, packaging was possible.

In this environment, the senior engineer could gather all the data needed in less than an hour, do the analysis based on facts, rather than making educated guesses. Convincing other departments of the findings was easy as there would be a complete audit trail supporting the findings.

CASE STUDY 2:

How long would it take and what investment would a company need to make to raise one expert-level yield analyst? Ascending the learning curve is slow. With proper training, experience, time, data and eagerness on the engineer's part, a company can raise one such analyst in a few years' time. The analyst can perhaps pass on the analysis know-how to junior analysts in the department by training them on how a certain analysis report works and what critical parameters to look for. But, what happens when other departments need the same set of analysis know-how? Do you lend the analysts out? Or worse yet, what does a company do when the analyst leaves, taking the know-how?

Many information technology (IT) departments make engineering databases, develop GUIs or buy multiple EDA tools in hopes of getting a "reporting system" and getting a head start on analytical know-how. The executives hope they can be freed from this "analysis know-how stored in individual people's brains", and hope these measures will raise the average engineers' analytic abilities.

Engineers with these GUIs and EDA tools would view data and make reports from what-was-in-the-GUI-menu. When they needed a new data item or new functionality, they would make a request of IT and wait for system upgrades or wait for the EDA tool vendor to release upgrades.

These measures may work for some engineers, but for those innovative engineers who just cannot wait half a year to see that certain bit of data matched against the data items in his report, it will not be good enough. They will continually request ad hoc data analysis tools where they can make reports independent of what was already offered in the GUI report's menu and ask whatever questions they wanted to ask, and get answers now. Companies striving to succeed cannot wait to see new correlations between data.

What was needed was a knowledge-based database where analytical knowledge of the company's own senior engineers' know-how could be stored. They already had many versions of engineering databases and data processing GUIs that they had made on their own. They already had many EDA tools they acquired from the market. They were already confident of their senior engineers' analytical experience and know-how about their company's analytical needs. They wanted this know-how captured for others to use. They wanted to save and share this analysis know-how between engineers.

In addition, report distribution throughout the organisation was time-consuming. The reports were made during the night shifts and printed out each morning to be reviewed by each engineer. More than 20 000 queries were being made against the databases to make 4000 reports (per Fab line) which came to an incredible count when summed factory-wide. Many of these reports had similar contents, so a more efficient way to share was needed.

In a data-warehousing environment, a ROLAP (relational OLAP) tool was provided to fully utilise the answer-sets stored in the data marts.

Most of the steps needed to do proper analysis would already be taken care of within the data mart's database design, as at the database creation time all the experts' analysis patterns are analysed and applied. What used to take two days and ten processing steps would now take fifteen minutes and two steps from the data mart's presentation layer, the OLAP tool.

Ad hoc data analysis and new report creation was simplified and only took minutes. A one-stop service – from automatic data gathering and analysis and manipulation to Web publishing in one process – is very possible under this architecture.

Important and frequently viewed reports would be published to the Web and refreshed with the latest data every morning. A "paperless office environment" can be achieved and users would only need their username and password to view the reports with their Web browser.

More importantly, a set of core reports used most by senior engineers would be created and stored into the data marts so that a novice engineer could select reports and learn the "how-to" skills and the analysis knowledge of senior engineers. What this means to the company is that the expert knowledge would stay with the company and remain a company asset.

In addition, the time required to raise one expert analyst was shortened, the learning curve will start not from the bottom but from the middle as time, data and analysis know-how will be available in the database to give engineers a jump start.

How many companies can say, "We have a way to store and share our own unique analytical knowledge and the means to grow upon it, together"?

CONCLUSION

From a data marts perspective, other EDA tools, MES and CIM systems are vertical. They have their primary focus on "data analysis for a certain set of processes or equipment".

They are good for a certain predefined purpose, but have the tendency to create "islands of analytic data and information". Islands in that each come with their own operating systems, their own databases, their own data formats, and their own data naming rules. These are fine and useful inside the island, but very confusing once outside. When one tries to merge data from islands, it becomes nearly impossible to put these bits together to make useful information.

Data marts differ from these tools in the "data analysis scope". While a "bin and bit data analysis tool" offers a specific bit and bin related process data analysis, data marts for semiconductor are "a data warehouse-based, overall engineering data analysis tool" that can analyse any and all kinds of data that exist in Fab, parametric testing, assembly and packaging. An overall aerial view is possible.

Take the example of lot history tracing. In the MES lot, history tracing is done to detail while the lot is moving within a production facility. But once the lot moves to another facility, like parametric testing, assembly, and packaging, the lot becomes untraceable. In a data mart lot history tracing is done supply chain wide from Fab to packaging.

Data marts and these vertical tools co-exist. Data marts do not replace the MES, EDA or CIM system. To a data mart these systems are just another data source that will be integrated into the data marts.

ABOUT THE AUTHORS:

Sern-Im Lee is a systems integration specialist for COMPAQ and has worked on CIM installations at Samsung, and data warehousing projects with Samsung, LG electronics and Korea Life Insurance. Presently Ms. Lee is working in the Professional Services Division of COMPAQ Korea developing data warehousing & data mining solutions. Ms. Lee's current project is implementing NeuroMART for Semiconductor at a foundry Fab in Southeast Asia.

In-Bo Sim is a systems integration specialist for COMPAQ and had worked in the Korea Advanced Institute of Science & Technology developing a data analysis system for non-linear surface. Mr. Sim has worked on CIM installations and data warehousing projects with Samsung and LG Electronics. Presently, Mr. Sim is working in the Professional Services Division of COMPAQ Korea developing data warehousing, data mining and artificial intelligence solutions. Mr. Sim's current project is implementing NeuroMART for Semiconductor at a foundry Fab in Southeast Asia.

Todd C. Williams is a solution architect for COMPAQ and has worked on CIM installations at TSMC, Winbond, WaferTech, MASCA and Komatsu Silicon America. Presently, Mr. Williams is working in the Industry Solutions Division of COMPAQ designing solution sets for the Semiconductor Industry.

IF YOU HAVE ANY ENQUIRIES REGARDING THE CONTENT OF THIS ARTICLE, PLEASE CONTACT:

Todd C. Williams
Compaq Computer Corporation
833 NW 24th Avenue
Camas
WA 98607-9365
USA

Tel: +1 (360) 834-7361
Fax: +1 (360) 834-0574
E-mail: todd.c.williams@compaq.com

Reader Ref [7]