Acquiring Data Warehousing Systems Guidelines

**Table of Contents**

[1 Acquiring Data Warehousing Systems Guidelines 2](#_Toc44074211)

[2 PROCESS CHECKLIST – PLAN & DEFINE PHASES 4](#_Toc44074212)

[3 PROCESS CHECKLIST – DESIGN & BUILD PHASE 5](#_Toc44074213)

# Acquiring Data Warehousing Systems Guidelines

The following table describes guidelines to implement data warehousing systems with the support of external suppliers. The list of considerations following this table can be used to ensure that unique data warehousing characteristics are factored into the planning and execution of each phase.

|  |  |
| --- | --- |
| **Phase** | **Interpretation Guidelines** |
| Plan & Define | The major challenge in setting up a data warehouse is defining the business needs it will support. Typically the project team engages users who are currently analyzing business data to  determine:   * How data is being used today * How today’s data could be used if data warehouse support were available * How data warehouse availability might impact data usage in the future * What business questions are being answered – and not answered – using today’s operational data.   User input is captured in system requirements specifications that are forwarded to candidate suppliers, usually in the form of a Request for Proposal (RFP). If internal expertise is not available to capture requirements for the data warehouse sufficiently, the project team may choose to engage external expertise to assist in the requirements specification effort.  By the end of the Define Phase, suppliers for the data warehouse should be under contract with requirements and acceptance criteria well understood. The Design and Build Phases then focus on monitoring these suppliers as the data warehouse is implemented.  *See* [*Plan & Define Phase Items to Consider*](#_PROCESS_CHECKLIST_–) |
| Design & Build | During the Design Phase, the system requirements are converted into system specifications by deriving logical (Entity-Relationship Diagrams) and physical data models (meta-models). From the specifications, the data warehouse functions can be designed that will extract source data, transform and integrate it, load it into the warehouse and present it to users in the desired forms. Numerous tools are available to support data warehousing implementation. These tools should be considered to minimize the amount of custom development and to expedite the launch of the data warehouse.  Implementation of the data warehouse proceeds on several paths with suppliers performing the bulk of the activities. The supplier’s activities typically include:   * Analyzing source data to determine the amount of cleansing that is required prior to loading it into the warehouse. * Developing procedures for moving data from its sources to the data warehouse. * Constructing the data warehouse infrastructure, including integrating selected tools. * Integrating data sources into the data warehouse, cleansing and transforming the data during the load process. * Documenting metadata, programs and applications.   An internal project personnel focus on monitoring the supplier’s activities and engaging data warehouse users periodically to ensure the solution will meet their expectations. When the data has been loaded, users assist in verifying that the data has been successfully loaded into the warehouse (often by comparing reports from the data warehouse with trusted reports from the source systems).  *See* [*Design and Build Phase Items to Consider*](#_PROCESS_CHECKLIST_–_1) |
| Deploy | In general, the deployment of a data warehouse should not differ significantly from the deployment guidance specified and used on other projects. Structured planning and tracking in concert with a Deployment Plan provide the project with the best approach for successfully launching the data warehouse.  The data warehouse should be transitioned into the user environment with well-defined content and update frequency to ensure that users trust query results. As more users with diverse needs are added, the data requirements for the warehouse evolve, often requiring integration of new data sources, new queries and customized reports. Planning and preparation for handling these evolving needs is crucial to the long-term success of the data warehousing project. |

# PROCESS CHECKLIST – PLAN & DEFINE PHASES

|  |  |
| --- | --- |
| **Intended use of this checklist** | To assist teams for acquiring Data Warehousing systems. Use this checklist in conjunction with other process assets during the Plan and Define Phases. |

|  |  |  |
| --- | --- | --- |
| **PLAN & DEFINE PHASES** | | |
| **ID** | **✓** | **Items to Consider** |
| 1 |  | Are the business users for the data warehouse identified? |
| 2 |  | Are the business questions that the users are trying to answer clearly understood? |
| 3 |  | Have you examined what current users of the data are doing to process and analyze the data? |
| 4 |  | Have users provided their ‘wish list’ of things they would like to do with the data today? |
| 5 |  | Have users described what they plan to do with data in the future? |
| 6 |  | Has the impact of the data warehouse on user workflow been examined? |
| 7 |  | Are business rules associated with the data well understood? Are there regional differences that need to be accounted for? |
| 8 |  | Have all data sources that will be integrated into the data warehouse and the specific data that will be extracted from them been identified? |
| 9 |  | Have external data sources (such as a Dow Jones feed) been specified? |
| 10 |  | Do plans and the architecture allow additional data sources to be added in the future? |
| 11 |  | Will the data warehouse interface with other data warehouses or data marts (subject- or business unit-oriented data warehouses)? |
| 12 |  | Is the life span of the data warehouse and its data well understood? |
| 13 |  | Has the amount of required historical data been determined? 6 months worth? A year’s worth? |
| 14 |  | Do plans and business cases comprehend the overhead associated with maintaining the data warehouse and its data sources to ensure that it is robust? |
| 15 |  | Are the integration costs and future value of additional data sources factored into the  business case? |
| 16 |  | Have you identified metadata, that is the patterns which trace data from its source to the data warehouse and then to user applications? (Metadata is often described as ‘data about data’) |
| 17 |  | Have required system queries been defined and documented? |
| 18 |  | Are the reports required from the system defined and documented? |
| 19 |  | Are graphical presentations of the data required? Is ‘Drill down’ capability? Data browsing? |
| 20 |  | Have you determined what standards and open interfaces will be used? |
| 21 |  | Have security requirements for the data been considered? |
| 22 |  | Have the deployment requirements been thoroughly defined including the tools and methods for accessing data, client platform configurations, and infrastructure requirements? |
| 23 |  | Do plans for implementing the data warehouse include data transformation (extraction, cleansing/scrubbing, mapping, demoralization, loading, etc.)? |
| 24 |  | Is external data warehousing expertise needed to supplement internal staff for Define Phase activities? |
| 25 |  | Will external resources be required to support deployment of the data warehouse? |

# PROCESS CHECKLIST – DESIGN & BUILD PHASE

|  |  |
| --- | --- |
| **Intended use of this checklist** | To assist teams for acquiring Data Warehousing systems. Use this checklist in conjunction with other process assets during the Design and Build Phases. |

| **Design & Build Phases** | | |
| --- | --- | --- |
| **ID** | **✓** | **Items to Consider** |
| 1 |  | Have user requirements been decomposed into Business entities and their attributes, entity relationships (E-R Diagrams), and hierarchies? |
| 2 |  | Is the data warehouse design documented (and kept current) to ensure that the data is robust? (Note: it is dangerous to delay documenting any information relative to the data warehouse that may lead to errors and cause users to lose faith in the data) |
| 3 |  | Does the proposed data warehouse design provide match the expected query needs of the user? (Data warehouse models often employ a star or snowflake schema with a single object in the center radially connected to other objects) |
| 4 |  | Will the data warehouse schema support integrating new sources of data? |
| 5 |  | Does the data warehouse provide users the capability to access indexed information (aggregated, summarized and integrated data) as well as view detail data by performing more sophisticated I/O functions on the database (e.g., data mining- looking for hidden patterns)? |
| 6 |  | Does the architecture support ‘real-time’ access to indexed aggregate data with slightly longer seek times for source data? Have these expected performance levels been specified? |
| 7 |  | Have data formats for the data warehouse been defined and do they comply with corporate IT standards and with any other relevant corporate data warehouse initiatives? |
| 8 |  | Do the warehouse’s data formats comprehend the transformations that occur on incoming source data? |
| 9 |  | Has operational data been mapped to the target warehouse formats? |
| 10 |  | Has the expected frequency of data uploads to the data warehouse been defined? |
| 11 |  | Does the performance of the proposed system support the desired upload frequency? |
| 12 |  | Does the data warehouse support verifications immediately after upload to ensure that no data has been corrupted? |
| 13 |  | Has the verification procedures for uploaded data been defined? Are all data sources accounted for? |
| 14 |  | Has the infrastructure for transporting data been fully defined? (examples include: network components, hardware/software platforms, network protocols, network, management systems, security (including firewall access for external data sources), and middleware) |
| 15 |  | Does network bandwidth support user performance requirements (fast response times) as data warehouses often require moving large amounts of data in relatively short timeframes? |
| 16 |  | Will the network architecture support expansion to accommodate growth in the number of users? |
| 17 |  | Have the tools necessary to extract required information from the data warehouse been identified if they are not being developed? |
| 18 |  | Are existing tools being used wherever feasible to minimize the amount of custom development in the application? |
| 19 |  | Have interim development releases of the data warehouse been reviewed with end users to ensure that the output of the data warehouse meets user needs? |
| 20 |  | Has user documentation that describes the specific data and associated queries within the data warehouse been provided to prevent misinterpretations? |
| 21 |  | Have processes for ongoing administration and housekeeping of the data warehouse been defined? |
| 22 |  | Are plans in place to support user requests for addition queries and query types? |
| 23 |  | Are plans in place to support the integration of additional data sources? |