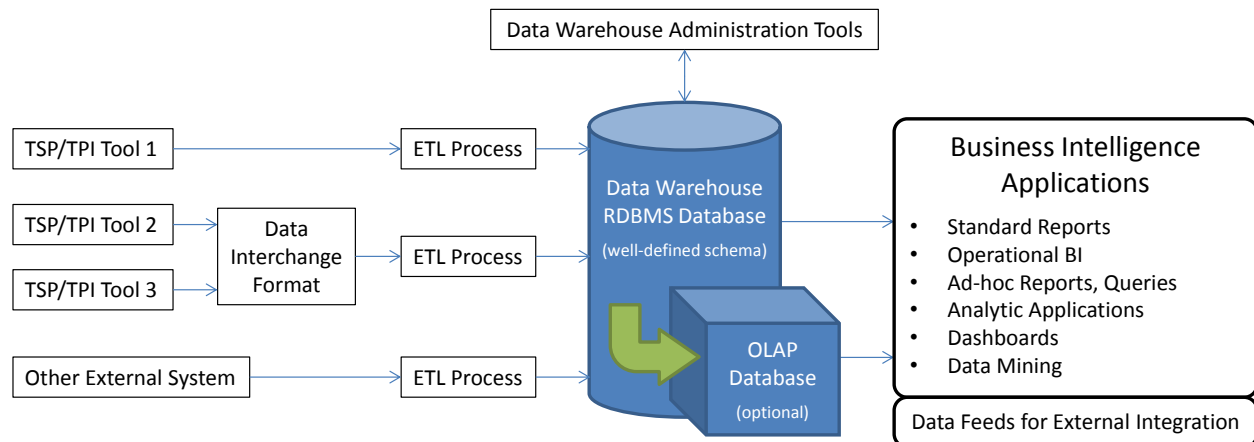


## Team Process\* Data Warehouse Components

The data warehouse components will be designed in a modular way. The primary subsystems in the warehouse are depicted below, and described in subsequent sections.



### RDBMS Database

The central component of the data warehouse will be a RDBMS database. This database will hold metrics, metadata, and other content as described in the requirements document.

Modern data warehouse architectures provide various options for the data store, including specialized products that are designed to efficiently handle massive amounts of data in multidimensional OLAP cubes. However, the volume of TSP data collected by a large organization is still likely to be small in comparison to other modern data warehouse workloads, so specialized data store products may not be necessary to achieve adequate performance. With this in mind, and with the goal of maximizing the reusability of the data store for other reporting/integration purposes, a standard relational database will be used to hold the data. However, the design of this relational database will follow best practices for dimensional modeling (for example, star-schemas and data bus architectures), allowing third-party OLAP products to provide additional query speed and scalability on top of the provided relational data.

To the extent possible, the design of this database schema should strive for portability. Organizations will prefer to deploy to different database platforms (for example: MySQL, Oracle, or Microsoft SQL Server) based on their scalability needs or their pre-existing enterprise software license purchases. Accordingly, the core schema of the RDBMS database should not rely on features that are specific to a particular database product.

---

\* TSP<sup>SM</sup> is used by teams working in a wide variety of problem domains (e.g. software, hardware, services). Since these activities are not limited to software, the name "Team Process Integrated" and the acronym "TPI" are used in this document to describe the full range of TSP-inspired high-maturity processes, and to avoid improper use of Carnegie Mellon service marks. TSP is a service mark of Carnegie Mellon University. Carnegie Mellon University has neither contributed to nor evaluated the contents of this document.

Without question, platform-specific SQL statements will be needed to create the objects in the RDBMS database and to load data from an ETL process. Where possible, database abstraction layers will be used to shield other system components from these platform-specific constraints. However, the core schema will strive to be consistent across platforms, to best enable the creation of standard/reusable reports and ETL processes.

Extensibility and customizability will also be important characteristics of the data warehouse. The primary mechanism for extensibility will be through the creation of new/custom database objects. For example, an organization with specialized analysis needs might create additional tables in the database which store custom attributes for projects or teams. If an organization adds custom columns to a table in the core schema, those custom columns should be nullable or should have default values to avoid breaking standard ETL logic.

### **ETL (Extract-Transform-Load) Logic**

In keeping with best practices for data warehouse architecture, standalone components will be written to extract data from source systems, transform it as necessary, and load it into the data warehouse.

Some TSP/TPI tool vendors may explicitly be data-warehouse aware. These vendors may choose to write ETL logic that extracts data directly from their tool and loads the data into the warehouse.

Other vendors may not wish to be coupled to the data warehouse. These tool vendors may prefer to export data from their tool into some agreed-upon data interchange format. A reusable ETL process could be written that would read these files and load them into the warehouse.

Finally, some organizations might choose to write custom ETL logic that reads data from external systems and loads that data into the warehouse. For example, an organization might:

- Read data from their corporate directory to populate the hierarchical list of suborganizations in the warehouse
- Read data from a corporate defect tracking system and load this data into corresponding warehouse tables
- Populate custom warehouse tables/columns with data read from some other system

The data warehouse should allow flexibility in the design and operation of ETL processes. For example, the warehouse design should not preclude the operation of ETL logic that loads data in near-real-time.

### **Business Intelligence Applications**

Analysis and reporting are fundamentally important to the data warehouse. The primary value proposition for the warehouse is data mining and analysis, and reports will be a critical link in this value chain.

Since many organizations will make use of the data warehouse, there will be significant ROI from the creation of a set of standard reports. For example, reports could be created which display data similar to the TSP Project Plan Summary forms (but with data from an entire organization, rather than a single

project). As new reporting requirements are identified over time, new reports could be added to the standard collection to address these requirements.

Of course, organizations will always identify special reporting needs that are unique to their organization, or that are relevant to their senior management. Since the data is stored in a relational database, these organizations can make use of open-source or COTS products for ad-hoc reporting, data mining, or business intelligence. Examples include Crystal Reports, Pentaho Analysis Services, and many others.

Organizations may also wish to feed TPI data into other systems. For example, they might wish to update a corporate reporting system with the “% Complete” status of various components in a TPI project. Since the warehouse data is stored in a simple RDBMS format, organizations can read the data using simple SQL queries and transfer that data into other systems as needed.

Since the TPI data warehouse will be designed around a standardized schema, this also creates the possibility for an entirely new market of products. Third-party vendors could design value-added reporting packages that read data from the standard TPI data warehouse schema and generate specialized reports, such as:

- Reports that view data from an entire organization and produce a “dashboard” for consumption by senior executives
- Reports that analyze data from a particular project, and:
  - Identify problems, such as missing data or inadequate reviews
  - Identify opportunities for process improvement
  - Provide analysis assistance to planning managers, quality managers, and coaches

Vendors could create these analysis packages to operate against the common warehouse schema, and simultaneously allow the analysis of data from any number of TSP/TPI tools that were warehouse-aware.

## Administration Tools

Within the warehouse, certain common tools will be needed to:

- Configure data sources for standard ETL processes
- Configure settings for security and data privacy
- Perform maintenance activities, such as backups or purges of old data
- Optionally configure static data (such as hierarchical org structure) which might not be populated by an automated ETL process in a particular organization