

Analysis: The process of exploring data and reports in order to extract **meaningful insights**, which can be used to better understand and improve business performance. Analysis transforms data and information into **insights**.

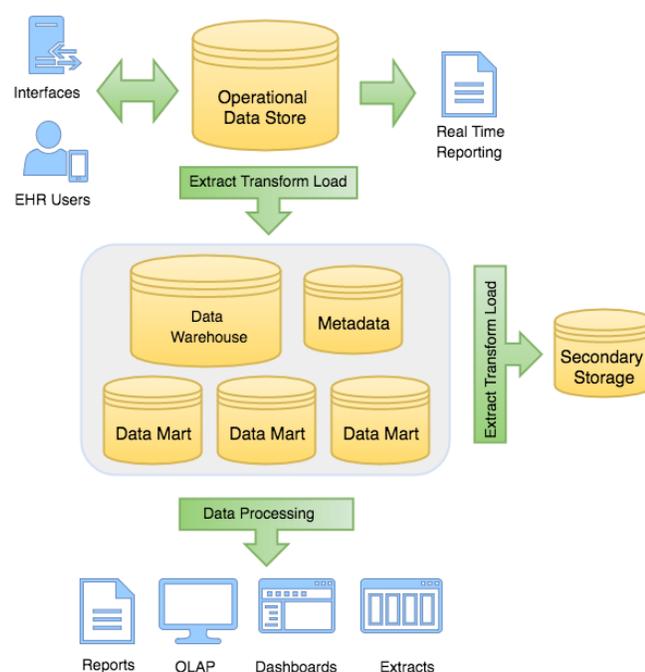
Reporting and Analysis activities result a number of **artifacts**. These artifacts are for example, report files that are stored on network shares, or other entities produces within corresponding analytical infrastructure – such as cubes, views, and dashboards. Other types of artifacts associated with analytical activities are those that facilitate these activities – such as stored procedures or scripts for data extraction.

The analytical activities are usually performed per requests from sites, departments, clinicians and researchers. These requests are also artifacts associated with analytical activities. The organizational unit that handles such requests is responsible for managing the demand and the **lifecycle** of requests starting from the point when a report was requested, to the point when the report was delivered to the requestor.

All analytical activities, the artifacts, and their lifecycles has a variety of associated **metadata** – any information that describes and qualifies analytical artifacts. Such metadata includes ownership, network drives where reports are stored, reporting metrics, systems where analysis is performed, relationships between artifacts and much more.

The artifacts of the analytical activity and their metadata cumulatively represent analytical assets.

A High-Level View at the Analytical Infrastructure and the Workflow



An Enterprise-grade EHR system typically has a **central operational data store** designed to support high-volume and real-time transaction processing. The data stored in the operational data store is updated in real time by users working with the EHR applications and by the interfaces with external systems and devices. While some real time reporting may be necessary and possible **the operational data store is not optimized for analysis or reporting**. Running complicated, resource-intensive queries against operational data store is likely to adversely affect the performance of the EHR system as a whole.

EHR vendors usually supply a data-warehousing storage optimized for analytical data processing. The data warehouse is refreshed - typically daily - from the operational data store using ETL (Extract, Transform, Load) processes. A data warehousing solution usually includes **data marts** – data and infrastructure layers used to perform specific analytical processing. **Secondary data warehouses** may also exist – for example, to expose data to analytical tools that don't work with the primary storage.

The output of analytical processing is delivered in a variety of ways and stored in different formats. The rough list of analytical artifacts may include:

- Formatted reports delivered as documents (e.g. PDF files). These reports may be stored on the network for subsequent retrieval or may be sent (e.g. emailed) to users directly.
- Raw extracts delivered as CSV files for subsequent post-processing by users (e.g. in Excel spreadsheets)
- Dashboards within EHR system or some third party applications that aggregate multiple reports in organized views.
- OLAP (Online Analytical Processing) results and supporting transient data structures (e.g. cubes, views) within Business Intelligence (BI) systems which in turn may include its own internal dashboards and other artifacts.

Reports and extracts are generated from the data warehouse store using complex sequence of queries and/or stored procedures which often take hours to complete.

The metadata repository in the data warehouse includes information on the contents of the data warehouse, location and their structure information on the processes that take place in the data warehouse back-stage, information on the implicit semantics of data with respect to a common enterprise model, along with any other kind of data that aids in exploiting the information of the warehouse.

Challenges in Managing Analytical Assets

Let's take a closer look at a few of the challenges organizations face when taking on the task of managing analytical assets.

Ever Growing Inventory of Analytical Assets

Over time even a small healthcare organization would accumulate thousands of

analytical assets in different formats and with different delivery mechanisms. It will run dozens to thousands of daily analytical and data extraction processes potentially clogging infrastructure pipes and filling up network shares.

Backlog of Requests for Analytical Processing

The increase in data availability from a modern EHR system leads to increased demand from sites, departments, clinicians and researchers for reports and data analysis. Transition from multiple legacy systems for individual sites to a single EHR system for the entire organization also means moving away from decentralized data management to the centralized one. This transition results a single queue of requests for reporting and analysis and a potentially significant delay in fulfilling the requests if necessary supporting infrastructure and processes are not in place.

In addition to delays in delivering vital insights to the organization, extensive request backlog and lack of supporting infrastructure results in cumulative burden on IT staff responsible for fulfilling the requests.

Excessive Inventory, Unreliable Analytics Due to Poor Data Quality

Poor data quality in EHR systems has many negative implications. From data analytics perspective it impedes quality and performance analysis and affects strategic insights for the institution. From the analytical asset management point of view poor data quality leads to excessive inventory.

Data quality loss happens on many levels:

- before data gets to the EHR system – for example, from external or legacy sources),
- within the EHR system, due to poor configuration and analytical work
- in the data warehouse, for example, due to lack of structural and semantic metadata

Identifying and eliminating duplication and unnecessary similarities in the assets and metadata is a complex process that require extensive **data governance** efforts across the enterprise.

Difficulty Finding and Organizing Analytical Assets

As we pointed out earlier analytical assets are entities of various types that may reside on distributed network shares, located in secondary warehouse data storages or within the proprietary business intelligent systems. In such distributed environment finding and accounting for the artifacts and its metadata without a cataloging solution that reaches beyond the boundaries of individual systems is a nearly impossible task.

EHR Vendors Don't Offer Much Help for Analytical Asset Management

At least for now, EHR vendors are not focused on providing a robust asset management infrastructure. Let's take a look at what Epic Systems – a dominant EHR vendor on the market - has to offer.

Managing Analytical Assets with Epic EHR System

Epic's operational database is Chronicles – a “nosql” data repository based on the InterSystems' Cache. ETL processes from Chronicles usually run daily to export data to Clarity – a relational database that could either run on Oracle or Microsoft SQL Server.

Epic has a built-in report management system - the Reporting Workbench. Reports and some other assets can be generated directly from the Workbench in real time or offline mode and often referred as “operational reports”. These reports are restricted on how much data they will return or how large the report can be. The number of reports available from the Reporting Workbench is measured in thousands.

In addition to Workbench reports organizations on Epic EHR generate formatted reports and extracts directly from Clarity database. Neither these extracts nor extensive metadata associated with them are accessible from the Reporting Workbench.

In an effort to help a healthcare facility maximize the use of Clarity data for clinical and financial use, Epic has developed other tools - such as a data warehouse called Cogito, SAP BusinessObjects (BO) Universes (which makes it easier to select data for reporting in Clarity), and finally BO Web Intelligence (a business intelligence software that allows for the creation of analysis and ad-hoc reports). These tools generate a significant number of additional analytical assets – more reports, but also cubes, dashboards and other types of assets. SAP Business Objects platform is loosely integrated into the Workbench – user can browse external assets. There is no integrated search and discovery across all assets in the Workbench, no option to view and analyze common metadata, no authoring of additional metadata. There is no capability to extend inventory types.

Reporting Workbench has no workflow management capabilities – such as change management, version control, selective publishing.

For demand management of analytical assets institutions may opt for third party CRM tools. Such tools usually don't integrate with the Workbench or SAP BO BI Portal – at least out of the box.

Epic offers multitude of means to perform analytical processing but little support for managing analytical artifacts and metadata. Epic has plans to address at least some of the shortcomings in the future versions.

The Solution - Asset Management Infrastructure

Enterprise Analytical Asset Management Infrastructure is a central repository of information about all analytical assets and a set of tools to support every tasks associated with analytical asset and metadata management.

Such infrastructure should include the following components:

- Searchable asset inventory catalog
- Data extraction processes and configuration management
- Request lifecycle management and self-service capabilities

In addition to the above components this infrastructure provides the means to facilitate Data Governance.

Asset Inventory Catalog

The Inventory Catalog is the cornerstone of the Asset Management infrastructure. It brings together metadata about the assets from all sources, links other components together, and facilitates the workflow for inventory management. The Catalog should include the following functionality:

- Import and re-import inventory metadata from multiple sources
- Automatically add certain metadata to imported assets
- Synchronize metadata that comes from external sources
- Create and manage inventory in the catalog
- Track, approve and monitor changes in the catalog
- Notify Catalog owners about workflow events associated with inventory
- Organize inventory by clinical portfolios and around other clinical content
- Detect duplicates and similarities in data and metadata
- Tag assets with desired identifiers, link to terminology domains
- Intelligent search and ad-hoc queries for inventory using variety of terms and concepts

Job Management and Configuration

Job Management and Configuration component provides users with all necessary means to manage report extraction jobs. It should include the following capabilities:

- Starting and stopping the extraction jobs, running jobs on different server nodes
- Monitoring running job statuses and rating the statuses against historical values (e.g. "slow", "hung")
- Impact analysis - traversing relationships between reports and different elements of the report extraction process, such as tables in the database and stored procedures.
- Analytical insights on report extraction jobs

- Monitor disk space for storing extracted files
- Monitoring the status of ETL processes from the operational database
- Managing and generating manual and automatic notifications about events associated with extraction jobs, such as failed jobs, downtime notifications, low space on network shares.

The Job Management and Configuration is linked to the Reporting Catalog and provides the capabilities to request and display necessary details about reports that are being generated.

Request Lifecycle Manager

Request Lifecycle Manager is a CRM component for tracking requests for reports and analysis and keeping requestors informed about the progress of their requests. Owners can easily find their requests, monitor their progress status, view associated assets and processes and receive notifications relevant to their requests.

The Request Lifecycle Manager reduces communication overhead between the IT staff responsible for analytical infrastructure and the requesting organizational entities and individuals.

Support for Data Governance

In the context of the Analytical Asset Management the data governance effort is about taking account of enterprise content and normalizing it for the purpose of reducing duplication in analytical assets and improving quality of analytics. We believe that this effort includes the following processes:

- Identifying and collecting and tagging concepts and data elements that are used in analytical processing
- Normalizing the content – identifying and creating structural and semantic relationships,
- Checking assets for similarities and duplication; reducing the duplication

Substantial part of the data governance work can be automated. The extent of automation depends on the number of factors, including the level of exposure of the harvestable content in the assets, and the quality of the content. For example, the system can potentially automatically extract data elements from report definitions (e.g. report columns), identify similar assets and metadata, highlight assets and metadata that may require manual review.

Conclusion

Analytical Asset Management is a critical element of the IT infrastructure of any Healthcare institutions that has a modern EHR system. Organizations that are in transition to a centralized EHR system should take in consideration the implications of a significant amount of analytical assets and associated processes on their technical and human IT resources and be prepared to address challenges associated with the management of analytical assets.

About

First Line Software is a premier provider of custom software development, tech enabled services and value-add consulting in big data engineering, digitalization, intellectual integration, industrial Internet and IoT, digital media and marketing, and enterprise content management as well as healthcare IT.

Headquartered in the US, First Line employs 600+ staff globally. First Line team and company culture is centered around subject matter expertise, technical excellence, consulting capabilities and proven methodologies, with a strong focus on Agile and Intellectual Integration.

The company has been recognized with multiple annual rankings and awards by the International Association of Outsourcing Professionals (IAOP), Global Services, CorporateLiveWire, Insights Success, Ventana Research and CNews. We were the first to be awarded with the Scrum Capability Medallion by Scrum, Inc. Most recently, research firm Gartner included FirstLine in their first ever Market Guide for Technology Integrators (2014) and the Cool Vendor in Applications Services 2015 Report. We are active members in Object Management Group and Industrial Internet Consortium. FLS is also an EPiServer Solutions Partner.

First Line ran over 400 projects in 15 countries in 14 different industries. We have global clients across North America, Europe and Australia, including Accenture, Bonnier Group, Clinic to Cloud, Dell Software, InnerWorkings, Partners HealthCare, Odysseus Data Services, Solita, Viastore, and others.

Contacts

USA

+1.877.737.7178

1 Broadway, 14th Floor
Cambridge MA 02142

The Netherlands

+31 (70) 512 1899

Louis Couperusplein 2
2514HP's-Gravenhage
The Hague

The Czech Republic

+420 (2) 28884855

Na Havránce 1508/14
143 00 Praha 12
Czech Republic

sales@firstlinesoftware.com

www.firstlinesoftware.com

www.firstlinesoftware.com