

Ein mandantenfähiges Data-Warehouse-System für die bayerischen Fachhochschulen - Anforderungen, Konzeption und Realisierung

Tim-Oliver T. Förtsch, Stefan Hartmann, Carsten Jürck, Julia Ringler

Wissenschaftliches Institut für Hochschulsoftware
der Universität Bamberg (ihb)
Feldkirchenstr. 21
96045 Bamberg
{tim.foertsch | stefan.hartmann | carsten.juerck | julia.ringler}@ihb.uni-bamberg.de

Abstract: Seit einigen Jahren befindet sich das deutsche Hochschulwesen im Umbruch, was sich aktuell in der Einführung der leistungsorientierten Mittelverteilung oder der Studienbeiträge zeigt. Der Wandel in Richtung New Public Management führt zu einem erhöhten Bedarf an fundierten Informationen über die eigenen Hochschulprozesse. Zur Unterstützung von Entscheidungen haben sich mittlerweile auch im Hochschulumfeld Data-Warehouse-Systeme etabliert. Um dieser Entwicklung gerecht zu werden, wurde im Rahmen des Projekts CEUS ein mandantenfähiges Data-Warehouse-System für die bayerischen Fachhochschulen realisiert. Ziel des neu konzipierten Systems war es, eine Basis für eine standardisierte, flexible Informationsversorgung bei einheitlicher Begriffsdefinition für alle Fachhochschulen zu schaffen. Das entwickelte mandantenfähige Data-Warehouse-System stellt hierfür eine geeignete Lösung dar.

1 Einleitung

Der Bedarf an gestaltungs- und steuerungsrelevanten Informationen ist in den letzten Jahren im Hochschulumfeld stark angestiegen. Hierfür sind u. a. die Festlegung von Zielvereinbarungen und die Überwachung der Zielerreichung oder die leistungs- und belastungsorientierte Mittelverteilung verantwortlich (New Public Management; vgl. z. B. [Ku08]). Eine wesentliche Voraussetzung zur Bewältigung dieser Herausforderungen stellt die Verfügbarkeit von validen und vergleichbaren Daten für das Hochschulmanagement dar.

Das Bayerische Staatsministerium für Wissenschaft, Forschung und Kunst (StMWFK) fördert mit dem Projekt CEUS¹ (Computerbasiertes Entscheidungsunterstützungssystem für die Hochschulen in Bayern) die Entwicklung und den Betrieb von Data-Warehouse-Systemen (DWH-Systemen) zur umfassenden Informationsversorgung aller Entscheidungsbereiche im bayerischen Hochschulwesen. Das wissenschaftliche Institut für Hochschulsoftware der Universität Bamberg (ihb²) ist seit 2003 mit der Einführung und Weiterentwicklung von CEUS am StMWFK und an den staatlichen Hochschulen betraut. Das ihb versteht sich hierbei als Kompetenzzentrum für die Entwicklung und Einführung von Softwarelösungen zur Unterstützung von Hochschulprozessen. Zum Leistungsspektrum des ihb gehören ebenfalls Beratungsdienstleistungen und Schulungsmaßnahmen.

Eine grundlegende Erweiterung erfährt das Projekt CEUS mit der im Jahr 2006 begonnenen Einführung an den 17 bayerischen Fachhochschulen. Dabei galt es einen Lösungsansatz zu entwickeln, der allen Fachhochschulen jeweils ein lokales DWH-System offeriert, das zugleich in die bestehende CEUS-Gesamtarchitektur eingebunden ist. Wichtige Ziele waren dabei auf Seiten des ihb, den Implementierungs- und Wartungsaufwand über alle lokalen DWH-Systeme so weit wie möglich zu reduzieren sowie auf Seiten der Fachhochschulen, die Vergleichbarkeit des Berichtswesens untereinander und – im Hinblick auf die Zielvereinbarungen – mit dem StMWFK sicherzustellen.

Ziel des vorliegenden Beitrags ist es, die Anforderungsanalyse, Konzeption und Realisierung eines mandantenfähigen DWH-Systems für die CEUS-Einführung an den bayerischen Fachhochschulen aufzuzeigen. Hierzu werden in Kapitel 2 Anforderungen an die Architektur eines solchen Systems aufgezeigt. Kapitel 3 umfasst die Konzeption und Kapitel 4 darauf aufbauend die Realisierung des DWH-Systems, welches den beschriebenen Anforderungen genügt. In Kapitel 5 folgen eine Zusammenfassung sowie ein kurzer Erfahrungsbericht über den mit dieser Architektur erzielten Nutzen.

¹ <http://www.ceushb.de>.

² <http://www.ihb.uni-bamberg.de>.

2 Anforderungen an ein DWH-System für die Fachhochschulen

Ausgangspunkt einer Analyse der Anforderungen an ein DWH-System für die bayerischen Fachhochschulen ist die Einbindung der einzelnen Systeme in die CEUS-Gesamtarchitektur. Diese besteht aus einer hierarchisch verteilten Data-Warehouse-Architektur mit autonomen Teilsystemen auf Hochschul- und auf Landesebene (vgl. [Si01]). Die lokalen DWH-Systeme der bayerischen Hochschulen basieren auf Detaildaten zu den Domänen *Studierende und Prüfungen*, *Finanzmittel* sowie *Personal und Stellen*. Das Data-Warehouse auf Landesebene beinhaltet aggregierte, konsolidierte und damit vergleichbare Informationen über alle Hochschulen des Landes. Mit dieser Architektur ist es möglich, sowohl den Informationsbedarf einzelner Hochschulen (unter Berücksichtigung derer individuellen Organisations- und Fächerstrukturen) als auch den Bedarf des StMWFK an hochschulübergreifenden Vergleichswerten zu berücksichtigen. Die DWH-Systeme stellen hiermit eine Dienstleistung für die Hochschuleinrichtungen (z. B. Studiendekane, Fakultäten) bei gleichzeitiger Entlastung der Verwaltung zur Verfügung.

Rückblickend betrachtet führte die sukzessive CEUS-Einführung an den bayerischen Universitäten zu einer Diversifikation der lokalen DWH-Systeme. Von den Universitäten wurde zu Projektbeginn ein eigenständiger Betrieb der lokalen DWH-Systeme vor Ort zwingend gefordert. Diese Systeme basieren grundlegend auf den Konzepten gemäß der CEUS-Referenzarchitektur. Jede Universität hat jedoch entsprechend ihrer individuellen Anforderungen und Spezifika unterschiedliche Schwerpunkte bei der Ausgestaltung und Weiterentwicklung ihres lokalen DWH-Systems gesetzt. Obwohl für die oben genannten Domänen eine landesweit weitgehend homogene operative IT-Landschaft vorliegt, entstanden nennenswerte Unterschiede in den darauf aufbauenden DWH-Systemen. Um den individuellen Informationsbedarf einer Universität bedienen zu können, wurden fortwährend die jeweiligen Schnittstellen zu deren operativen Systemen sowie die daran angeschlossenen ETL-Prozesse (Extract, Transform, Load; vgl. z. B. [BG04, S. 49 ff.]) des lokalen DWH-Systems modifiziert. Dementsprechend wurden auch die zugehörigen Datenbank- und Analyse-Schemata an diese Änderungen oder Erweiterungen angepasst. Der isolierte Betrieb der lokalen DWH-Systeme erforderte zudem die Bereitstellung entsprechend qualifizierter Personen an den einzelnen Universitäten vor Ort. Synergieeffekte konnten daher nur zum Teil genutzt und Ressourcenersparnisse bei Pflege und Wartung der Systeme lediglich in geringem Umfang erzielt werden. Die eigenständige Verwaltung der lokalen DWH-Systeme weist zudem Nachteile insbesondere bei der Implementierung allgemeingültiger Anforderungen im Berichtswesen auf (z. B. landesweite Berichtsanforderungen vom StMWFK). Jedes DWH-System muss hierzu für sich angepasst und erweitert werden. Der parallele Betrieb von isolierten DWH-Systemen für einen einzigen Aufgabenbereich führte im Projektverlauf letztlich zu ähnlichen Aufwänden, wie der Betrieb von DWH-Systemen unterschiedlicher Aufgabenbereiche.

In einer weiteren Ausbaustufe werden seit 2006 die bayerischen Fachhochschulen in die landesweite CEUS-Gesamtarchitektur mit eingebunden. Den einzelnen Fachhochschulen soll hierbei jeweils ein lokales DWH-System, zunächst mit Detaildaten zu Studierenden und Prüfungen, zur Verfügung gestellt werden. Dabei sollen Synergieeffekte bei der Konzeption sowie beim Betrieb und der Weiterentwicklung dieser DWH-Systeme bestmöglich genutzt werden. Die bei den bayerischen Universitäten verfolgte Strategie, eigenständige DWH-Systeme anhand eines Referenzmodells zu realisieren, erschien bei den Fachhochschulen aus wirtschaftlichen Gründen (höhere Anzahl, durchschnittliche Größe der Fachhochschulen etc.) und den genannten Erfahrungen als nicht adäquat.

Aufgrund der oben geschilderten Rahmenbedingungen ergeben sich die nachfolgend dargestellten Anforderungen des IHB an die lokalen DWH-Systeme der Fachhochschulen:

- *Standardisierte Schnittstelle:* Die Datenextraktion soll durch die Fachhochschulen gemäß einer einheitlichen, verbindlichen Schnittstelle erfolgen. Hierdurch soll Unabhängigkeit von den jeweiligen operativen Systemen erreicht werden. Auf diese Weise wird ebenfalls gewährleistet, dass personenbezogene Daten im Vertrauensbereich der Fachhochschulen extrahiert und vor der Übergabe an die DWH-Systeme eliminiert oder anonymisiert werden. Dies erlaubt die Einhaltung datenschutzrechtlicher Vorgaben. In einem DWH-System ist der Datenschutzproblematik besondere Aufmerksamkeit zu widmen, da die dort angebotenen Navigationsoperatoren auf einer umfassenden Datenbasis vielfältige Kombinationsmöglichkeiten von Einzeldaten ermöglichen (vgl. z. B. [Mö99], [SPU02]).
- *Einheitliches Datenbankschema:* Als Voraussetzung für einen einheitlichen ETL-Prozess soll ein einheitliches Datenbankschema konzipiert werden, das die Informationsbedarfe aller Fachhochschulen erfüllt.
- *Einheitlicher ETL-Prozess:* Die zu entwickelnden lokalen DWH-Systeme sollen über einen einheitlichen ETL-Prozess versorgt werden, da der Betrieb und die Wartung von 17 eigenständigen Ladeprozessen im Rahmen der verfügbaren personellen Ressourcen nicht effizient durchführbar sind. Auch bei einer sukzessiven Einführung der einzelnen Systeme kann auf diese Weise ein homogener ETL-Prozess nachhaltig gewährleistet werden. Anpassungen und Weiterentwicklungen sind entsprechend nur einmalig an einer Stelle durchzuführen.
- *Zentrale Steuerung des ETL-Prozesses:* Der einheitliche ETL-Prozess soll von zentraler Stelle administriert und gesteuert werden können. Die Steuerung des Prozesses ist durch eine Parametrisierung zu ergänzen, um die DWH-Systeme einzeln, domänenspezifisch und zeitspezifisch laden zu können.

- *Standardisiertes Berechtigungskonzept:* Das Berechtigungskonzept umfasst eine genaue Definition der Zugriffsberechtigungen auf die Datenbasis des DWH-Systems. Um dessen Pflege zu erleichtern, sollen die Berechtigungen mittels eines für die einzelnen Fachhochschulen gleichartig definierten Rollenkonzepts verwaltet werden.

In Zusammenarbeit mit den Fachhochschulen kommen darüber hinaus folgende Anforderungen an die Konzeption des DWH-Systems hinzu:

- *Hochschulübergreifendes Berichtswesen:* Um die Vergleichbarkeit des Berichtswesens im Hinblick auf die Zielvereinbarungen mit dem StMWFK sicherzustellen, soll ein fachhochschulübergreifendes Standardberichtswesen geschaffen werden. Dies erfordert, die Semantik der Kennzahlen und Merkmale für alle Fachhochschulen eindeutig festzulegen.
- *Hochschulspezifisches Berichtswesen:* Neben dem hochschulübergreifenden Standardberichtswesen sollen auch spezifische Anforderungen berücksichtigt werden können, um den individuellen Informationsbedarf der Fachhochschulen zu decken.
- *Minimierung der Implementierungs- und Betriebskosten:* Der Implementierungs-, Wartungs- und Pflegeaufwand ist auf Seiten der Fachhochschulen möglichst gering zu halten. Der Betrieb der DWH-Systeme soll daher nicht an jeder Fachhochschule separat erfolgen, da hierfür redundant Hardware, Software und Personalressourcen vorgehalten werden müssen.

Im folgenden Kapitel wird – auf Grundlage der genannten Anforderungen – die vom ihb entwickelte Architektur für die lokalen DWH-Systeme der bayerischen Fachhochschulen vorgestellt.

3 Konzeption eines mandantenfähigen DWH-Systems

Mit Bezug zu den vorausgegangenen Ausführungen wird nachfolgend die Architektur des entwickelten DWH-Systems vorgestellt. Aufgrund der in Kapitel 2 genannten Anforderungen stellten sich eigenständige DWH-Systeme für die Fachhochschulen als ungeeignet heraus. Aus diesem Grund wurde nach Konzepten für eine zentrale Architektur gesucht. Das in der Praxis bei transaktionalen Systemen bewährte Prinzip der Mandantenfähigkeit lässt sich auf die Konzeption eines DWH-Systems für die Fachhochschulen übertragen. Im Folgenden wird zunächst die Grundidee eines Mandantensystems eingeführt und dessen Eignung für die Umsetzung der hier vorliegenden, spezifischen Anforderungen aufgezeigt. Anschließend wird der Begriff eines mandantenfähigen DWH-Systems definiert und die vom ihb entwickelte mandantenfähige DWH-Architektur vorgestellt.

Nach *Schulze* wird als Mandantensystem Software bezeichnet, die so konzipiert ist, dass gleichartige Arbeiten für unterschiedliche Mandanten durchgeführt werden können [Sc89, S. 1843]. Im SAP-Umfeld kommt der Begriff des Mandanten dem Begriff eines Konzerns nahe. Der Mandant im System R/3 bildet eine logische Gruppierung von handelsrechtlichen, organisatorischen und administrativen Einheiten und ist damit die höchste Organisationseinheit [EGN96, S. 45]. Auf diese Weise lassen sich in einem gemeinsamen System mehrere eigenständige Unternehmen getrennt voneinander abbilden. Die datentechnische Abtrennung wird hierbei durch die Aufnahme des Mandanten als Schlüssel in die Tabellen des R/3-Repositorys erreicht [Wi98, S. 26]. Im Allgemeinen kann in einem Mandantensystem zwischen mandantenabhängigen und mandantenunabhängigen Strukturen unterschieden werden. Mandantenabhängige Strukturen werden für jeden Mandanten individuell gestaltet. Mandantenunabhängige Strukturen sind übergreifend für alle am System Beteiligten verfügbar.

Der Begriff des Mandanten und die in diesem Zusammenhang beschriebene datentechnische Abtrennung lassen sich auf die Fachhochschulen übertragen, die weitgehend eigenständig und autonom agieren. Die Grundidee eines Mandantensystems wird im Folgenden für die Konzeption der lokalen DWH-Systeme aufgegriffen und auf die Spezifika eines DWH-Systems im Hochschulumfeld ausgerichtet.

Im Rahmen dieser Arbeit wird Mandantenfähigkeit eines DWH-Systems folgendermaßen definiert:

Ein mandantenfähiges Data-Warehouse-System ist ein Data-Warehouse-System, das für eigenständige Organisationseinheiten entscheidungsrelevante Daten aus unterschiedlichen Datenquellen über einen einheitlichen ETL-Prozess in einer konsolidierten Datenbasis bereitstellt und das darauf aufbauende multidimensionale Schema sowohl mandantenabhängige als auch mandantenunabhängige Strukturen enthält.

An dieser Stelle ist darauf hinzuweisen, dass zum einen die Quellsysteme der beteiligten Mandanten einem logischen Aufgabenbereich entstammen sollten. Die Integration stark heterogener Quellsysteme im Sinne der abgebildeten Diskurswelt führt zu einer hohen Anzahl mandantenabhängiger Strukturen und würde die Vorteile eines mandantenfähigen DWH-Systems egalisieren. Zum anderen ist für dessen Einführung eine hohe Übereinstimmung bzgl. der Analyseanforderungen der einzelnen Mandanten (z. B. ein weitgehend homogenes Merkmals- und Kennzahlensystem) eine entscheidende Voraussetzung.

Die logische Architektur eines DWH-Systems lässt sich in die Ebenen *Datenerfassung*, *Datenhaltung* und *Datenbereitstellung* gliedern (vgl. z. B. [BU00, S. 17 ff.]). Die Datenerfassungsebene stellt Werkzeuge zur Extraktion der Daten aus den operativen Systemen sowie zur Datenbereinigung zur Verfügung. Das eigentliche Data-Warehouse auf der Datenhaltungsebene bildet eine konsolidierte Datenbasis. Die Datenbereitstellungsebene besteht aus einem OLAP-Server, der die multidimensional aufbereiteten Daten den Präsentationswerkzeugen zur Verfügung stellt (vgl. z. B. [BU00, S. 2 ff.]). Im Folgenden wird das vom ihm entwickelte mandantenfähige DWH-System entlang dieser Ebenen erläutert. Hierbei werden insbesondere die Anpassungen bzw. Erweiterungen im Vergleich zu einem nicht mandantenfähigen Konzept aufgezeigt.

Datenerfassungsebene

Auf der Datenerfassungsebene ist eine einheitliche Schnittstellenbeschreibung für die Anbindung der operativen Systeme zu erarbeiten. Diese bildet den Ausgangspunkt für einen gemeinsamen Ladeprozess aller Mandanten. Die zu definierende Schnittstelle muss zum einen die informationellen Anforderungen aller beteiligten Mandanten abdecken und zum anderen die Bedeutung der einzelnen Elemente der Schnittstelle eindeutig bestimmen. Durch sie werden die Extraktion sowie eine Teilaufgabe der Homogenisierung der Daten an den Datenlieferanten ausgelagert. Die Schnittstellendefinition ist für alle Mandanten verpflichtend. Sie umfasst mandantenübergreifende Elemente, die von allen Mandanten gefüllt werden, sowie Felder, die individuelle Erweiterungen darstellen und nicht von allen Mandanten bedient werden müssen.

Die Datenextraktion findet im Vertrauensbereich der einzelnen Mandanten statt. Diese bestimmen selbstständig den Zeitpunkt sowie die Häufigkeit der Extraktion aus den eigenen operativen Systemen und sind für die Qualität der Daten verantwortlich. Neben der Datenextraktion erfolgt eine Datentransformation entsprechend der Konventionen der Schnittstellenbeschreibung. Hierbei werden u. a. auch mandantenspezifische Identifikationsnummern hinzugefügt und alle personenbezogenen Angaben eliminiert oder anonymisiert. Die verschlüsselten Daten werden im Anschluss über eine gesicherte Verbindung per Push-Prinzip an einen zentralen Upload-Server (vgl. Abb. 1) übertragen.

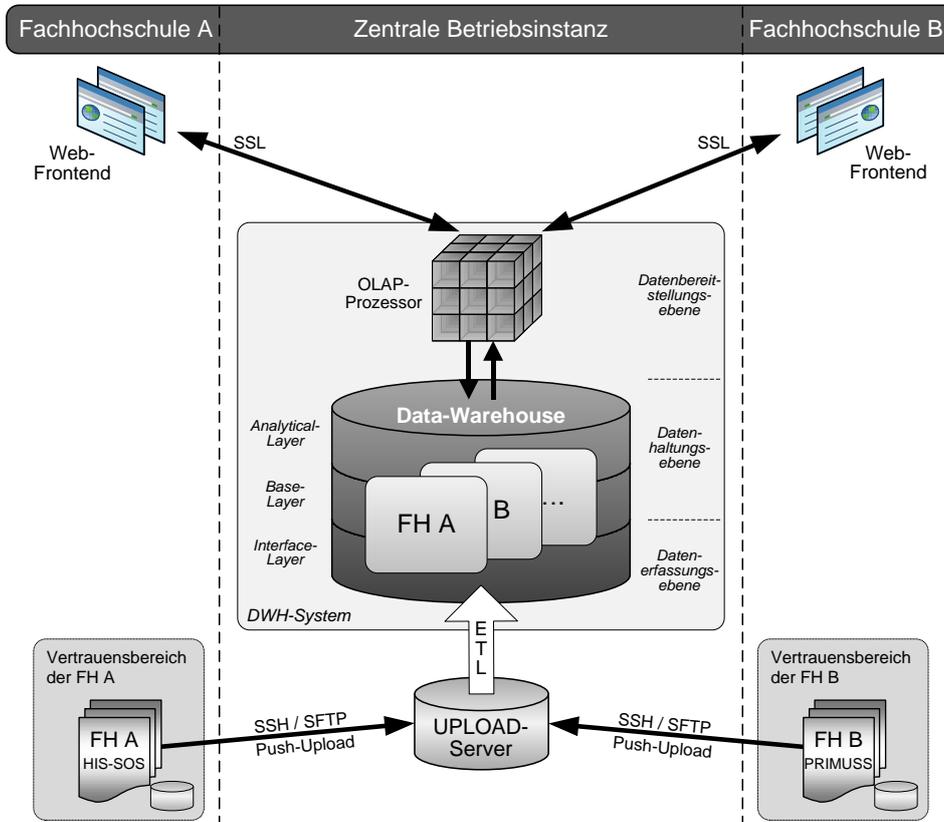


Abbildung 1: Architektur des mandantenfähigen DWH-Systems für die bayerischen Fachhochschulen

Über einen zentralen ETL-Prozess werden die von den einzelnen Mandanten gemäß der Schnittstellenbeschreibung bereitgestellten Daten in das Data-Warehouse geladen. Durch die Vorgabe einer einheitlichen und verbindlichen Schnittstelle wird der Ladeprozess erheblich vereinfacht, da ein Teil der syntaktischen und semantischen Homogenisierung der Daten bereits im Verantwortungsbereich des Mandanten erfolgt. Die notwendigen Schritte zur Datentransformation sind zudem nur einmal zu spezifizieren und gelten für alle Mandanten gleichermaßen. Der ETL-Prozess ist darüber hinaus parametrisierbar, um abhängig von Mandant, Domäne und Zeitpunkt in unterschiedlichen Varianten ausgeführt werden zu können. Für die Administration steht ein eigener Steuerungs- und Überwachungsprozess zur Verfügung. Dieser regelt die jeweilige Kombination der oben genannten Parameter.

Datenhaltungsebene

Die Datenhaltungsebene des mandantenfähigen DWH-Systems ist in drei Ebenen gegliedert (vgl. z. B. [Si01, S. 12 ff.]). Der Interface-Layer ist als temporärer Speicher konzipiert, dessen Datenstruktur bei der vorliegenden Architektur exakt der Schnittstellendefinition entspricht. Eine konsistente, normalisierte Datenhaltungsschicht bildet der Base-Layer, der die konsolidierten und historisierten Daten permanent speichert. Der Analytical-Layer ist für die Bereitstellung der Daten in der von den jeweiligen Anfrage-Werkzeugen benötigten Form vorgesehen. Im Gegensatz zur Konzeption mehrerer eigenständiger DWH-Systeme sind bei einem mandantenfähigen Ansatz diese Ebenen mit jeweils einem mandantenübergreifenden Datenschema nur einmal vorhanden.

Ein Datenschema besteht hier grundsätzlich aus mandantenabhängigen und mandantenunabhängigen Schemaobjekten. Beispielsweise ist das *Semester* fachhochschulübergreifend definiert und damit mandantenunabhängig. Beim *Studiengang* hingegen handelt es sich um ein mandantenabhängiges Objekt, da an jeder Fachhochschule andere Studiengänge angeboten werden. Unabhängig davon, ob eine einzelne Fachhochschule ein bestimmtes Schemaobjekt verwendet oder nicht, steht jedes allen Mandanten gleichermaßen zur Verfügung. So wird der *Schwerpunkt* nur von einem Teil der Fachhochschulen für Auswertungen benötigt und daher auch nur von diesen Hochschulen mit Daten versorgt. Um die Daten später mandantenspezifisch auswerten zu können, wird der Mandant als Teil des Schlüssels in alle mandantenabhängigen Schemaobjekte aufgenommen. Über diese Schemaobjekte können somit bis zu einem gewissen Grad individuelle Anforderungen abgebildet werden. Des Weiteren stehen durch das gemeinsame Datenschema Erweiterungen in der Implementierung allen am System Beteiligten gleichzeitig zur Verfügung, wodurch sich die Pflege und Wartung des Systems vereinfacht.

Datenbereitstellungsebene

Die Datenbereitstellungsebene bildet die Schnittstelle zu den Endanwendern. Ein OLAP-Server stellt die multidimensionalen Strukturen zur Verfügung und ermöglicht den interaktiven Zugriff auf das Data-Warehouse. Der Zugriff auf das Analyseschema wird für alle Mandanten über ein einheitliches Berechtigungskonzept realisiert. Dazu sieht das Analyseschema eine Dimension zur Aufnahme des Mandanten vor. Diese Dimension wird im Berechtigungskonzept genutzt um die datentechnische Abtrennung der einzelnen Mandanten durch entsprechende Filterkriterien zu garantieren. Hierdurch sind mandantenspezifische Inhalte nur für den jeweiligen Mandanten verfügbar.

Das Analyseschema beruht auf einem multidimensionalen Datenschema. Dieses umfasst sowohl mandantenabhängige als auch mandantenunabhängige Dimensionen und Kennzahlen. Generell stehen die Dimensionen und Kennzahlen für alle Mandanten gleichermaßen zur Verfügung. Per Berechtigungskonzept kann einem Mandanten jeweils nur auf die für ihn relevanten Dimensionen und Kennzahlen Zugriff gegeben werden. Insgesamt ergibt sich die Notwendigkeit in Bezug auf Benennung, Definition und Abhängigkeiten der Schemaobjekte einen Konsens zwischen den Mandanten zu erarbeiten. Dies genügt dem Anspruch, die Vergleichbarkeit der Schemaobjekte sicherzustellen. Zudem bildet das gemeinsame Analyseschema die Grundlage für den Aufbau eines standardisierten Berichtswesens.

4 Realisierung des mandantenfähigen DWH-Systems

Im Jahr 2006 begann das ibh mit der Entwicklung des mandantenfähigen Prototyps für die Domäne *Studierenden und Prüfungen*. Wichtigstes Ziel des ibh war dabei, den Implementierungs- und Wartungsaufwand für die sukzessive Einführung an den 17 bayerischen Fachhochschulen so gering wie möglich zu halten. Das umgesetzte mandantenfähige Konzept erforderte initial einen relativ hohen Entwicklungsaufwand. Der anschließende Aufwand für die Realisierung der einzelnen Mandanten ist jedoch im Vergleich zu einem nicht-mandantenfähigen Konzept deutlich geringer (vgl. Abb. 2).

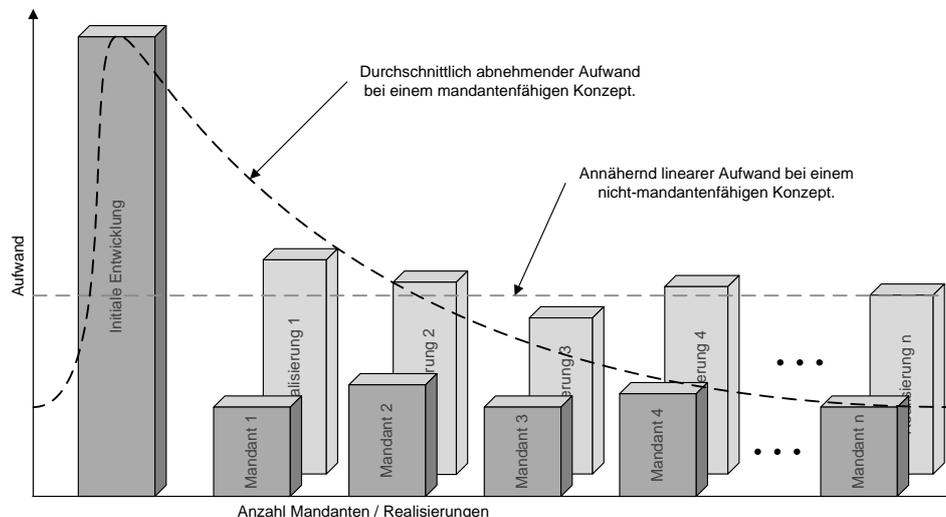


Abbildung 2: Gegenüberstellung der Realisierungsaufwände für ein DWH-System nach einem mandantenfähigen und einem nicht-mandantenfähigen Konzept

Basierend auf dem im vorangegangenen Kapitel erarbeiteten Konzept eines mandantenfähigen DWH-Systems wird nachfolgend dessen Realisierung für die lokalen DWH-Systeme der bayerischen Fachhochschulen vorgestellt. Die zentrale Betriebsinstanz (vgl. Abb. 1) wird derzeit durch das ihb bereitgestellt. Diese besteht aus einer gemeinsamen Hardware-Architektur für alle Fachhochschulen. Die Datenlieferungen erfolgen über einen zentralen Upload-Server. Für die gesicherte Übertragung wird hierbei das SSH-Protokoll genutzt. Als Datenbankverwaltungssystem kommt der Microsoft SQL Server 2005 zum Einsatz. Bei der Implementierung des ETL-Prozesses wird auf die SQL Server Integration Services (SSIS) zurückgegriffen. Die Datenbereitstellung wird mit Hilfe des MicroStrategy Intelligence Servers realisiert. Der Zugriff auf das Berichtswesen findet dabei ausschließlich über das MicroStrategy Web-Frontend (SSL-verschlüsselt) statt.

Die nachstehenden Ausführungen greifen die Unterteilung in Datenerfassungs-, Datenerhaltung- und Datenbereitstellungsebene wieder auf.

Datenerfassungsebene

Die Verwaltung der Studierenden- und Prüfungsdaten erfolgt bei den bayerischen Fachhochschulen entweder mithilfe des Anwendungsprogramms HIS-SOS/POS oder PRIMUSS. Die den beiden Systemen zugrunde liegenden Datenstrukturen unterscheiden sich sowohl auf syntaktischer als auch semantischer Ebene. Um die konzeptionell vorgesehene einheitliche Schnittstelle zu ermöglichen und einen gemeinsamen ETL-Prozess realisieren zu können, wurden auf der Datenerfassungsebene folgende Vorbereitungen getroffen:

- Die Heterogenität der Datenstrukturen von HIS-SOS/POS und PRIMUSS wurde bei der Analyse der jeweiligen Datenschemata hinsichtlich Gemeinsamkeiten und Differenzen aufgelöst. Basierend auf diesen Ergebnissen wurde die für beide Vorkomponenten einheitliche und verbindliche Schnittstellenbeschreibung festgelegt. Diese enthält neben allgemeinen Anweisungen zur Datenextraktion und -lieferung für jede Satzart die exakte Struktur und das genaue Format, in dem die Daten bereitzustellen sind. In jeder Satzart ist zudem ein Feld vorgesehen, um die jeweilige Fachhochschule (über die Hochschulnummer) zu identifizieren.
- Die Einhaltung datenschutzrechtlicher Vorgaben kann aufgrund der detaillierten Schnittstellenbeschreibung durch den örtlichen Datenschutzbeauftragten gut überprüft werden. Felder, die direkt personenbezogene Daten enthalten (z. B. Name, Adresse, Telefonnummer), sind in der für die Fachhochschulen entwickelten Satzartenbeschreibung nicht enthalten. Identifizierende Merkmale (z. B. Matrikelnummer), die zur korrekten Verknüpfung der Daten aus unterschiedlichen Satzarten erforderlich sind, werden u. a. mittels eines vom ihb bereitgestellten Werkzeugs verschlüsselt (z. B. RSA-Algorithmus).

- Jede Fachhochschule stellt ihre Studierenden- und Prüfungsdaten gemäß der Satzartenbeschreibung als so genannte Flatfiles auf dem zentralen Upload-Server zur Verfügung (Push-Prinzip). Auf diese Weise bestimmen die Fachhochschulen zugleich die Aktualisierungsstrategie für ihre Daten im Data-Warehouse.
- Bei der Ausgestaltung des für alle Fachhochschulen im Wesentlichen einheitlichen ETL-Prozesses galt es, neben den allgemeingültigen Übertragungsregeln auch einige wenige logische und semantische Besonderheiten einzelner Mandanten zu berücksichtigen. Bei mandantenspezifischen Daten werden daher bei jedem Transformationsschritt der jeweilige Mandant und dessen Vorsystem als Parameter mit eingebunden. Auf diese Weise können Spezifika (z. B. unterschiedliche Zuordnung von gemeinsamen Schwerpunktfächern zu Studiengängen) explizit behandelt und auf das gemeinsame Datenschema abgebildet werden.
- Um den über SSIS-Pakete abgebildeten ETL-Prozess parametrisieren zu können, wurden Tabellen zur Steuerung des gesamten Ladeprozesses erstellt. Diese beinhalten Angaben zu Fachhochschule, Domäne (z. B. Studierenden-daten) sowie Zeitraum der zu ladenden Daten. Die Protokollierung des Ladeprozesses wurde dabei so realisiert, dass anhand eines Identifiers in der Steuerungstabelle die einzelnen Ladeschritte sowie ggf. aufgetretene (Daten-)Fehler automatisiert aufgelistet und über Fehlerlisten per E-Mail an die Administratoren versendet werden.
- Die Transformations- und Ladeprozesse für Interface-, Base- und Analytical-Layer können individuell angesteuert werden. Ein übergeordnetes SSIS-Paket erlaubt den durchgängigen Ablauf des gesamten ETL-Prozesses.

Datenhaltungsebene

Zur Realisierung der Datenhaltungsebene wurde für die Interface-, Base- und Analytical-Layer jeweils ein Schema angelegt. Diese dienen der Speicherung der Daten aller Fachhochschulen. Im Folgenden werden die zur Realisierung der Mandantenfähigkeit notwendigen Anpassungen aufgezeigt (vgl. Abb. 3):

- Das Schema des Interface-Layer entspricht genau den Strukturen der Satzarten. Die von den Fachhochschulen bereitgestellten Daten werden durch die SSIS-Pakete ohne Transformation in den Interface-Layer übertragen. Dabei stellt eine Tabelle einen temporären Speicher für die Daten einer Satzart aller beteiligten Fachhochschulen dar. Das Kennzeichen Hochschulnummer dient der Unterscheidung der mandantenspezifischen Daten.

- In allen Tabellen des Base-Layer mit mandantenspezifischen Daten wurde die Hochschulnummer als Teil des identifizierenden Schlüssels aufgenommen. Dadurch können die Daten einer Hochschule einzeln und unabhängig von anderen Hochschulen verwaltet werden. In Kombination mit den in allen Schlüsselstabellen vorgesehenen Zeitstempeln, kann darüber hinaus eine individuelle Historisierung der Datensätze realisiert werden. Zur Durchführung der korrekten, mandantenspezifischen Historisierung sowie der Datentransformation wurden für den Übergang von Interface- auf Base-Layer entsprechende Sichten angelegt, welche die vollständige Historisierungs- bzw. Transformationslogik abbilden.
- Analog zum Base-Layer führen auch im Analytical-Layer alle Fakttabellen und mandantenabhängigen Dimensionstabellen das Attribut Mandant mit sich. Dies bildet einen elementaren Baustein für das Berechtigungskonzept, um einer Fachhochschule exklusiven Zugriff auf ihre Stamm- und Bewegungsdaten zu gewährleisten.
- Zur Verarbeitung der Daten durch den auf dem ROLAP-Prinzip (vgl. z. B. [BU00, S. 27 f.]) basierenden Intelligence Server von MicroStrategy wurde dem Analytical-Layer ein denormalisiertes Snowflake-Schema zugrunde gelegt.

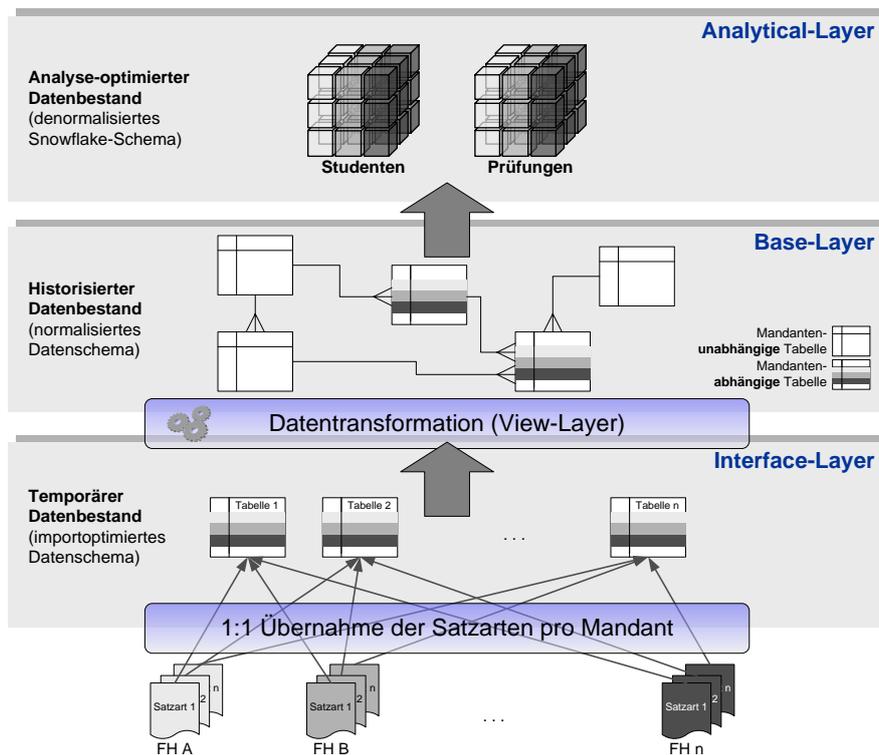


Abbildung 3: Logische Architektur des mandantenfähigen Data-Warehouse

Aufgrund der Implementierung eines zentralen, mandantenfähigen DWH-Systems, können die Aufgaben zur Administration des Datenbankverwaltungssystems für alle Fachhochschulen einheitlich und zeitgleich erfüllt werden (z. B. Verwaltung der Indizes, Datenbanksicherungen, Archivierung der Datenbestände).

Datenbereitstellungsebene

Der auf der Datenbereitstellungsebene durch den MicroStrategy Intelligence Server unterstützte Analyseprozess operiert auf den Daten des Analytical-Layer. Die Daten werden den Endanwendern über fachbezogene Schemaobjekte präsentiert. Die Komplexität der Datenstrukturen eines Data-Warehouse bleibt daher für den Endanwender hinter der semantischen Schicht der Terminologie verborgen [LJ99, S. 7-1 f.]. Die Besonderheiten der mandantenspezifischen Implementierung der Datenbereitstellungsebene werden nachstehend erörtert:

- Analog zur Datenhaltungsebene musste ein mandantenspezifischer Zugriff auf der Datenbereitstellungsebene realisiert werden. Im Analyseschema wurde daher ein Schemaobjekt zur Aufnahme des Mandanten angelegt. Jeder hochschulspezifischen Anwendergruppe wird dieses Schemaobjekt mit der jeweils entsprechenden Ausprägung (hier die Hochschulnummer) als Sicherheitsfilter hinterlegt. Anwender sind jeweils nur einer Anwendergruppe zugeordnet. Auf diese Weise wird gewährleistet, dass nur die für eine Fachhochschule relevanten Schemaobjekte zur Datenanalyse bereitgestellt und nur die mandanteneigenen Daten in einem Bericht ausgegeben werden.
- Um das Berechtigungskonzept leicht administrieren zu können, wurden fachhochschulübergreifend gültige Zugriffsrollen (Standard-, Advanced-, und Power-User) entworfen. Diese unterscheiden sich hinsichtlich der auf den Datenbeständen erlaubten Operationen (von der einfachen Berichtsausführung bis hin zur eigenständigen Berichtsdefinition). Jedem Endanwender wird entsprechend seiner Profildimension (vgl. [HM07, S. 582]) eine Rolle zugewiesen.
- Aufbauend auf der für alle Fachhochschulen einheitlichen Definition der Dimensionen und bestimmter Kennzahlen, erfolgte die Ausgestaltung eines konsolidierten, übergreifenden Standardberichtswesens. Der Aufbau und die Semantik der von allen Fachhochschulen benötigten Berichte wurden in einer Arbeitsgruppe, bestehend aus Fachverantwortlichen der einzelnen Fachhochschulen, gemeinsam definiert und anschließend vom ihb implementiert. Zur Ausführungszeit eines Berichts werden über den Sicherheitsfilter des Endanwenders die jeweils zugehörigen Dateninhalte ermittelt. Die identischen Berichtsstrukturen erlauben auf diese Weise eine Vergleichbarkeit der fachhochschulspezifischen Berichtsergebnisse.

- Um die individuellen Informationsbedarfe vollständig zu erfüllen, kann jede Fachhochschule zudem eigene Berichte erstellen. Die fachlichen Gegebenheiten einer Fachhochschule werden durch die mandantenspezifischen Anteile in der Behandlung der Daten auf Datenerfassungs-, Datenhaltungs- und Datenbereitstellungsebene abgebildet. Neben einem standardisierten Berichtswesen steht damit jeder Fachhochschule eine flexible Informationsversorgung zum Aufbau eines personalisierten Informationsportals zur Verfügung.

Die gewonnenen Erkenntnisse im Rahmen der Realisierung des mandantenfähigen DWH-Systems für die Fachhochschulen legen offen, dass ein entscheidendes Kriterium bei der Ausgestaltung eines mandantenfähigen DWH-Systems in der Differenzierung zwischen mandantenabhängigen und mandantenunabhängigen Schemaobjekten liegt. Trotz des bei den Fachhochschulen vorliegenden gleichen Domänenbezugs, entstand während der Konzeptionsphase ein nicht unerheblicher Aufwand bei der Bewertung und Definition der Stammdaten als mandantenabhängig bzw. mandantenunabhängig.

Mit steigender Anzahl von Schemaobjekten, die nur von einem Teil der Mandanten verwendet werden, sinkt der Mehrwert, der mit einem derartigen DWH-Konzept erzielbar ist. Davon abhängig ist zu entscheiden, ob die Implementierung eines mandantenfähigen DWH-Systems unter betriebswirtschaftlichen Gesichtspunkten sinnvoll ist. Grundsätzlich besteht hier ein Trade-off zwischen der Berücksichtigung individueller Anforderungen der Mandanten und der insgesamt potentiell realisierbaren Betriebskostensparnis.

5 Zusammenfassung

Der vorliegende Beitrag erläutert die Anforderungen an die Architektur der lokalen DWH-Systeme für die bayerischen Fachhochschulen. Als Lösung wurde ein mandantenfähiges DWH-System mit speziellen Architekturmerkmalen vorgestellt, das am ihb konzipiert und entwickelt wurde. Zu den wesentlichen Aufgaben des ihb bei der Einführung der lokalen DWH-Systeme zählten die Unterstützung der Fachhochschulen bei der Bereitstellung der Daten gemäß der Satzartenbeschreibung, die Prüfung der gelieferten Daten auf Konsistenz und Plausibilität, das Laden der Daten sowie die Rückmeldung über aufgetretene Datenfehler, die Anpassung und Weiterentwicklung des MicroStrategy-Schemas und die Unterstützung bei der Berichtsdefinition. Das mandantenfähige DWH-System befindet sich gegenwärtig für fünf Fachhochschulen im Betrieb. Die sukzessive Einbindung weiterer Fachhochschulen erfolgt parallel.

Die Mandantenfähigkeit des DWH-Systems wirkt sich insbesondere vorteilhaft durch den reduzierten Aufwand für Pflege und Wartung sowie die gemeinsame Nutzung von technischen und personellen Ressourcen für alle Fachhochschulen aus. Die Architektur ermöglicht zudem eine leichte Einbindung weiterer Fachhochschulen. Dies wird insbesondere durch den gemeinsamen Ladevorgang (ETL-Prozess) und das gemeinsame Datenbankschema erreicht. Nachteilig zeigte sich in der Konzeptionsphase vor allem der notwendige Abstimmungsaufwand zwischen den Fachhochschulen. Darüber hinaus sind zukünftige Änderungen zwischen allen Mandanten zu koordinieren.

Die bayerischen Fachhochschulen erhalten durch das vorgestellte mandantenfähige DWH-System ein zukunftssicheres und erweiterbares Anwendungssystem, mit dem auf spezifische Informationsbedarfe individuell reagiert werden kann. Es steht eine standardisierte, flexible Informationsversorgung durch ein umfangreiches Berichtswesen (hochschulübergreifend und hochschulspezifisch) zur Verfügung. Dies ermöglicht eine gezielte und schnelle Informationsrecherche bei einheitlicher Begriffsdefinition sowie eine Vergleichbarkeit der Berichte. Hierdurch konnte bei den Fachhochschulen in wenigen Monaten eine hohe Akzeptanz des Systems CEUS erzielt werden. Dies zeigt sich auch anhand folgender Punkte: Die zunächst prototypische Implementierung konnte erfolgreich in den Systembetrieb überführt werden. Des Weiteren wurde das System innerhalb von zwei Monaten drei weiteren Fachhochschulen zur Verfügung gestellt. Auch der Nutzerkreis innerhalb einer Fachhochschule wird permanent erweitert (z. B. Roll-out an den Fakultäten). Zudem haben bereits mehrere Fachhochschulen ein starkes Interesse an einer Einbindung in die mandantenfähige DWH-Lösung von CEUS bekundet.

Literaturverzeichnis

- [BG04] Bauer, A.; Günzel, H.: Data-Warehouse-Systeme - Architektur, Entwicklung, Anwendung. dpunkt Verlag, Heidelberg, 2004.
- [BU00] Böhnlein, M.; Ulbrich-vom Ende, A.: Grundlagen des Data Warehousing: Modellierung und Architektur. Bamberger Beiträge zur Wirtschaftsinformatik Nr. 55, Bamberg, 2000.
- [EGN96] Engels, A.; Gresch, J.; Nottenkämper, N.: SAP R/3 kompakt – Einführung und Arbeitsbuch für die Praxis. Tewi Verlag, München, 1996.
- [HM07] Henrich, A.; Morgenroth, K.: Kontextbasiertes Information Retrieval zur Unterstützung von Entwicklungsprozessen. In (Oberweis, A.; Weinhardt, C. et al. Hrsg.): eOrganisation - Service-, Prozess-, Market-Engineering: 8. Internationale Tagung Wirtschaftsinformatik - Band 2. Universitätsverlag, Karlsruhe, 2007; S.575-592.
- [Ku08] Kurz, D.: New Public Management und Verwaltungskomplexität. In (von Maravic, P.; Priddat, B.P. Hrsg.): Öffentlich – Privat: Verwaltung als Schnittstellenmanagement. Metropolis Verlag, Marburg, 2008; S. 211-232.
- [LJ99] Lehmann, P.; Jaszewski, J.: Business Terms as a Critical Success Factor for Data Warehousing. In: Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'99), Heidelberg, 1999; S.7-1 - 7-5.
- [Mö99] Möncke, U.: Sicherheit im Data Warehouse – Profilbildung und Anonymität. In (Horster, P.; Fox, D. Hrsg.): Datenschutz und Datensicherheit – Konzepte, Realisierungen, Rechtliche Aspekte, Anwendungen. Vieweg, Braunschweig/Wiesbaden 1999, S. 30-59.
- [Sc89] Schulze, H.H.: Computer Enzyklopädie. Rowohlt Taschenbuch Verlag, Hamburg, 1989.
- [Si01] Sinz, E.J. et. al.: Architekturkonzept eines verteilten Data-Warehouse-Systems für das Hochschulwesen. Bamberger Beiträge zur Wirtschaftsinformatik Nr. 59, Bamberg, 2001.
- [SPU02] Sinz E.J., Plaha M., Ulbrich-vom Ende A.: Datenschutz und Datensicherheit in einem landesweiten Data-Warehouse-System für das Hochschulwesen. Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik Nr. 62, Bamberg, 2002.
- [Wi98] Will, L.: SAP R/3-Systemadministration: Basiswissen für das SAP R/3-Systemmanagement. Addison-Wesley-Longman, Bonn, 1998.