Introduction to Operational Data Provisioning

Purpose

This is the main page to all content concerning the topic introduction to Operational Data Provisioning.

Overview

- Purpose
- Overview
- Introduction
- History and Background
- Qualities
- Scenarios
  - Embedded Analytics in Business Suite 7
  - Consolidation
  - Distribution
  - Transformation
  - Extraction
- Architecture
  - Model-Based Approach
  - ODP Consumer Interface
- Implementation
  - Operational Data Providers
    - SAP DataSources / Extractors
    - SAP HANA Information Views
    - SAP Business ByDesign
    - SAP Landscape Transformation Replication Server
    - SAP NetWeaver Business Warehouse
  - Operational Data Provisioning Consumer
    - SAP NetWeaver Embedded Analytics
    - SAP NetWeaver Business Warehouse
    - SAP BusinessObjects DataServices
- Outlook - what's next?
- References

Introduction

Data created within business processes can be referred to as operational data: business transaction documents, master data, configuration data. Data provisioning of operational data has the purpose of making this data available for further processing in other contexts. This is a well-known topic with a long history for integration scenarios spread across multiple systems. For example, master data are replicated between systems based on harmonized data structures and standardized interfaces. Another area is the classical process integration, where the data provider sends messages via A2A interfaces, always based on the concrete structures of the transaction documents to be processed. However, as soon as the concrete data structure cannot be cited anymore, so, if basically any kind of application data shall potentially be provided, a more general approach is required, which is referred to as operational data provisioning in the following.

Analytics represents such an area consisting of infrastructures, tools and applications that process data made available by source systems. A lasting assumption cannot be made about the data to be provided, and thus the data structures to be supported, because it is totally dependent on the application use case. On the other hand, analytical software consuming extracted data is a hot spot of innovation with varying needs and evolving approaches how to access data sources. Therefore, data exchange must be based on well-cut services communicating over widely available protocols. Operational data provisioning as described in this document does not claim to be the exclusive approach for data integration between applications. A decision has to be made case-by-case, the alternatives mentioned above are to be considered. For analytical contexts, operational data provisioning is the defined standard for extracting from SAP's ABAP platforms, and it may also be well-suited for certain other application scenarios as well.

History and Background

The concepts for Operational Data Providers, abbreviated ODPs, originate from earlier work done in the NetWeaver platform. Based on NetWeaver Business Warehouse (BW) technology, data were extracted from Business Suite applications to a Data Warehouse. DataSources were the first interface abstraction for heterogeneous application data used for defining data extraction structures and linking with extractor implementations. NetWeaver BW technology connected these DataSources with its data acquisition layer using a generic API (BW Service API). This approach formed a solid basis for reporting and analysis in a BW system operated as a side-car or as a central Enterprise Data Warehouse hub. Data extracted from the application DataSources are fed into modeled warehouse layers for data integration and harmonization. The warehouse layer adds analytical semantics and makes data available for multi-dimensional processing.
With the rising demand for operational analytics by embedding analytical capabilities into Business Suite applications, this approach for a unified access to application data was not yet sufficient to address local reporting requirements. For example, DataSources with transactional data are not linked with DataSources exposing referenced master data; DataSource field structures don't convey the analytical semantics needed for analytic queries.

This led to the development of the Operational Data Provisioning infrastructure for extracting data from applications built on the NetWeaver platform. It offers reuse of existing DataSources and provides a design-time tool and an API for creating ODPs as part of Search & Analytics models that capture relations between isolated data structures and the analytical semantics of their components. The ODP runtime component integrates with HANA or the BW Accelerator (BWA) in-memory technology for high performance access to large data volumes.

With this, the NetWeaver platform architecture introduced a unified data provisioning layer with rich metadata and analytical semantics. Based on this layer, operational reporting embedded into Business Suite 7i2011 could be established by defining analytic queries with multi-dimensional access to ODP data. This unified provisioning layer eliminated the need for several other analytics frameworks in former releases of the Business Suite, which all required their own vertical stack and established a tight coupling between data provider and consumer that was not reusable and did not scale to multiple frontends consuming the same data source. At the same time, the ODP concept protects and leverages investments in the large number of DataSources developed by Business Suite and industry solutions, which can be reused in this new infrastructure.

The next step in the evolution of this concept was triggered by SAP's ambition to increase presence in the on-demand and mobile space and reach one billion people with its products. This requires opening up the existing business applications for integration with other applications and infrastructures to enable desirable new scenarios. As a result, SAP business solutions often span across several products, residing on different systems and technology stacks, being deployed on-premise in the customer landscape, on-demand in a cloud, or on a mobile device. In such a heterogeneous environment, it is of the utmost importance for SAP’s ability to execute and for the overall SAP product experience that analytical application data can be provided once, with high technical and semantic quality, and then exchanged freely and reliably between applications and systems.

For the ABAP-based platforms NetWeaver and NGAP (Next Generation ABAP Platform), ODPs have been designated as the standard for extraction, because they offer advanced extraction capabilities including those especially relevant for data exchange between remote systems in hybrid landscapes: reliable transfer of very large data volumes and sophisticated delta handling procedures. Access to these ODPs is enabled via a standard ODP consumer interface, which is provided by the platform infrastructure.

ABAP-based business applications therefore shall provide data relevant for analytics as ODPs. The underlying application platform contains the infrastructure for Operational Data Provisioning including an adapter for the standard ODP consumer interface.

Data integration components such as a NetWeaver BW-based data warehouse and Data Services consume application data exposed as ODPs.
In addition, NetWeaver BW shall make its warehousing data again available for extraction by other consumers via the ODP contract.

Qualities

The driving forces to develop operational data provisioning were given by specific goals addressing the evolving needs for architectures supporting analytical capabilities on the NetWeaver platform. Initially formulated with requirements for a unified local reporting on operational data in mind, the goals remained stable when applied to heterogeneous system landscapes combining multiple products. The quality attributes associated to these goals cover both local reporting as well as data extraction in heterogeneous landscapes.

The first goal is openness to foster interoperability between data providers and extraction consumers.

Investments in developed data providers and consuming applications must be protected:

- A stable and unified contract decouples data providers from consumers in order to limit the impact of changes on either side. Extraction capabilities are made available by a standardized ODP consumer interface. With this, vertical stacks for every data flow are avoided meaning consumers do not need dedicated components with extraction logic to interact with data providers.
- The contract enables interoperability across technology platforms. It defines a protocol for the interactions between consumers and providers and formats for the exchanged data.

The acceptance of the concept highly depends on the level of consistency assured by the contract. Consumers expect extracted data to be coherent, complete, up-to-date and received in a timely manner, processed reliably also for large data volumes:

- The contract enables providers to describe exposed data structures and semantics as metadata.
- The contract enables consumers to define the scope of data to be extracted. It supports direct access to data, full extraction of a data snapshot as well as automated and reliable delta loads of data that have been changed and created on provider side.

Finally, flexible deployment enables easy integration of providing and consuming products in complex system landscapes:

- The ODP consumer interface defines a common set of operations that constitute the data provisioning contract. Such a generic interface can be applied to any ODP model and therefore results in low development and integration costs when setting up integration scenarios.
- A Web-based implementation of the ODP consumer interface supports integration scenarios across different SAP products and across different landscapes.
- The decision whether to extract data physically by copying them to a target in the consumer or to leave them in the source and directly access them at the data provider is deferred to the time of interaction.

The unified contract opens up the possibility to introduce central infrastructure components (NetWeaver BW, DataServices) between ODP providers and consumers. They are only needed if their specific functionality on top of the provided extraction capabilities is required.

Scenarios

Providing the qualities described above, the ODP concept enables extraction for different application scenarios, where NetWeaver-based applications or tools act as data provider or consumer. Five main archetypes can be distinguished:

**Embedded Analytics in Business Suite 7**

Create unified analytical views of operational data for local reporting and analytics.
Applications create ODPs based on a common meta model.

An ODP contains a link to the underlying operational application data source and describes the analytical semantics of the exposed data structure.

The ODP infrastructure provides uniform and direct access to data and metadata of an ODP, without replication to a data warehouse.

Given the rich semantics of ODPs, transient BW InfoProviders emulate multi-dimensional structures for ODPs and form the basis for operational reporting on the local system.

Applications can create BEx queries on top of these transient InfoProviders.

Consolidation

Build analytical views integrating operational data from multiple applications.

- NetWeaver BW implements an ODP adapter connecting to application ODPs via any of the two supported communication channels.

- A NetWeaver BW-based Enterprise Data Warehouse acts as central data hub providing an integrated view on data from multiple applications to avoid point-to-point connections between applications for data replication and to provide user access to integrated data that does not require direct access to data of contributing applications.

- The data warehouse contains analytical data models for integration and harmonization of data from multiple ODP sources.

Distribution
Supply downstream systems with up-to-date data.

- Basically any kind of NetWeaver-based application can use ODPs to distribute its data to multiple targets. The ODP infrastructure ensures that data and deltas are correctly supplied to all target systems.

**Example:**

- A NetWeaver BW warehouse system makes its data available to multiple target BW warehouse systems. While the source system often holds enterprise-wide data, the receiving systems may be so-called data marts that typically hold only a subset of the original data.

- Data marts implement independent and decentralized data stores for specific business, on a different location, or for better performance. Their operation requires a controlled, cost-effective method to replicate data from the enterprise system.

For data transfer, the source BW system releases the InfoProviders with relevant data as ODPs. In the target BW systems, data transfer and propagation (DTP) makes use of the ODP consumer interface to replicate these data into the local InfoProviders.
Connect with non-ODP data providers and consumers.

- BusinessObjects Data Services acts as central data distribution hub allowing integration of data from multiple sources, to transform, standardize and de-duplicate them and to load them into multiple different targets.

Data Services integrates:
- ODP providers with data consumers that are not able to consume ODPs,
- ODP consumers with data providers that are not able to expose data via ODPs.

This applies to non-SAP systems as well as SAP systems with lower releases. It should be noted that at the time of writing, the integration with ByDesign is a prototype and not generally available.

Extraction

Make operational application data available for other scenarios.

Applications expose their local analytical views that are suitable for extraction as ODPs. The application platforms provide different implementations: The Business Suite applies the approach also taken for embedded analytics (see above), applications based on the ByDesign platform transform their multi-dimensional analytical views (MDAVs) to ODPs and expose them as extraction data sources, and in a NetWeaver-on-HANA stack a specific ODP provider exists to expose HANA views as ODPs.

- Access to ODPs is enabled by a standard consumer interface that exposes a unified representation of analytical views implemented with different platform-specific architectures.
- Extraction with ODPs embraces both, direct access as well as full and delta replication of operational data.
- The ODP infrastructure supports two communication channels: RFCs for ABAP consumers and a SOAP Web service for non-ABAP consumers and consumers in hybrid system landscapes.

Architecture

The NetWeaver and NGAP (Next Generation ABAP Platform) platforms implement the Operational Data Provisioning concepts using the architecture depicted in the following diagram.

For the purpose of analytics within the application, data is represented as analytical views implemented with local capabilities of the platform (Business Suite: DataSource, ByDesign: MDAV, HANA: view).

An ODP context is a technical component within the application platform that transforms analytical views of a given technology to ODPs. The application owns a registry to control which of the available local analytical views are released for extraction via this ODP context. Note that in case a system includes multiple types of local views such as a Business Suite-on-HANA system offering ODPs based on DataSources and on HANA views, multiple ODP contexts may be active.

The ODP consumer interface decouples consumers from the providing business applications. Consumers access ODPs of a business application via this interface. It offers a unified contract for accessing ODP extraction capabilities and includes operations for reading the ODP catalog, reading ODP metadata and extracting data from ODPs.
The ODP infrastructure manages access to exposed ODPs. Catalog and metadata requests as well as direct access to ODP data is federated to the respective ODP contexts. The infrastructure maintains a delta queue for extraction of full and delta data with support for concurrent access by multiple consumers. For ensuring the extraction reliability, the delta queue implements the EOIO data delivery and preserves the transactional data coherence.

**Model-Based Approach**

Information about the capabilities and data structure of ODPs is communicated between providing business application and ODP consumer based on a common meta model (see figure below).

A **context** represents a source of ODPs. Known context identifiers exist for all technologies whose analytical views can be exposed as ODPs: for DataSource-based ODPs, for HANA views, for ByDesign MDAVs (Multidimension Analytical View), and for ODPs provided by the ODP infrastructure for metadata introspection.

**ODPs** are always accessed within a context. They represent an analytical view provided by the application, which shall become visible to external consumers. ODPs encompass metadata and data, both of which are provided by the containing context. An ODP is always classified according to the data it represents. It is either a **transactional ODP** with facts from business transaction documents, or a **master data ODP**, an ODP with language-dependent **texts** for IDs and code values used in other ODPs, or an ODP representing a parent/child **hierarchy**. Data for all but transactional ODPs may be time-dependent. Common attributes of an ODP include an identifier, a language-dependent name, and the supported access modes (full extraction, delta extraction, direct access), and the supported modes for delta extraction.

The ODP data structure consists of one or more **fields**. The description of a field includes an identifier, a language-dependent name, a technical data type, key field indicator, its analytical type (characteristic/dimension, key figure/measure with aggregation behavior, navigation attribute), selection properties (mandatory, supported options), and authorization relevance.
ODPs can be related to other ODPs by associations linking some of their fields. Typical examples are transactional ODPs with fields holding IDs for customers, countries, etc. They can be associated to corresponding key fields of other ODPs providing access to the corresponding master data or texts.

**ODP Consumer Interface**

The platform provides the standard interface for accessing ODPs exposed by applications and consuming their data.

The interface supports different extraction modes:

- **Direct access** for querying selected data directly at the source. This kind of extraction is performed with a synchronously executed operation returning the complete query result in a single response.

- **Full replication** for copying a selective snapshot of the entire ODP data. Replication involves orchestration of several interface operations. The interface implements a cursor concept for reliable paged access to packages of limited size with the option to repeat individual fetch operations.

- **Delta replication** for an initial load of a subset of the entire ODP data (comparable to a full replication) followed by a continuous feed of changes to these data (modifications, deletions, new records).

For this kind of extraction, consumers subscribe to ODPs, and the ODP infrastructure maintains the extraction state per consumer. Replication is performed as a sequence of requests each of which is an orchestration of several interface operations. The first request performs the initial full replication, subsequent requests at later times get the incremental deltas to these data.

The interface operations and signatures adhere to the terms introduced by the meta model. The interface offers:

- **Catalog** operations to retrieve available contexts and their ODPs.

- **Metadata** operations to get an overview description of a given ODP. For accessing full details about the structure and semantics of ODPs and their associations, the ODP infrastructure offers the context ODP_SELF with certain ODPs returning these metadata. If you plan to use this particular context, please communicate your use case to the PI HANA Platform P&D BW team before.

- **Extraction** operations for directly accessing a selected set of ODP data and replicating ODP data. Extracted data can be returned in an ABAP internal table to RFC (Remote Function Call) consumers or as XML document, optionally compressed. Two XML formats are supported, the open ABAP XML and the optimized Binary XML, and can be negotiated with the consumer.

The interface is implemented for two communication channels. ABAP consumers can access the interface on NetWeaver via ABAP RFCs. The NGAP (Next Generation ABAP Platform) implementation of the interface can be accessed via SOAP (Simple Object Access Protocol) Web service, which enables secure data transmission to other cloud infrastructures and to systems in the customer landscape.

At the time of writing, the interface has not been generally released, but only for specific consumers and integration scenarios: Embedded Analytics in Business Suite 7, BW Enterprise Data Warehouse and Data Services.

**Implementation**

This section gives an overview of the currently available implementations of ODP providers and consumers.

**Operational Data Providers**

Currently, the following ODP contexts are available:

**SAP DataSources / Extractors**

This context exposes BW DataSources as ODPs. A BW DataSource defines an extraction structure that is filled by an extractor implementing the logic to retrieve the relevant data from the ABAP system. There are application-specific extractors, each of which is hard-coded. In addition, there are generic extractors for tables, views and application areas such as LIS, CO-PA, FI-SL and HR.

Without additional configuration, the ODP framework supports DataSources that have been released by the application owner since SAP NetWeaver 7.0 SPS 24. Most released DataSources are part of SAP Business Suite. Consumers access these ODPs via the RFC implementation of the ODP consumer interface.
The context “Search and Operational Analytics” is used for embedding analytical capabilities into the SAP Business Suite 7i2011+. It enables operational analytics on the application data locally in the application system without replicating the data to a data warehouse system. The Search and Analytics Modeler provides a design time tool to create ODPs by importing BW DataSources and defining their ODP properties and associations between ODPs. The modeler can be launched in the SAP GUI with transaction code ESH_MODELER. The transient BW InfoProviders resulting from ODP definitions can be identified by their name prefix ‘2O’.

SAP HANA Information Views

This context makes analytic views, calculation views and associated attribute views defined in a HANA system release 1.0 SP4+ available as (transient) ODPs in a connected NetWeaver 7.30 SP5+ system. It allows replicating the view data into another data warehouse system or to perform BW queries directly on HANA views. Such queries can be defined on top of transient BW InfoProviders, which are provided for the transient ODPs and can be identified by their name prefix ‘2H’. Consumers access these ODPs via the RFC implementation of the ODP consumer interface.

This implementation opens up the possibility to shift business logic from the ABAP-based BW warehousing and query layers to the HANA database. Examples include calculations of key figures, restricted key figures, currency conversions, joins and unions, which can be defined in HANA views. The results of such calculations can then also be accessed via ODP.

SAP Business ByDesign

With the ODP context for SAP Business ByDesign, data represented with ByDesign’s multi-dimensional analytical views (MDAVs) can be extracted via ODP from the system. This functionality is available for an extended ByDesign solution scope covering the integration with SAP ERP for Subsidiaries. The set of MDAVs that are to be released as ODPs can be controlled with the Business Analytics work center, where they are marked as exposed DataSources. ByDesign currently implements the full extraction mode with the SOAP Web Service for the ODP consumer interface.

SAP Landscape Transformation Replication Server

Using the ODP infrastructure and the trigger-based replication of SAP Landscape Transformation Replication Server (SAP LT Replication Server), data can be transferred in real time from an ABAP system to one or more data warehouse systems. SAP LT Replication Server acts as a provider for the ODP infrastructure. It can make tables from SAP sources available as delta queues. This provider implementation is available with DMIS 2011 SP5 and SAP NetWeaver 7.30 SP8.

SAP NetWeaver Business Warehouse

SAP NetWeaver Business Warehouse exposes most of its data as ODPs for replication purposes. Starting with release 7.40 SP5, the following object types are supported: DataStore-Object, InfoCube, Semantic partitioned object, HybridProvider, MultiProvider, InfoSet, Queries released as InfoProvider and InfoObjects for master data, texts and hierarchies. The ODPs are accessible via the RFC implementation of the ODP consumer interface.

Operational Data Provisioning Consumer

SAP NetWeaver Embedded Analytics

If an ODP context offers its Operational Data Providers for embedded analytics, it is possible to define and execute a BW query directly on the Operational Data Provider. It is not necessary to replicate the data into BW. The BW query is defined and executed locally on top of a transient InfoProvider, which is implicitly derived from the ODP. The list of available ODP-based transient InfoProviders and their details can be displayed in the SAP GUI with transaction code RSRTS_ODP_DIS. Embedded Analytics for Operational Data Provider is supported since SAP NetWeaver release 7.30.

SAP NetWeaver Business Warehouse

Beginning with release 7.30 SP8 / 7.31 SP5, the SAP NetWeaver Business Warehouse consumes data exposed via ODP from other systems as described in chapter 4 in section Consolidation. Both communication channels RFC and HTTP/SOAP are supported. Starting with NW 7.40 SP5, the context for DataSources / Extractors also becomes available for BW consumption.

SAP BusinessObjects DataServices

SAP DataServices release 4.2 SP1 can extract data from ODPs, transform and standardize them, and then load them to different target systems. It uses the RFC-based implementation of the consumer interface to connect to NetWeaver systems.

Outlook - what’s next?
ODPs could be consumed in many more consumption scenarios, if the entry barriers for accessing them would be lowered. For this purpose, the ODP infrastructure could be wrapped by an **OData** provider making ODPs available via REST-based OData services. This would also support SAP’s overall strategy to adopt OData for lightweight consumption of data in SAP business systems. At the time of writing, a first prototype exposing BW InfoProviders via ODP via OData demonstrates that OData extraction services nicely complement the existing ODP infrastructure and preserve the main qualities associated with Operational Data Provisioning. It opens up a lightweight method for data extraction by applying ubiquitous Web-oriented technologies based on open standards OData and HTTP.

**Core Data Services** (CDS) enhance SQL to allow defining and consuming semantically rich data models natively in HANA (High-Performance Analytic Appliance) applications, thereby improving productivity, consumability, performance and interoperability. The key concept for expressing semantics of CDS models are annotations that are attached to elements of a model (entities, views). Applying annotations with analytical semantics, CDS views can be annotated such that the resulting metadata would allow generating transient ODPs. Therefore, one future direction might be an ODP Provider implementation exposing CDS views with matching analytical annotations.

**References**

“Modeling – Enterprise Data Warehouse Layer – DataSource”, NetWeaver platform 7.3. Published on SAP Help

“Search and Operational Analytics – Operational Data Provisioning”, NetWeaver platform 7.3. Published on SAP Help

Operational Data Provisioning Published on SAP Help

Operational Data Provisioning - Troubleshooting