Northwind BI solution

Оглавление

[Planning a BI Solution 4](#_Toc129296698)

[Project Scope 4](#_Toc129296699)

[Required BI Software 4](#_Toc129296700)

[Project conceptual schema 5](#_Toc129296701)

[Planning SQL Server Business Intelligence Infrastructure 6](#_Toc129296702)

[BI Topology 6](#_Toc129296703)

[Single BI server 6](#_Toc129296704)

[Dedicated report server 7](#_Toc129296705)

[Dedicated ETL server 8](#_Toc129296706)

[Dedicated master data server 9](#_Toc129296707)

[Distributed report server 10](#_Toc129296708)

[Planning Data Warehouse Hardware 11](#_Toc129296709)

[Designing Data Warehouse 11](#_Toc129296710)

[Designing Data Warehouse Logical Schema 11](#_Toc129296711)

[Identify business processes 11](#_Toc129296712)

[Designing Dimension Models and Data Warehouse Tables 12](#_Toc129296713)

[Data source profiling 13](#_Toc129296714)

[Dimension 15](#_Toc129296715)

[Fact 16](#_Toc129296716)

[Designing Data Warehouse Physical Implementation 16](#_Toc129296717)

[Designing File Storage 16](#_Toc129296718)

[Partitioning 17](#_Toc129296719)

[Indexes 17](#_Toc129296720)

[Compression 17](#_Toc129296721)

[Views 17](#_Toc129296722)

[Designing ETL Solution 18](#_Toc129296723)

[Xmlcalendar 18](#_Toc129296724)

[Incremental load package 18](#_Toc129296725)

[Customer 19](#_Toc129296726)

[Date 20](#_Toc129296727)

[Employee 21](#_Toc129296728)

[Product 22](#_Toc129296729)

[Order 23](#_Toc129296730)

[Data Warehouse maintenance 24](#_Toc129296731)

[Planning BI Delivery Solution 25](#_Toc129296732)

[Reporting Requirements 25](#_Toc129296733)

[Monitoring and Optimizing BI Solution 25](#_Toc129296734)

[Creating a Performance Baseline 25](#_Toc129296735)

[Monitoring Data Warehouse 25](#_Toc129296736)

[Monitoring Power BI Report Server 26](#_Toc129296737)

[Workload Priorities 26](#_Toc129296738)

[Operating BI Solution 27](#_Toc129296739)

[ETL Operations 27](#_Toc129296740)

[Deploying and Configuring Packages 27](#_Toc129296741)

[Environments and Variables 27](#_Toc129296742)

[Considerations for Package Execution 28](#_Toc129296743)

[Package Execution Reports 28](#_Toc129296744)

[Backing Up the SSIS Catalog 28](#_Toc129296745)

[Data Warehouse Operations 28](#_Toc129296746)

[Reporting Services Operations 28](#_Toc129296747)

[Automating Operational Tasks 28](#_Toc129296748)

[SQL agent 28](#_Toc129296749)

[Subsystems maintenance plan 29](#_Toc129296750)

# Planning a BI Solution

[Northwind BI Plan.mpp](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/Docs/Northwind%20BI%20Plan.mpp)

[Northwind BI Plan with prototype.mpp](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/Docs/Northwind%20BI%20Plan%20with%20prototype.mpp)

## Project Scope

|  |  |  |
| --- | --- | --- |
| High business value |  | * See sales amount with discount by customers and products. * Compare sales amount with discount with previous year/quarter/month/week. * See discount by employees and their geography. * See new and returned customers. * Cluster customers based on spending volume. * Product ABC classification. * Product basket analysis. * See number of orders, sales amount, sales amount with discount in progress |
| Low business value |  |  |
|  | Low feasibility | High feasibility |

## Required BI Software

|  |  |
| --- | --- |
| Software | Rationale |
| SQL Server Database Engine | * The CEO wants a consistent view of all business data, and a centralized data warehouse in SQL Server would provide this. |
| SQL Server Integration Services | * The business data required for analysis and reporting is currently spread across a range of data sources. Integration Services will provide an ETL platform to populate and refresh the data warehouse. |
| SQL Server Master Data Services | * The Sales VP has complained about inconsistent data, which could potentially be caused by a lack of central data management for key business entities. |
| SQL Server Data Quality Services | * The specialists have difficulty ensuring that their analysis of sales data is accurate because of quality issues in the data. Data Quality Services could be used to cleanse records and improve the consistency, thereby reducing the inaccuracy of the sales analysis. |
| Power BI Report Server | * Business users need the centralized portal for analysis and reporting. * The database administrators team needs to receive email notifications about errors in the ETL process. * Data stewards need to be notified of errors and raw values in data cleansing processes. |

## Project conceptual schema

Data sources

Northwind (упрощённая)



xmlcalendar.ru

ETL



Reporting and Analysis

Data Cleansing



Data management



Data

Warehouse



# Planning SQL Server Business Intelligence Infrastructure

## BI Topology

### Single BI server



* SQL Server DB Engine
  + Data Warehouse
  + Landing database
  + SSIS catalog
  + Reporting Services catalog
  + DQS catalog
  + MDS catalog
* SQL Server Integration Services
* Power BI Report Server
* Data Quality Services
* Master Data Services

SRVBI



CLIENT PC

* Microsoft Excel
  + Power Query
  + Power Pivot
  + Power View
* Report Builder
* Power BI Desktop
* Microsoft Edge

Notes:

* Minimal server hardware and software license requirements, but the server would require significant memory, CPU, and disk resources for all but the most lightweight BI workloads.
* The range of different workload types on the server would make it difficult to specify and configure hardware resources appropriately.
* The server could be clustered to provide high availability.

### Dedicated report server



* SQL Server DB Engine
  + Data Warehouse
  + Landing database
  + SSIS catalog
  + DQS catalog
  + MDS catalog
* SQL Server Integration Services
* Data Quality Services
* Master Data Services

SRVDWH



CLIENT PC

* Microsoft Excel
  + Power Query
  + Power Pivot
  + Power View
* Report Builder
* Power BI Desktop
* Microsoft Edge



* SQL Server DB Engine
  + Reporting Services catalog
* Power BI Report Server

SRVPBIRS

Notes:

* The reporting and DWH workloads are full separated.
* Using a dedicated reporting server makes it easier to manage different types of workloads.
* Additional report servers could be added to scale out reporting.
* One or both servers could be clustered to provide high availability.

### Dedicated ETL server



* SQL Server DB Engine
  + Data Warehouse
  + Landing database

SRVDWH



CLIENT PC

* Microsoft Excel
  + Power Query
  + Power Pivot
  + Power View
* Report Builder
* Power BI Desktop
* Microsoft Edge



* SQL Server DB Engine
  + Reporting Services catalog
* Power BI Report Server

SRVPBIRS



* SQL Server DB Engine
  + SSIS catalog
  + DQS catalog
  + MDS catalog
* SQL Server Integration Services
* Data Quality Services
* Master Data Services

SRVETL

Notes:

* The DWH, reporting and ETL workloads are full separated.
* The landing database could be hosted on the ETL server or data warehouse server depending on workloads.
* SRVDWH, SRVPBIRS, and SRVETL could be clustered to provide high availability.
* Additional report servers and ETL could be added to scale out.

### Dedicated master data server



* SQL Server DB Engine
  + Data Warehouse
  + Landing database

SRVDWH



CLIENT PC

* Microsoft Excel
  + Power Query
  + Power Pivot
  + Power View
* Report Builder
* Power BI Desktop
* Microsoft Edge



* SQL Server DB Engine
  + Reporting Services catalog
* Power BI Report Server

SRVPBIRS



* SQL Server DB Engine
  + DQS catalog
  + MDS catalog
* Data Quality Services
* Master Data Services

SRVMDM



* SQL Server DB Engine
  + SSIS catalog
* SQL Server Integration Services

SRVETL

Notes:

* The DWH, reporting, ETL and MDM workloads are full separated.
* The landing database could be hosted on the ETL server or data warehouse server depending on workloads.
* Servers could be clustered to provide high availability.
* Additional report servers and ETL could be added to scale out.

### Distributed report server



* SQL Server DB Engine
  + Data Warehouse
  + Landing database

SRVDWH



CLIENT PC

* Microsoft Excel
  + Power Query
  + Power Pivot
  + Power View
* Report Builder
* Power BI Desktop
* Microsoft Edge



* Power BI Report Server

SRVPBIRS



* SQL Server DB Engine
  + DQS catalog
  + MDS catalog
* Data Quality Services
* Master Data Services

SRVMDM



* SQL Server DB Engine
  + SSIS catalog
* SQL Server Integration Services

SRVETL



* SQL Server DB Engine
  + Reporting Services catalog

SRVPBIRSDB

Notes:

* The DWH, reporting, ETL and MDM workloads are full separated.
* The landing database could be hosted on the ETL server or data warehouse server depending on workloads.
* The report server uses a separate database server to host the report catalog.
* Servers could be clustered to provide high availability.
* Additional report servers and ETL could be added to scale out.

## Planning Data Warehouse Hardware

[DW Hardware Spec.xlsx](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/Docs/DW%20Hardware%20Spec.xlsx&version=GBmaster&_a=contents)

# Designing Data Warehouse

## Designing Data Warehouse Logical Schema

### Identify business processes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Business processes** | **Dimensions** | | | |
| Date | Product | Customer | Employee |
| Orders | X | X | X | X |

Notes:

* There is no interest in shipment geography, so there is no reason to create dedicated geography dimension. Employee and Customer geography will be implemented as attributes.
* In Date dimension will be used opensource project [Производственный календарь в XML (xmlcalendar.ru)](http://xmlcalendar.ru/) for getting information about production calendar.

### Designing Dimension Models and Data Warehouse Tables

#### Order dimensional model

Date

(Order, Required, Shipped)

Year

Quarter

Month

Day

Year

Quarter

Mon

ISO Week

Weekday

Year

Quarter

Month

Week

Weekday

Customer

Country

City

Customer

Contact name

Contact title

Phone

Fax

DayOfQuarterNumber

DayOfYearNumber

StartOfYear

EndOfYear

YearQuarter

StartOfQuarter

EndOfQuarter

YearMonth

StartOfMonth

EndOfMonth

StartOfWeek

EndOfWeek

Holiday

Workday type

Workday hours

Product

Category

Product

Employee

Country

City

Employee

Title

Title of courtesy

### Data source profiling

Data sources

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Type | Description | Profiling |
| Northwind (упрощённая) | SQL Server | Data about customers, employees, products, orders | [Northwind Data Profile.xml](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/Docs/Northwind%20Data%20Profile.xml)[[1]](#footnote-1) |
| [Производственный календарь в XML (xmlcalendar.ru)](http://xmlcalendar.ru/) | XML | Information about holidays in Russia | [xmlcalendar/data (github.com)](https://github.com/xmlcalendar/data) |

#### Customer

Table 1. Northwind (упрощённая).dbo.Customers.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| CustomerID | CompanyName | ContactName | ContactTitle | City | Country | Phone | Fax |
| ALFKI | Alfreds Futterkiste | Maria Anders | Sales Representative | Berlin | Germany | 030-0074321 | 030-0076545 |
| ANATR | Ana Trujillo Emparedados y helados | Ana Trujillo | Owner | Mexico D.F. | Mexico | (5) 555-4729 | (5) 555-3745 |
|  | SCD1 | SCD1 | SCD1 | SCD2 | SCD2 | SCD1 | SCD1 |
| Business key | Drill-thought detail | Drill-thought detail | Slicer | Hierarchy | | Drill-thought detail | Drill-thought detail |

#### Product

Table 2. Northwind (упрощённая).dbo.Products.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ProductID | ProductName | SupplierID | CategoryID | UnitPrice |
| 1 | Chai | 1 | 1 | 18,00 |
| 2 | Chang | 1 | 1 | 19,00 |
|  | SCD1 |  |  |  |
| Business key | Drill-thought detail | Foreign key | Foreign key (Hierarchy) | Drill-thought detail |

Table 3. Northwind (упрощённая).dbo.Categories.

|  |  |  |
| --- | --- | --- |
| CategoryID | CategoryName | Description |
| 1 | Beverages | Soft drinks, coffees, teas, beers, and ales |
| 2 | Condiments | Sweet and savory sauces, relishes, spreads, and seasonings |
|  | SCD1 |  |
| Business key | Drill-thought detail | Drill-thought detail |

#### Employee

Table 4. Northwind (упрощённая).dbo.Employees.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EmployeeID | LastName | FirstName | Title | TitleOfCourtesy | City | Country |
| 1 | Davolio | Nancy | Sales Representative | Ms. | Seattle | USA |
| 2 | Fuller | Andrew | Vice President, Sales | Dr. | Tacoma | USA |
|  | SCD1 | SCD1 | SCD2 | SCD1 | SCD2 | SCD2 |
| Business key | Drill-thought detail | Drill-thought detail | Drill-thought detail | Drill-thought detail | Hierarchy | |

#### Order

Table 5. Northwind (упрощённая).dbo.Orders

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| OrderID | CustomerID | EmployeeID | OrderDate | RequiredDate | ShippedDate | ShipCity | ShipCountry |
| 10248 | VINET | 5 | 1996-07-04 00:00:00.000 | 1996-08-01 00:00:00.000 | 1996-07-16 00:00:00.000 | Reims | France |
| 10249 | TOMSP | 6 | 1996-07-05 00:00:00.000 | 1996-08-16 00:00:00.000 | 1996-07-10 00:00:00.000 | Münster | Germany |
| Degenerate dimension | Dimension key | Dimension key | Dimension key | Dimension key | Dimension key |  |  |

Table 6. Northwind (упрощённая).dbo.Order Details.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| OrderID | ProductID | UnitPrice | Quantity | Discount |
| 10248 | 11 | 14,00 | 12 | 0 |
| 10248 | 42 | 9,80 | 10 | 0 |
| Degenerate dimension | Dimension key | Measure | Measure | Measure |

### Dimension

#### Customer

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| COLUMN NAME | COLUMN TYPE | SCD TYPE | DATA TYPE | NULL |
| CustomerKey | Surrogate key |  | INT | NOT |
| CustomerAlterKey | Business key |  | NVARCHAR(5) | NOT |
| Customer | Drill-thought detail | 1 | NVARCHAR(50) | NOT |
| ContactName | Drill-thought detail | 1 | NVARCHAR(50) |  |
| ContactTitle | Slicer | 1 | NVARCHAR(50) |  |
| Country | Hierarchy | 2 | NVARCHAR(25) |  |
| City | Hierarchy | 2 | NVARCHAR(25) |  |
| Phone | Drill-thought detail | 1 | NVARCHAR(30) |  |
| Fax | Drill-thought detail | 1 | NVARCHAR(30) |  |
| StartDate | Metadata |  | DATETIME2 | NOT |
| EndDate | Metadata |  | DATETIME2 |  |
| Current | Metadata |  | BIT | NOT |
| LineageKey | Metadata |  | INT | NOT |

#### Date

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| COLUMN NAME | COLUMN TYPE | SCD TYPE | DATA TYPE | NULL |
| DateKey | Surrogate key |  | INT | NOT |
| AlterDateKey | Business key |  | DATE | NOT |
| DayOfMonth | Hierarchy |  | TINYINT | NOT |
| DayOfWeek | Hierarchy |  | NVARCHAR(5) | NOT |
| DayOfWeekNumber | Sort by column |  | TINYINT | NOT |
| DayOfQuarterNumber | Drill-thought detail |  | TINYINT | NOT |
| DayOfYearNumber | Drill-thought detail |  | SMALLINT | NOT |
| Year | Hierarchy |  | SMALLINT | NOT |
| StartOfYear | Drill-thought detail |  | DATE | NOT |
| EndOfYear | Drill-thought detail |  | DATE | NOT |
| Quarter | Hierarchy |  | NVARCHAR(5) | NOT |
| YearQuarter | Hierarchy |  | NVARCHAR(10) | NOT |
| YearQuarterNumber | Sort by column |  | INT | NOT |
| StartOfQuarter | Drill-thought detail |  | DATE | NOT |
| EndOfQuarter | Drill-thought detail |  | DATE | NOT |
| Month | Hierarchy |  | NVARCHAR(10) | NOT |
| Mon | Hierarchy |  | NVARCHAR(5) | NOT |
| MonthNumber | Sort by column |  | TINYINT | NOT |
| YearMonth | Hierarchy |  | NVARCHAR(10) | NOT |
| StartOfMonth | Drill-thought detail |  | DATE | NOT |
| EndOfMonth | Drill-thought detail |  | DATE | NOT |
| Week | Hierarchy |  | NVARCHAR(50) | NOT |
| IOSWeekNumber | Hierarchy |  | TINYINT | NOT |
| StartOfWeek | Drill-thought detail |  | DATE | NOT |
| EndOfWeek | Drill-thought detail |  | DATE | NOT |
| Holiday | Drill-thought detail |  | NVARCHAR(100) | NOT |
| WorkDayType | Slicer |  | NVARCHAR(25) | NOT |
| WorkDayHours | Measure |  | TINYINT | NOT |

#### Employee

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| COLUMN NAME | COLUMN TYPE | SCD TYPE | DATA TYPE | NULL |
| EmployeeKey | Surrogate key |  | INT | NOT |
| EmployeeAlterKey | Business key |  | INT | NOT |
| Employee | Drill-thought detail | 1 | NVARCHAR(35) | NOT |
| Title | Drill-thought detail | 1 | NVARCHAR(30) |  |
| TitleOfCourtesy | Drill-thought detail | 1 | NVARCHAR(10) |  |
| City | Hierarchy | 2 | NVARCHAR(25) |  |
| Country | Hierarchy | 2 | NVARCHAR(25) |  |
| StartDate | Metadata |  | DATETIME2 | NOT |
| EndDate | Metadata |  | DATETIME2 |  |
| Current | Metadata |  | BIT | NOT |
| LineageKey | Metadata |  | INT | NOT |

#### Product

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| COLUMN NAME | COLUMN TYPE | SCD TYPE | DATA TYPE | NULL |
| ProductKey | Surrogate key |  | INT | NOT |
| ProductAlterKey | Business key |  | INT | NOT |
| Product | Hierarchy | 1 | NVARCHAR(50) | NOT |
| Category | Hierarchy | 1 | NVARCHAR(50) | NOT |
| LineageKey | Metadata |  | INT | NOT |

### Fact

#### Order

|  |  |  |  |
| --- | --- | --- | --- |
| COLUMN NAME | COLUMN TYPE | DATA TYPE | NULL |
| OrderKey | Degenerate dimension | INT | NOT |
| ProductKey | Dimension key | INT | NOT |
| CustomerKey | Dimension key | INT |  |
| EmployeeKey | Dimension key | INT |  |
| OrderDateKey | Dimension key | INT | NOT |
| RequiredDateKey | Dimension key | INT |  |
| ShippedDateKey | Dimension key | INT |  |
| UnitPrice | Measure | MONEY |  |
| Quantity | Measure | INT |  |
| Discount | Measure | MONEY |  |
| SalesAmount | Measure | MONEY |  |
| SalesAmountWithDiscount | Measure | MONEY |  |
| LineageKey | Metadata | INT |  |

## Designing Data Warehouse Physical Implementation

### Designing File Storage

|  |  |  |
| --- | --- | --- |
| Logical Drive | Filegroups | Notes |
| Disk1 | 1. PRIMARY 2. Default\_FG | * The system tables are in a dedicated filegroup to minimize fragmentation of data files. * The default filegroup is initialized with a single file of 100MB, which should be large enough for the tables it will contain for the foreseeable future. |
| Disk2 | Dimention\_Data | * Putting different tables used in the same join queries in different filegroups. This step will improve performance, because of parallel disk I/O searching for joined data. |
| Disk3 | Dimention\_Index | * Using different filegroups for heavily accessed tables and the nonclustered indexes will improve performance, because of parallel I/O if the files are located on different physical disks. |
| Disk4[[2]](#footnote-2) | Order\_Unkown\_Member\_Data | * This table is likely to grow extremely large, and so it is assigned to a dedicated logical drive to avoid competing for space. |
| Disk5 | Order\_1996\_Data | * As above |
| Disk6 | Order\_1997\_Data | * As above |
| Disk7 | Order\_Unkown\_Member\_Index | * Using different filegroups for heavily accessed tables and the nonclustered indexes will improve performance, because of parallel I/O if the files are located on different physical disks. |
| Disk8 | Order\_1996\_Index | * As above |
| Disk9 | Order\_1997\_Index | * As above |
| Disk10 | 1. Landing 2. Staging DB data files 3. Logs DB data files | * A separate staging database, landing filegroup and logs database is used, and its data files are stored on a dedicated logical disk to minimize the effects of fragmentation and I/O on the data warehouse data and activity. |
| Disk11 | TempDB data files | * TempDB is given its own logical disk to minimize the effect of fragmentation on the data warehouse while enabling fast I/O to a suitable pre-sized TempDB. |
| Disk12 | Database log files | * The logfiles for TempDB, the Staging database, the Logs database and the data warehouse are stored on this drive to isolate log activity from data I/O. |
| Disk13 | Backup volume | * This logical drive will be used for backup file storage |

### Partitioning

The **Order** table are partitioned on the **ShippedDateKey** column. This column was chosen because orders are not considered complete until shipped, so it is shippedto assume that the ship date is used as the cut off for each month’s orders.

Orders for the current month are partitioned into daily partitions, as this is likely to reflect the ETL load frequency. Orders from previous month are merged to form a single partition per month and Orders from previous years are merged to form a single partition per year, as this reduces management complexity.

### Indexes

The dimension tables each have a clustered index on the surrogate key and a non-clustered index on the alternate key, which includes SCD metadata columns where they exist. Additional non-clustered indexes are created on attribute columns basing on [dimensional models](#_Designing_Dimension_Models) and Database Engine Tuning Advisor.

The fact table has clustered columnstore index to maximize aggregations performance. Additional nonclustered indexes are created on dimension key columns to improve performance for joins.

### Compression

All row-stored indexes are compressed using PAGE.

Column-stored indexes compressed using COLUMNSTORE for current and previous year partitions and COLUMNSTORE\_ARCHIVE for archived data.

### Views

Each table has a matching view in the **Reports** schema. These views have user-friendly names and use the NOLOCK query hint to minimize locking.

# Designing ETL Solution

[Source to target mapping.xlsx](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/Docs/Source%20to%20target%20mapping.xlsx&version=GBmaster&_a=contents)

## Xmlcalendar

For each xml files

Update local repository

xmlcalendar/data

Github.com

Landing.Holidays

xml file

Sort holidays

Sort days

Left join by holiday Id

Add year to holiday date change datatypes

## Incremental load package

Load Dimensions

Load Facts

Customer

Employee

Product

Date

DQS status subscription activation

DWH maintenance

Order

Truncate landing

## Customer

Northwind (упрощённая)

Customers

Filter on modified date

Clean data with DQS

Load and validate customers data with MDS

Filter on LastChgDateTime and succeeded validation

Add metadata columns

Lookup existing rows

Start audit

Insert new rows (generate surrogate key)

Update SCD1 rows

Customer

Update and Insert SCD2 rows (generate surrogate key)

End audit

DQS

MDS

Customer, ContactName, ContactTitle, Phone, Fax

## Date

xmlcalendar

.\ru

Clean data with DQS

Load and validate holidays data with MDS

Add unknown member

Start audit

Date

End audit

Get start and load dates from Landing.Orders

Populate Date dimension from 01-01 first year to 12-31 last +1 year

DQS

MDS

## Employee

Northwind (упрощённая)

Employees

Filter on modified date

Clean data with DQS

Load and validate employees data with MDS

Add metadata columns

Lookup existing rows

Start audit

Insert new rows (generate surrogate key)

Update SCD1 rows

Employee, TitleOfCourtesy

Employee

Update and Insert SCD2 rows (generate surrogate key)

End audit

Filter on LastChgDateTime and succeeded validation

DQS

MDS

## Product

Northwind (упрощённая)

Products

Filter on modified date

Clean data with DQS

Load and validate products & categories data with MDS

Add metadata columns

Lookup existing rows

Start audit

Insert new rows (generate surrogate key)

Update SCD1 rows

Product, Category

Product

End audit

Categories

Filter on modified date

Clean data with DQS

Filter on LastChgDateTime and succeeded validation

DQS

MDS

## Order

Northwind (упрощённая)

Orders

Create new partition

Create staging table

Load data to staging table

Start audit

Switch partition with unknown shipped date to fact table

Order

End audit

Order Details

Create filegroups if necessary

Get start load date for incremental load

Switch partition with new shipped date to fact table

Insert or update data for previous dates

Drop staging table

## Data Warehouse maintenance

MDS clean up staging tables

Backup

Copy Data Warehouse metadata to Logs DB

Update statistics

Set filegroups read only

MDS clean up transactions

Optimize indexes

Check DWH integrity

# Planning BI Delivery Solution

Power BI usage scenarios: [On-premises reporting](https://learn.microsoft.com/en-us/power-bi/guidance/powerbi-implementation-planning-usage-scenario-on-premises-reporting)

## Reporting Requirements

|  |  |  |  |
| --- | --- | --- | --- |
| Requirement | Audience | Tool | Notes |
| See sales amount with discount by customers and products | Sales VP | Power BI |  |
| Compare sales amount with discount with previous year/quarter/month/week | CEO, Sales VP | Power BI | [Standard time-related calculations](https://www.daxpatterns.com/standard-time-related-calculations/) |
| See discount by employees and their geography | Sales Representative | Power BI |  |
| See new and returned customers | Sales VP | Power BI | [New and returning customers](https://www.daxpatterns.com/new-and-returning-customers/) |
| Cluster customers based on spending volume | Sales VP | Power BI | [Dynamic segmentation](https://www.daxpatterns.com/dynamic-segmentation/) |
| Product ABC classification | Sales VP | Power BI | [ABC classification](https://www.daxpatterns.com/abc-classification/) |
| Product basket analysis | Inside Sales Coordinator | Power BI | [Basket analysis](https://www.daxpatterns.com/basket-analysis/) |
| See number of orders, sales amount, sales amount with discount in progress | Inside Sales Coordinator | Power BI | [Events in progress](https://www.daxpatterns.com/events-in-progress/) |
| Receive email notifications about errors in the ETL process | Database administrators team | Paginated report |  |
| Notification of errors and raw values in data cleansing processes | Data stewards | Paginated report |  |

# Monitoring and Optimizing BI Solution

## Creating a Performance Baseline

Before applying an effective health monitoring process, it is necessary to have a set of resource utilization metrics with which to compare future measurements. It is necessary to collect baseline resource utilization metrics for each server or service in the BI solution and collect the data over a sufficient period to reflect normal workload cycles. These workload cycles should include:

* Data warehouse load operations.
* Power BI reports processing operations.
* Normal everyday query activity.
* Peak-time query activity.

## Monitoring Data Warehouse

1. [Manage Data Collection](https://learn.microsoft.com/en-us/sql/relational-databases/data-collection/manage-data-collection?view=sql-server-ver15)
2. Performance monitor counters ([Performance Monitor Template](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/Monitoring/Performance%20Monitoring.xml&version=GBmaster))[[3]](#footnote-3):
   1. Memory: Available MBytes
   2. Paging File: % Usage
   3. Physical Disk:Avg. Disk sec/Read
   4. Physical Disk: Avg. Disk sec/Write
   5. Physical Disk: Disk Reads/sec
   6. Physical Disk: Disk Writes/sec
   7. Processor: % Processor Time
   8. SQLServer:Buffer Manager: Buffer cache hit ratio
   9. SQLServer:Buffer Manager: Page life expectancy
   10. SQLServer:General Statistics: User connections
   11. SQLServer:Memory Manager: Memory grants pending
   12. SQLServer:SQL Statistics: Batch requests/sec
   13. SQLServer:SQL Statistics: Compilations/sec
   14. SQLServer:SQL Statistics: Recompilations/sec
   15. System: Processor queue length
   16. SQLServer:Resource Pool Stats: CPU control effect %
   17. SQLServer:Workload Group Stats: CPU usage %
   18. SQLServer:Workload Group Stats: Reduced memory grants/sec
3. SQL Server Profiler ([SQL Server Profiler Template](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/Monitoring/Monitor%20Data%20Warehouse%20Query%20Activity.tdf&version=GBmaster)):
   1. events
      1. SQL: BatchCompleted
      2. SQL: StmtCompleted
   2. columns
      1. ApplicationName
      2. DatabaseName
      3. Duration
      4. EndTime
      5. LoginName
      6. Reads
      7. RowCounts
      8. SPID
      9. StartTime
      10. TextData
   3. filters
      1. DatabaseName LIKE %NorthwindDW%

## Monitoring Power BI Report Server

Objects counters names:

1. MSRS 2016 Web Service
2. MSRS 2016 Windows Service
3. ReportServer:Service

[Monitoring Report Server Performance](https://learn.microsoft.com/en-us/sql/reporting-services/report-server/monitoring-report-server-performance?view=sql-server-ver15)

## Workload Priorities

Table 7. SQL Server resource pools

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Minimum CPU % | Maximum CPU % | Minimum Memory % | Maximum memory % | Minimum I/O operations per second (IOPS) per disk volume | Maximum I/O operations per second (IOPS) per disk volume |
| Low Priority | 0 | 50 | 0 | 50 | 0 | 65 536 |
| High Priority | 20 | 90 | 20 | 90 | 65 536 | 0 |

Table 8. SQL Server workload groups

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Resource pool name | Importance | Maximum Requests | CPU Time (sec) | Memory Grant % | Memory Grant Time-out (sec) | Degree of Parallelism |
| User Queries | Low Priority | Low | 10 | 50 | 50 | 20 | 1 |
| ETL | High Priority | High | 100 | 80 | 80 | 30 | 4 |

Conditions for dbo.fn\_classify:

IF USER\_NAME IN

AzPipelineAgent

SQLAGENT

RDLexec

THEN Workload group “ETL”

ELSE Workload group “User Queries”

# Operating BI Solution

## ETL Operations

### Deploying and Configuring Packages

Table 9. Deployment model.

|  |  |
| --- | --- |
| Deployment Model | Project |
| Storage | SSIS catalog |
| Dynamic configuration | Maps environment variables in the SSIS catalog to project-level parameters and connection managers. |
| Troubleshooting | Logs events are and saves them to the catalog automatically. It can be viewed by using built-in reports, reports in [SSRS project Monitoring](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/NorthwindPBIRS/Monitoring), txt file in file system, views such as **catalog.executions** and **catalog.event\_messages** and tables in database Logs dbo. Lineage. |

Table 10. SSIS project parameters.

|  |  |  |
| --- | --- | --- |
| Project parameter | Data Type | Description |
| BackupFilesPath | String | Path for backup files. Used in DWH maintenance package to create backup for Data Warehouse. |
| DBFilesPath | String | Default path for new database files which are created in fact load packages. |
| DQSDatabaseName | String | Name of staging database where writing data after DQS cleansing. |
| DQSServerName | String | Name of server with staging database where writing data after DQS cleansing. |
| DWHDatabaseName | String | Name of Data Warehouse database. |
| DWHServerName | String | Name of Data Warehouse server. |
| EndLoadDate | String | Parameter for testing incremental load |
| ExternalFilesPath | String | Path for files needed for ETL process:   1. Data files (bcp) 2. Logs in txt format 3. Dimensional rows with no change 4. Maintenance SQL scripts |
| LogsDatabaseName | String | Name of database used for storing SSIS logs and Data Warehouse metadata |
| LogsServerName | String | Name of server with database used for storing SSIS logs and Data Warehouse metadata |
| MDSDatabaseName | String | Name of database used for Master data services. |
| MDSServerName | String | Name of server with database used for Master data services. |
| RetrainWeeks | Int16 | Storage depth for Data Warehouse backups in weeks. |

### Environments and Variables

To change execution context in easy way it is recommended to create environment with variables mapped to project-level parameters and to connection manager properties. With a release pipeline in Azure DevOps creates environment “Release”.

Project-level parameters are listed in Table 10.Additionally, it is necessary to create:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Data type | Description | Mapping |
| XMLCalendarFolder | String | Path to folder in file system with local copy of xmlcalendar repository | XMLCalendarFolder package-level parameter in the xmlcalendar package |
| OLTPNorthwindPassword | String, sensitive | Password to data source server. | Project-level connection OLTP\_ADO\_NET |

### Considerations for Package Execution

#### Timing of package execution

Every night load new portion of data.

Every first night between Friday and Saturday in month starts optimization partitions of fact tables.

Every night between Saturday and Sunday is created Full backup.

#### Execution account

SQL Server agent account on server with SQL Server integration services and SSIS catalog.

#### Environment

When executing the package automatically should be used environment “Release”.

### Package Execution Reports

SSIS catalog reports:

1. **Integration Services Dashboard**. This report provides a central summary that shows details of package executions. For each package execution listed in this report, you can drill into three sub-reports: Overview, All Messages, and Execution Performance.
2. **All Executions.** This report provides details of all package executions on the server, and you can filter it to show executions within a specified date range.
3. **All Connections.** This report shows details of all connections that the package used during executions, including connection strings and whether the connection failed or succeeded.
4. **All Operations.** This report shows details of all operations that occurred on the server, including package deployments, executions, and other administrative operations.
5. **All Validations.** This report shows details of all validations that SSIS performed for packages.

[SSRS project Monitoring](https://dev.azure.com/zinykov/20767/_git/Northwind_BI_Solution?path=/NorthwindPBIRS/Monitoring) reports:

1. **Monitoring**. This report provides information about loading duration, detailed information about Data Warehouse partitions, row validating statuses in DQS and MDS and errors in ETL process.
2. **SSIS log**. This report provides visual representation of sysssislog table.

### Backing Up the SSIS Catalog

## Data Warehouse Operations

Are implemented in [DWH maintenance package](#_Data_Warehouse_maintenance).

## Reporting Services Operations

## Automating Operational Tasks

### SQL agent

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Job | Steps | Description | Environment | Schedule | Notifications |
| Northwind (упрощённая) | Northwind (упрощённая) Package | Ingest data from Northwind (упрощённая) | Release | 01:00:00 AM every day | E-mail to DBA team |
| Xmlcalendar | [Xmlcalendar Package](#_Xmlcalendar) | Ingest data from xmlcalendar | Release | 01:00:00 AM on Sunday | E-mail to DBA team |
| Incremental load | [Incremental load Package](#_Incremental_load_package) | Load new portion of data | Release | 02:00:00 AM every day | E-mail to DBA team |

### Subsystems maintenance plan

#### Full backup

Steps:

1. Check database integrity task
   1. Databases:
      1. DQS\_MAIN
      2. DQS\_PROJECTS
      3. master
      4. MDS
      5. msdb
      6. ReportServer
      7. ReportServerTempDB
      8. SSISDB
   2. Include indexes
   3. Physical only
2. Backup database task
   1. Databases:
      1. DQS\_MAIN
      2. DQS\_PROJECTS
      3. master
      4. MDS
      5. msdb
      6. ReportServer
      7. SSISDB
   2. Type: Full
   3. Backup set will expire after 14 days
   4. Destination: Disk
   5. Compression: on
3. Maintenance cleanup task
   1. Cleanup Database Backup files
   2. Age: older than 4 weeks

Schedule:

1. Occurs every week on Sunday at 1:00:00.

#### Differential backup

Steps:

1. Check database integrity task
   1. Databases:
      1. DQS\_MAIN
      2. DQS\_PROJECTS
      3. master
      4. MDS
      5. msdb
      6. ReportServer
      7. ReportServerTempDB
      8. SSISDB
   2. Include indexes
   3. Physical only
2. Backup database task
   1. Databases:
      1. DQS\_MAIN
      2. DQS\_PROJECTS
      3. MDS
      4. msdb
      5. ReportServer
      6. SSISDB
   2. Type: Differential
   3. Backup set will expire after 14 days
   4. Destination: Disk
   5. Compression: on

Schedule:

1. Occurs every day every 12 hour(s) between 0:00:00 and 23:59:59.

#### Log backup

Steps:

1. Backup database task
   1. Databases:
      1. MDS
      2. ReportServer
      3. SSISDB
   2. Type: Transaction log
   3. Backup set will expire after 14 days
   4. Destination: Disk
   5. Compression: on

Schedule:

1. Occurs every day every 30 minute(s) between 0:00:00 and 23:59:59.

1. For view xml file is needed [Data Profile Viewer](https://learn.microsoft.com/en-us/sql/integration-services/control-flow/data-profile-viewer?view=sql-server-ver15) [↑](#footnote-ref-1)
2. For every fact partition is needed dedicated disk. [↑](#footnote-ref-2)
3. Additional information about performance counters is under the link [BI performance counters (sqlshack.com)](https://www.sqlshack.com/bi-performance-counters/). [↑](#footnote-ref-3)