Northwind BI solution

Project Scope

|  |  |  |
| --- | --- | --- |
| High business value |  | * See sales with and without discounts. * See average discount by customers and geography. * See sales on holidays, weekdays and weekends. * See sales by salespersons and employee geography. |
| 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 CEO 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 | * 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. |

# BI Topology

## Вариант установки на одном сервере



* SQL Server ядро СУБД
  + Хранилище данных
  + Staging база данных
  + SSIS база данных
  + База данных cервера отчётов
  + MDS база данных
  + DQS база данных
* SQL Server Integration Services
* SQL Server Analysis Services
  + Табличный режим
  + Многомерный режим
* Сервер отчётов Power BI
* Локальный шлюз данных
* Master Data Services
* Data Quality Services

SRVBI



Клиентский ПК

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

Заметки:

1. Минимальные требования к лицензиям на серверное оборудование и программное обеспечение, но серверу потребуются значительные ресурсы памяти, ЦП и диска для всех, кроме самых легких рабочих нагрузок бизнес-аналитики.
2. Диапазон различных типов рабочих нагрузок на сервере затруднит правильное определение и настройку аппаратных ресурсов.
3. Сервер может быть кластеризован для обеспечения высокой доступности.

## Вариант с выделенным сервером хранилища данных



* SQL Server ядро СУБД
  + Хранилище данных
  + Staging база данных
  + SSIS база данных
  + MDS база данных
  + DQS база данных
* SQL Server Integration Services
* Локальный шлюз данных
* Master Data Services
* Data Quality Services

SRVDWH



Клиентский ПК

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



* SQL Server ядро СУБД
  + База данных сервера отчётов
* SQL Server Analysis Services
  + Табличный режим
  + Многомерный режим
* Сервер отчётов Power BI

SRVBI

Заметки:

1. Использование выделенного сервера хранилища данных упрощает управление рабочими нагрузками ETL и хранилища.
2. Один или оба сервера должны быть кластеризованы для обеспечения высокой доступности.

## Вариант с выделенным сервером отчётов



* SQL Server ядро СУБД
  + Хранилище данных
  + Staging база данных
  + SSIS база данных
  + MSD база данных
  + DQS база данных
* SQL Server Integration Services
* Локальный шлюз данных
* Master Data Services
* Data Quality Services

SRVDWH



Клиентский ПК

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



* SQL Server ядро СУБД
  + База данных   
    сервера отчётов
* Сервер отчётов Power BI

SRVPBIRS



* SQL Server Analysis Services
  + Табличный режим
  + Многомерный режим

SRVANLYS

Заметки:

1. Рабочие нагрузки хранилища данных и ETL, отчётов и OLAP полностью разделены
2. Некоторые сервера должны быть кластеризованы для обеспечения высокой доступности.

## Вариант с распределённым сервером отчётов



* SQL Server ядро СУБД
  + Хранилище данных
  + Staging база данных
  + SSIS база данных
  + MSD база данных
  + DQS база данных
* SQL Server Integration Services
* Локальный шлюз данных
* Master Data Services
* Data Quality Services

SRVDWH



Клиентский ПК

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



* Сервер отчётов Power BI

SRVPBIRS



* SQL Server Analysis Services
  + Табличный режим
  + Многомерный режим

SRVANLYS



* SQL Server ядро СУБД
  + База данных   
    сервера отчётов

SRVPBIRSDB

Заметки:

1. Рабочие нагрузки хранилища данных и ETL, отчётов и OLAP полностью разделены.
2. Сервер отчётов использует выделенный сервер баз данных для размещения базы данных.
3. SRV-DWH, SRV-PBIRS-DB, SRV-ANLYS должны быть кластеризованы для обеспечения высокой доступности.
4. Могут быть добавлены дополнительные сервера отчётов для масштабирования.

## Вариант с выделенным сервером ETL



* SQL Server ядро СУБД
  + Хранилище данных
  + Staging база данных
  + MDS база данных
* Локальный шлюз данных
* Master Data Services

SRVDWH



Клиентский ПК

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



* Сервер отчётов Power BI

SRVPBIRS



* SQL Server Analysis Services
  + Табличный режим
  + Многомерный режим

SRVANLYS



* SQL Server ядро СУБД
  + База данных   
    сервера отчётов

SRVPBIRSDB



* SQL Server ядро СУБД
  + SSIS база данных
  + DQS база данных
* SQL Server Integration Services
* Data Quality Services

SRVETL

Заметки:

1. Рабочие нагрузки хранилища данных, ETL, отчётов и OLAP полностью разделены.
2. База данных Staging может быть размещена на сервере хранилища или ETL зависит от рабочих нагрузок.
3. Сервер отчётов использует выделенный сервер баз данных для размещения базы данных.
4. SRV-DWH, SRV-PBIRS-DB, SRV-ANLYS должны быть кластеризованы для обеспечения высокой доступности.
5. Могут быть добавлены дополнительные сервера отчётов для масштабирования.
6. SRV-ETL можно включить в группу высокой доступности AlwaysOn Availability Group.

## Вариант с выделенными аналитическими серверами



* SQL Server ядро СУБД
  + Хранилище данных
  + Staging база данных
  + MDS база данных
* Локальный шлюз данных
* Master Data Services

SRVDWH



Клиентский ПК

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



* Сервер отчётов Power BI

SRVPBIRS



* SQL Server Analysis Services
  + Многомерный режим

SRVANLYSMD



* SQL Server ядро СУБД
  + База данных   
    сервера отчётов

SRVPBIRSDB



* SQL Server ядро СУБД
  + SSIS база данных
  + DQS база данных
* SQL Server Integration Services
* Data Quality Services

SRVETL



* SQL Server Analysis Services
  + Табличный режим

SRVANLYSTAB

Заметки:

1. Рабочие нагрузки хранилища данных, ETL, отчётов и OLAP полностью разделены.
2. Многомерные (основная рабочая нагрузка на диск) и табличные (основная рабочая нагрузка на память) модели данных разделены.
3. База данных Staging может быть размещена на сервере хранилища или ETL зависит от рабочих нагрузок.
4. Сервер отчётов использует выделенный сервер баз данных для размещения базы данных.
5. SRV-DWH, SRV-PBIRS-DB, SRV-ANLYS-MD, SRV-ANLYS-TAB должны быть кластеризованы для обеспечения высокой доступности.
6. SRV-ANLYS-MD можно масштабировать с помощью доступных только для чтения баз данных служб Analysis Services.
7. Могут быть добавлены дополнительные сервера отчётов для масштабирования.
8. SRV-ETL можно включить в группу высокой доступности AlwaysOn Availability Group.

# Расчёт параметров оборудования для хранилища данных

Перед началом проектирования системы хранилища данных необходимо рассчитать maximum consumption rate (MCR) для одного ядра процессора в сочетании с ядром базы данных SQL Server.



# Проектирование хранилища данных

## Проектирование логической схемы ХД

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Бизнес-процессы | Измерения | | | | | | |
| **Время** | **Сотрудник** | **Оптовые покупатели** | **Розничные покупатели** | **Продукты** | **Регионы продаж** | **Маркетинговые Акции** |
| Оптовые продажи | Х | X | Х |  | Х | Х |  |
| Розничные продажи | Х |  |  | Х | Х |  | Х |

**Примечания:**

* Географические данные (страна, регион, город) используются в нескольких измерениях, так как при анализе продаж географические данные могут быть важны рекомендуется вынести в отдельное измерение.

### Модель данных бизнес-процесса «Оптовые покупатели»

Время

(Заказа, Доставки)

Год

Квартал

Месяц

День

Оптовый покупатель

Сотрудник

Товар

Регион продаж

Категория

Подкатегория

Товар

Регион продаж

Страна продаж

Территория продаж

Менеджер

Продавец

Тип

Магазин

Страна

Провинция

Город

Магазин

### Модель данных бизнес-процесса «Розничные покупатели»

Время

(Заказа, Доставки)

Год

Квартал

Месяц

День

Розничный покупатель

Маркетинговая акция

Товар

Категория

Подкатегория

Товар

Тип

Маркетинговая

акция

Пол

Материальный

статус

Расстояние до

магазина

Количество машин

Страна

Провинция

Город

Магазин

## Проектирование физической реализации ХД

# Проектирование ETL

# Проектирование аналитических моделей данных

# Проектирование решения для доставки бизнес-аналитики

# Мониторинг и оптимизация решения бизнес-аналитики

# Управление решением бизнес-аналитики