Adventure Works Data Warehouse

AWDataWarehouse Physical Design

# Storage

|  |  |  |
| --- | --- | --- |
| Logical Drive | Usage | Notes |
| E: | Data Warehouse system and data table filegroups | * The system tables are in a dedicated filegroup to minimize fragmentation of data files. * The data filegroup is initialized with a single file of 1GB, which should be large enough for the tables it will contain for the foreseeable future. |
| F: | Filegroups for the FactResellerSales table | * This table is likely to grow extremely large, and so it is assigned to a dedicated logical drive to avoid competing for space. |
| G: | Filegroups for the FactInternetSales table | * As above |
| H: | Staging database data files | * A separate staging 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. |
| I: | 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. |
| J: | Database log files | * The logfiles for tempDB, the Staging database, and the data warehouse are stored on this drive to isolate log activity from data I/O. |
| K: | Backup volume | * This logical drive will be used for backup file storage |
| L: | Unused |  |
| M: | Unused |  |
| N: | Unused |  |

# Partitioning

The **FactResellerSales** and **FactInternetSales** tables are partitioned on the **ShipDateKey** column. This column was chosen because reseller orders are not considered complete until shipped, so it is reasonable to assume that the ship date is used as the cut off for each month’s orders. For consistency, the same key is used for Internet sales.

Orders for the current year are partitioned into monthly partitions, as this is likely to reflect the ETL load frequency. 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 that are likely to be queried frequently.

The smaller fact tables have a clustered index on the most commonly queried date key, and non-clustered indexes on other dimension key columns.

The two larger (partitioned) fact tables both have columnstore indexes on all columns to maximize query performance.

# Compression

The larger fact tables are compressed using PAGE compression for all partitions.

# Views

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