Module 2

Planning Data Warehouse Infrastructure

Zyead Ahmed.
Aspiring To Learrning Data Engineer.

Module Overview

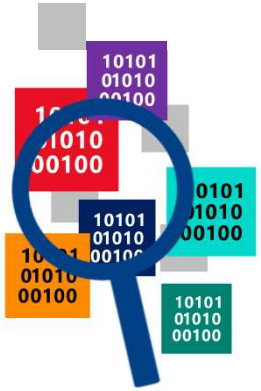
- Considerations for Data Warehouse Infrastructure Planning Data Warehouse Hardware

Lesson 1: Considerations for Data Warehouse Infrastructure

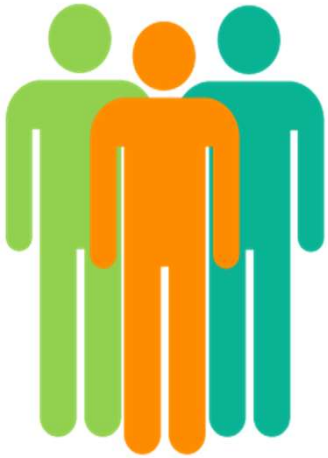
- System Sizing Considerations
  Data Warehouse Workloads
  Typical Server Topologies for a BI Solution
  Scaling Out a BI Solution
  Planning for High Availability

System Sizing Considerations

Data Volume



Analysis/Report Complexity



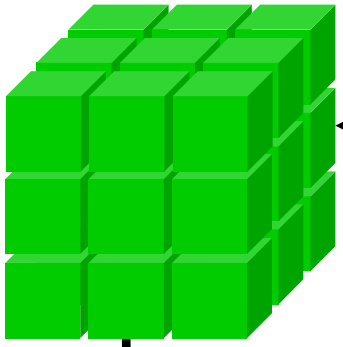
Number of Users

Availability Requirements

Data Warehouse Workloads

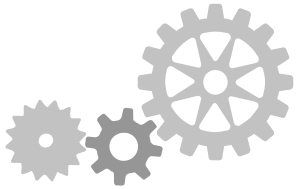
Data Models

- Processing
- Aggregation storage
  - Multidimensional on disk
  - Tabular in memory
- Query execution



ETL

- Control flow tasks
- Data query and insert
- Network data transfer
- In-memory data pipeline
- SSIS Catalog or msdb I/O



DW

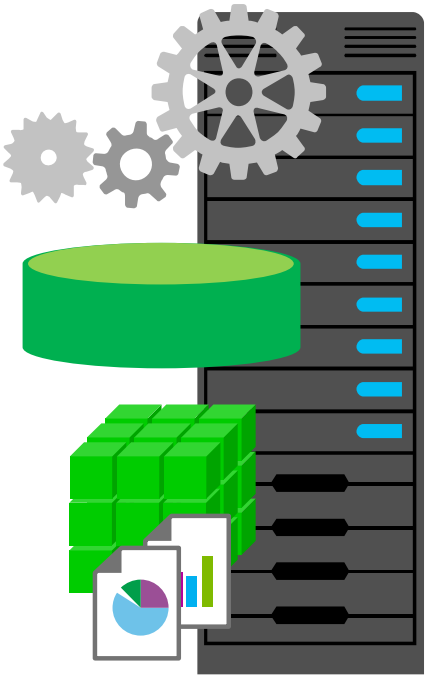
Reporting

- Client requests
- Data source queries
- Report rendering
- Caching
- Snapshot execution
- Subscription processing
- Report Server Catalog I/O

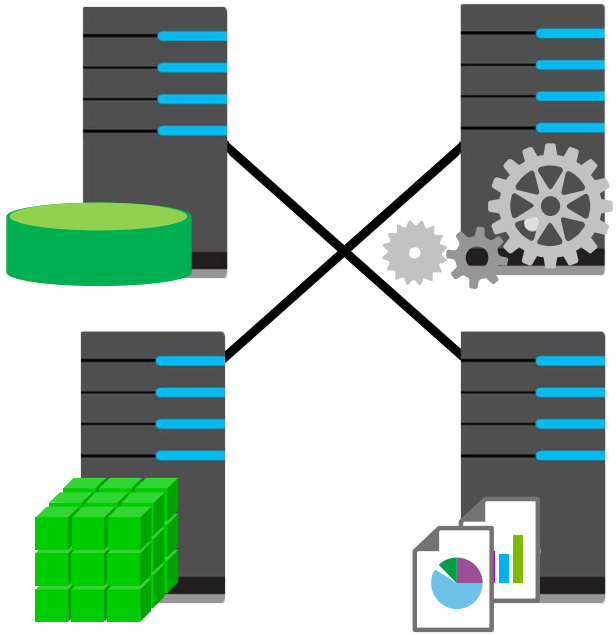
Operations and Maintenance

- OS activity
- Logging
- SQL Server Agent Jobs
  - SSIS packages
  - Indexes
  - Backups

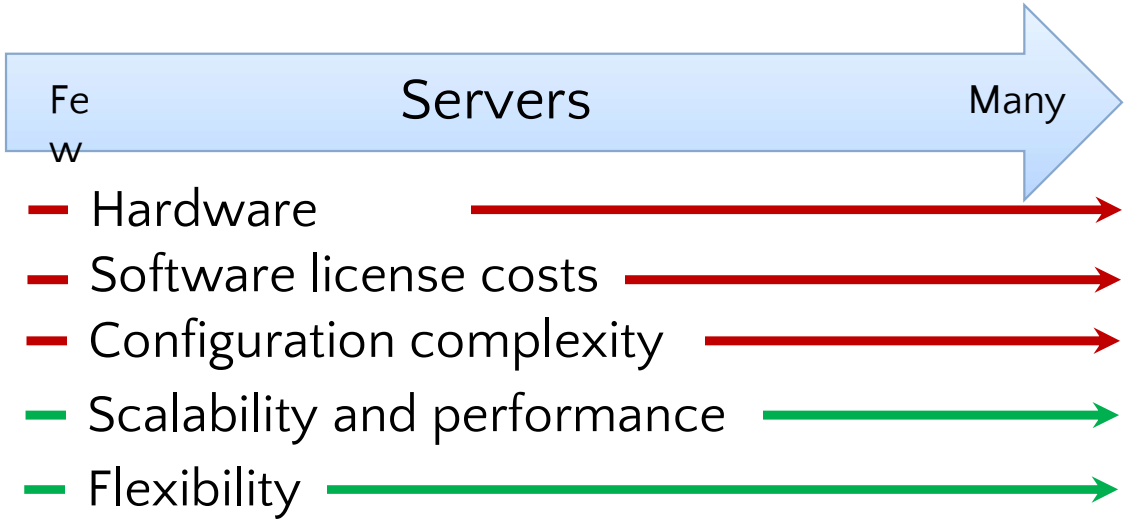
Typical Server Topologies for a BI Solution



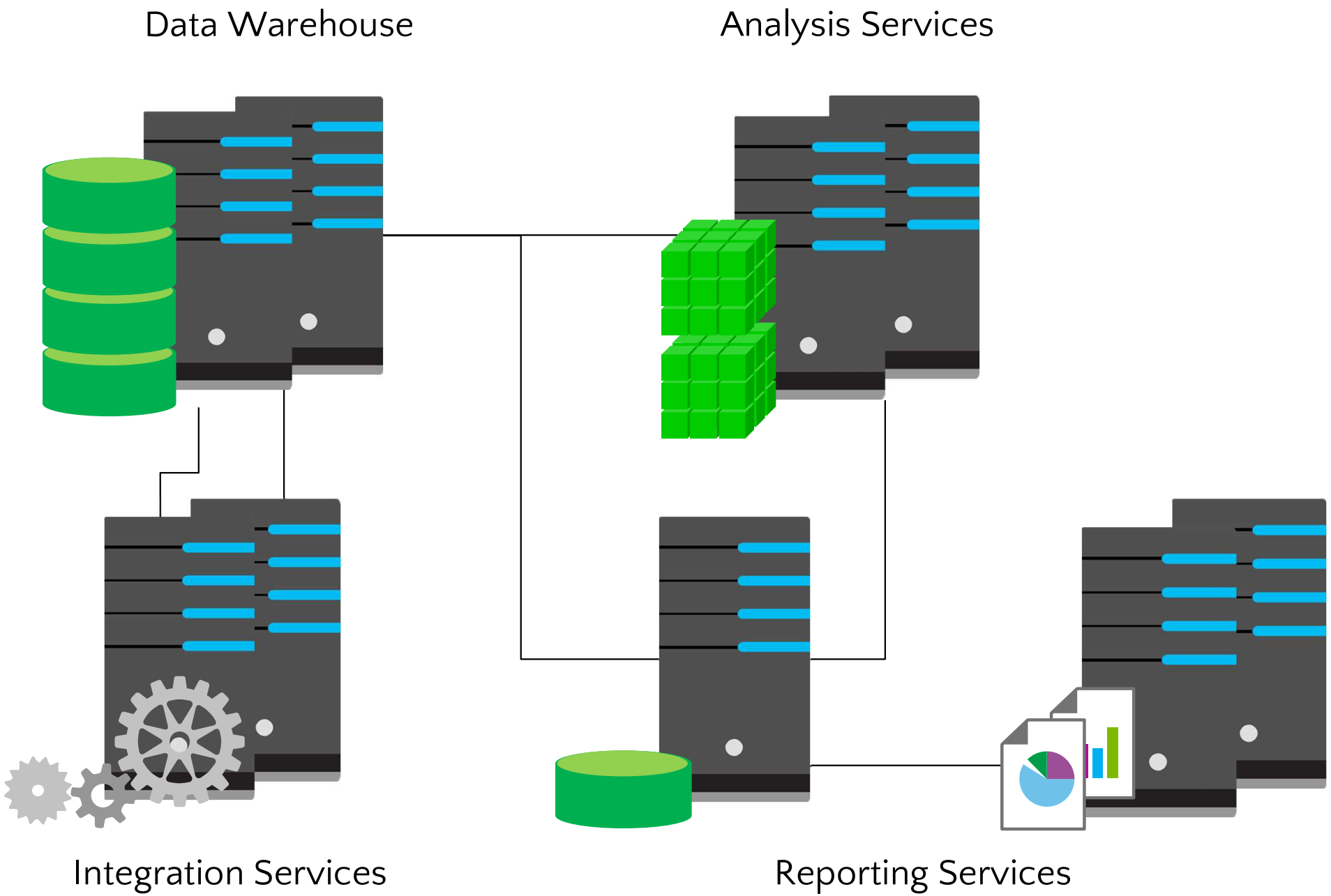
Single-server



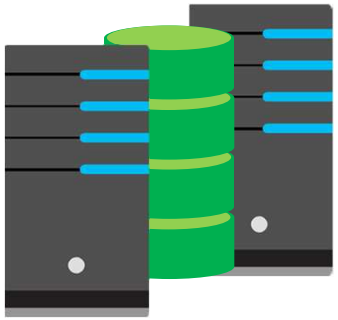
Distributed



Scaling Out a BI Solution

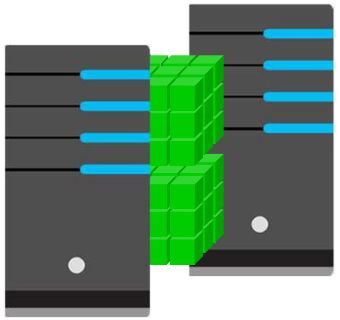


Planning for High Availability



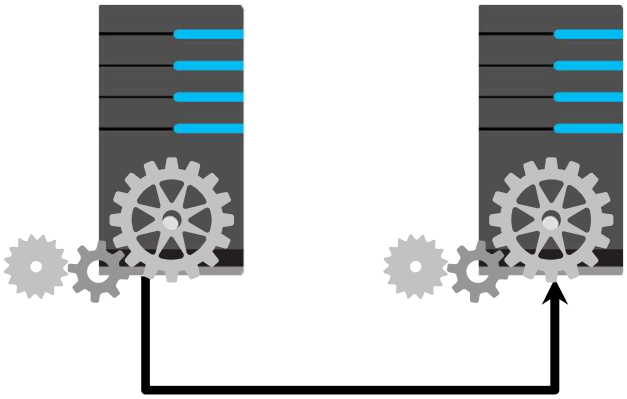
Data Warehouse

- AlwaysOn Failover Cluster
- RAID Storage



Analysis Services

- AlwaysOn Failover Cluster



Integration Services

- AlwaysOn Availability Group

Reporting Services

- NLB Report Servers
- AlwaysOn Availability Group

Or

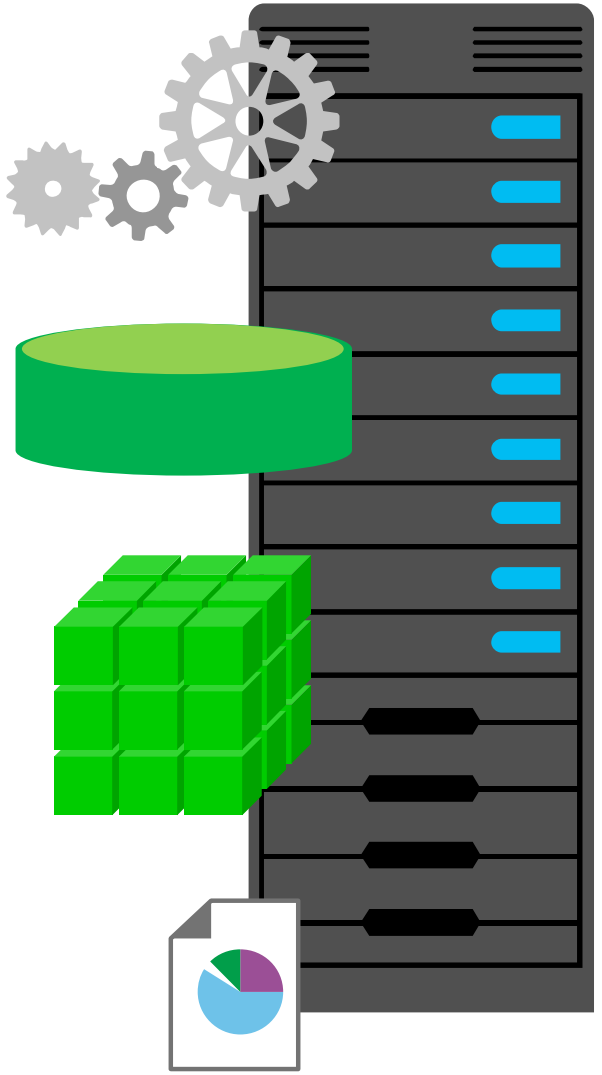
- AlwaysOn Failover Cluster

Lesson 2: Planning Data Warehouse Hardware

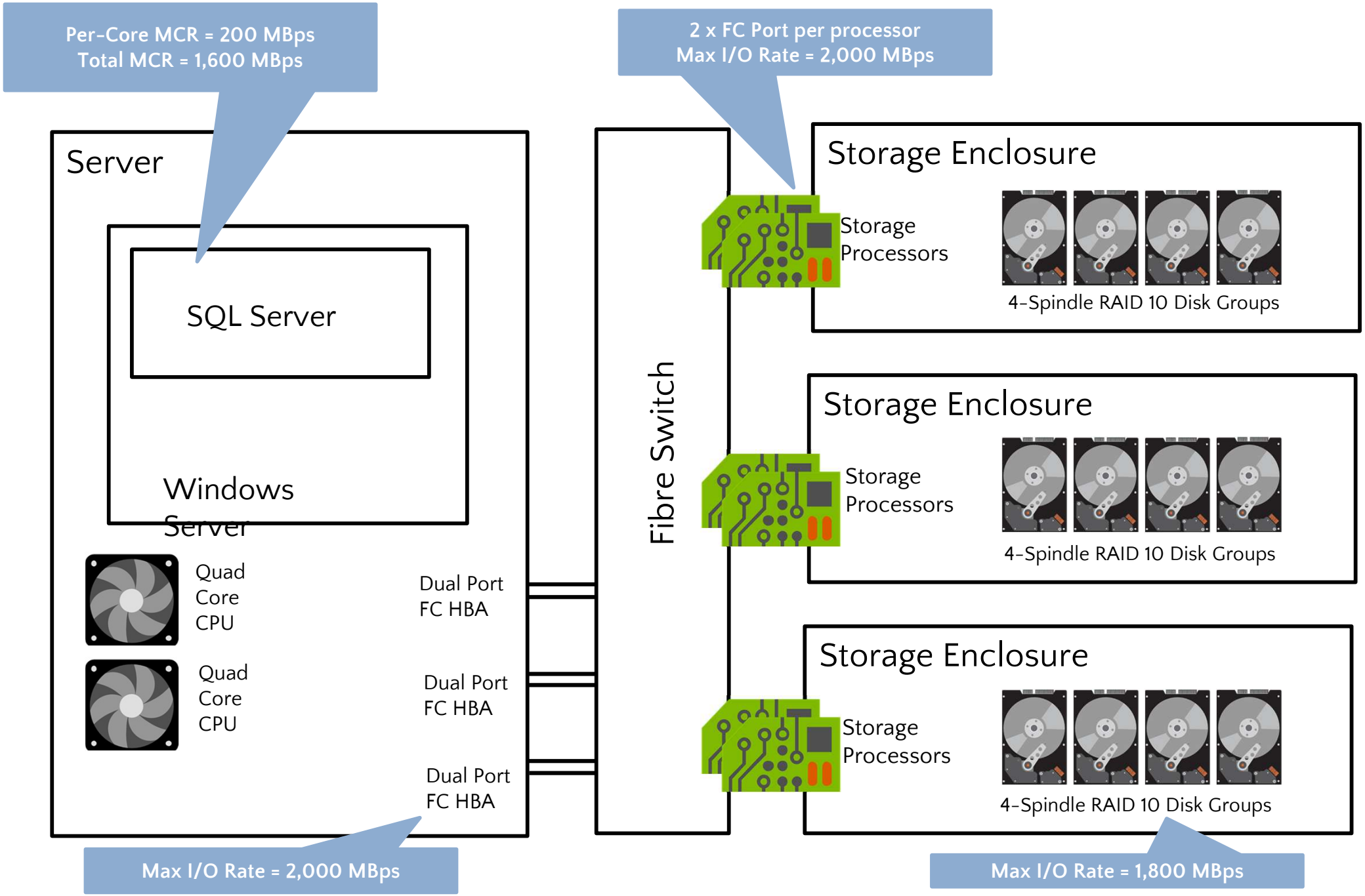
- SQL Server Fast Track Data Warehouse Reference Architectures
  Core-Balanced System Architecture
  Demonstration: Calculating Maximum Consumption Rate (MCR)
  Determining Processor and Memory Requirements
  Determining Storage Requirements
  Considerations for Storage Hardware
  SQL Server Data Warehouse Appliances
  SQL Server Parallel Data Warehouse

SQL Server Fast Track Data Warehouse Reference Architectures

- Pre-tested and approved hardware specifications and guidance
- Available from multiple hardware vendors in partnership with Microsoft
- Support for a range of data warehouse sizes
- Tools provided to calculate required specification



Core-Balanced System Architecture



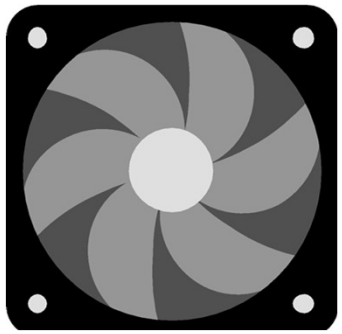
Demonstration: Calculating Maximum Consumption Rate (MCR)

In this demonstration, you will see how to:

- Create tables for benchmark queries
- Execute a query to retrieve I/O statistics
- Calculate MCR from the I/O statistics

Determining Processor and Memory Requirements

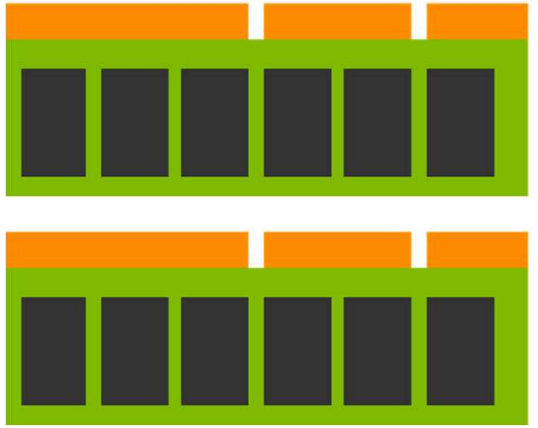
Estimating CPU Requirements:



- Determine core MCR
- Apply formula to estimate required number of cores:

  ((Average query size in MB ÷ MCR) x Concurrent users) ÷ Target response time
- Spread cores across CPUs based on the number of storage arrays

Estimating RAM Requirements:



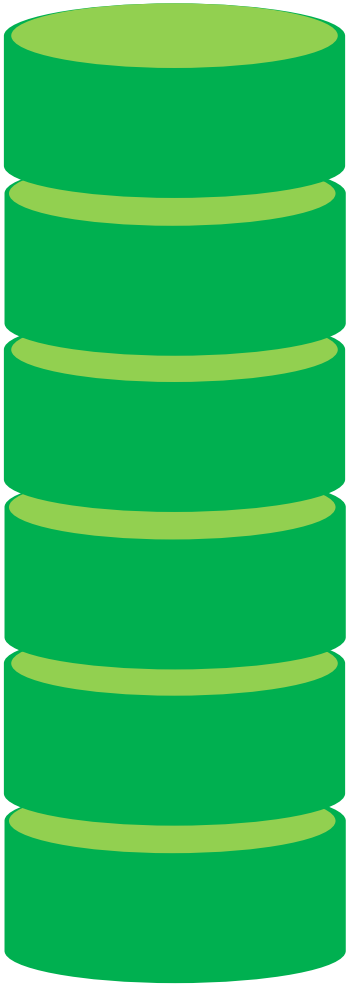
- Use a minimum of 4 GB per core (or 64–128 GB per socket)
- Target 20% of data volume

Determining Storage Requirements

Data Warehouse

Estimating Data Volumes for the Data Warehouse

1. Estimate Initial Fact Data
   - Number of fact table rows × row size
     - Use 100 bytes per row as an estimate if unknown
2. Allow for Indexes and Dimensions
   - Add 30–40% for dimensions and indexes
3. Project Fact Data Growth
   - Number of new fact rows per month
4. Factor in compression
   - Typically 3:1

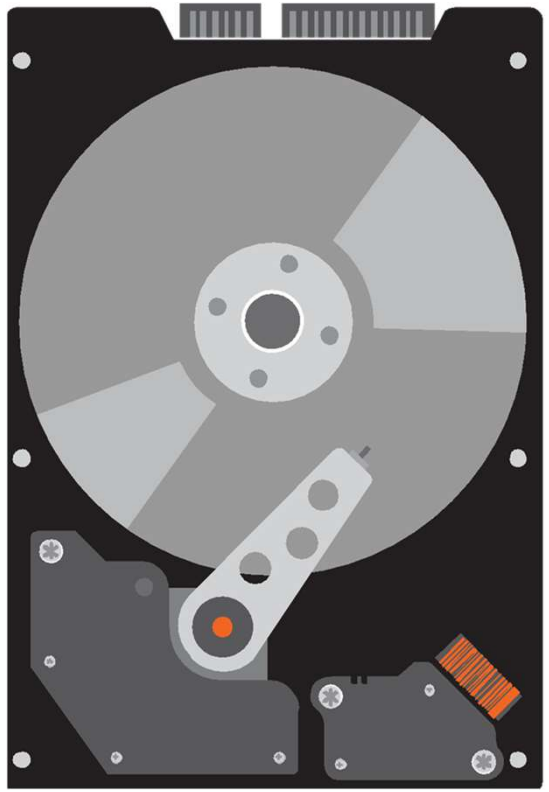


Other storage requirements

- Configuration databases
- Log files
- tempdb
- Staging tables
- Backups
- Analysis Services models

Considerations for Storage Hardware

- Use more smaller disks instead of fewer larger disks
- Use the fastest disks you can afford
  - Consider solid state disks—especially for random I/O
- Use RAID 10, or minimally RAID 5
- Consider a dedicated storage area network for manageability and extensibility
  - Balance I/O across enclosures, storage processors, and disk groups



SQL Server Data Warehouse Appliances

- Pre-built hardware and software solutions, based on tested configurations
- Part of a range of appliances that are based on SQL Server
- Available from multiple hardware vendors

SQL Server Parallel Data Warehouse

- A special SQL Server edition only available in hardware appliances
- Shared-nothing architecture
- Massively parallel processing
- Dedicated control nodes, compute nodes, and storage nodes