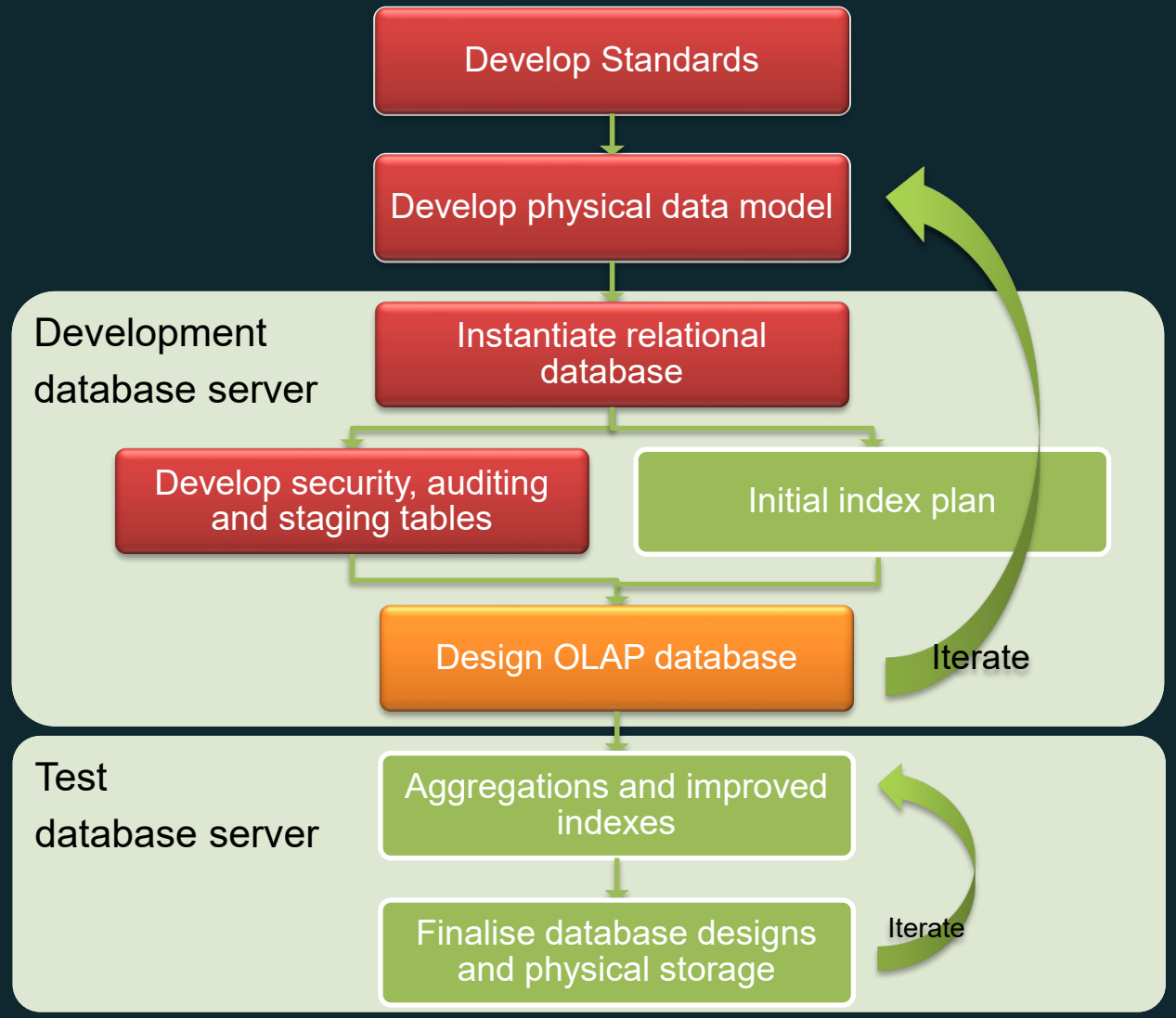
Designing the dimensional model and the physical database

Part 2

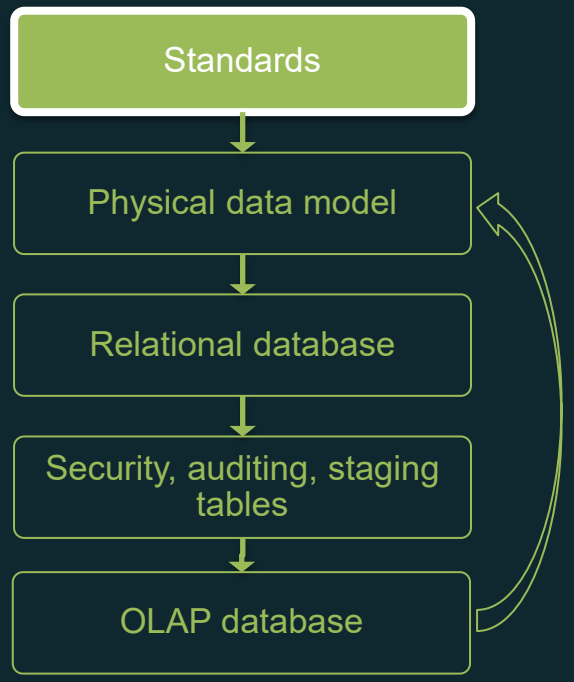
Physical database design process

- Logical to physical design
- General guidelines:
  - Plan & tie to project plan
  - No temporary solutions
  - Follow standards
  - Use tools

Develop for performance!
Do it right the first time!



Dimensional Models – Logical or Physical?



Develop system-wide standards:

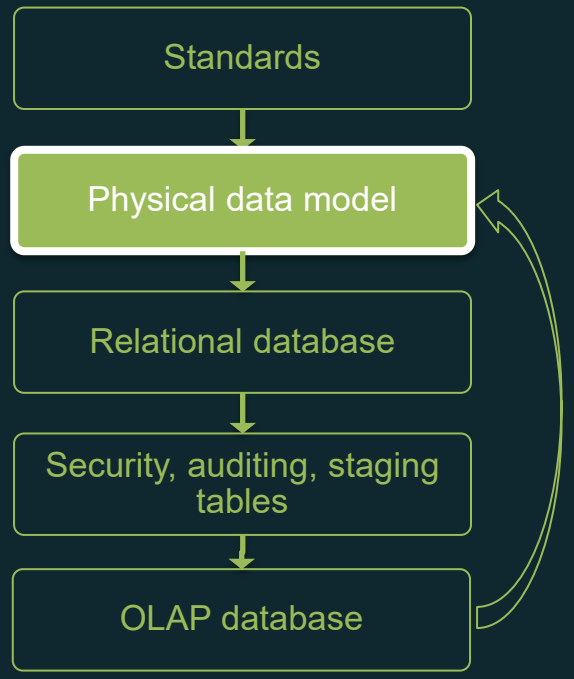
1. Naming conventions
   - ✓ Consistent style
   - ✓ Descriptive
   - ✓ User-oriented

custVatNo ✗

Customer VAT Number ✓

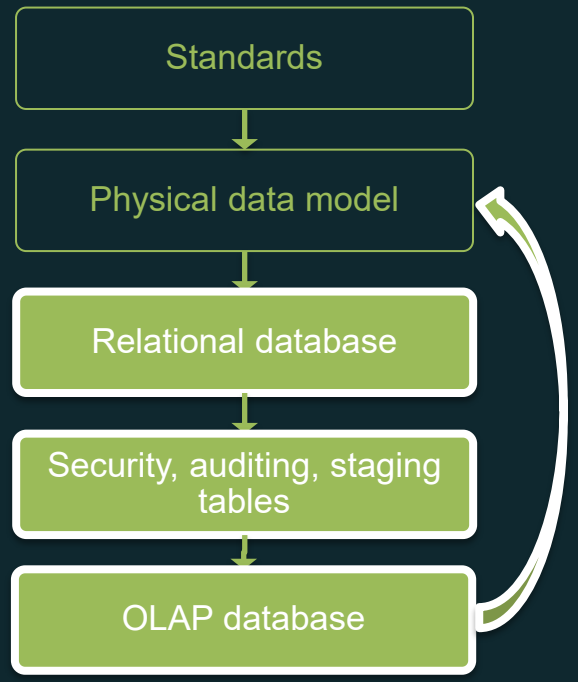
2. Avoid NULL values in descriptive fields
   - ✓ Replace with *Unknown*, *No value*, *Not applicable*, etc.
3. Staging tables – separate db/schema
4. File location standards – base structure on DW functions
5. Use synonyms and views for users – more flexible
6. Primary keys
   - ✓ integer surrogate PKs for dims

   SQL Server: IDENTITY keyword
   - ✓ composite key = subset of the FKs for fact tables
7. Foreign keys
   - ✓ Best to declare and enforce (RI & query optimisation)
   - ✓ Can consider drop before a big load



Develop the physical database model:

1. Design physical data structure
   - Based on the logical model
   - Create scripts to create db objects -> part of documentation
2. Finalise source-to-target maps
   - For each table and column
   - Standards, views, descriptions, valid ranges, default values, audit and maintenance columns, transformations, ….
3. Star vs. snowflake – prefer flat star schema
   - Simple + fast **+ easier to manage in the ETL process**
4. Use a data modelling tool (e.g. SQL server design tools)
   - Generate metadata & object definition scripts
5. Initial sizing estimates
   - Fact tables will take up most space -> don't try to compromise dimension details



Build the development database

Star schema & Processing data stores:

Security tables

Access monitoring tables

Dimension and fact tables

With views and other related objects

Staging tables

to support the ETL system

Auditing tables

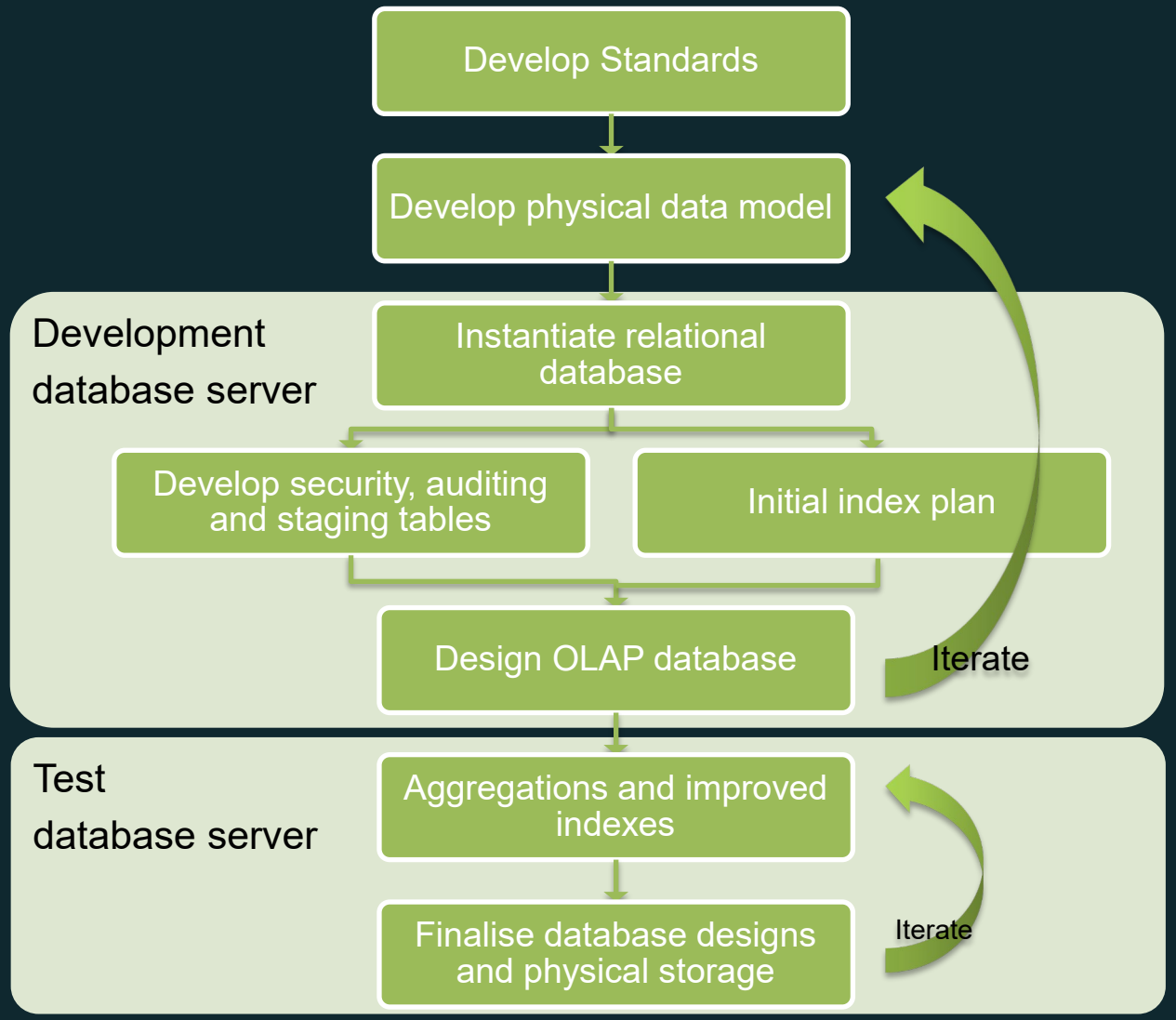
for ETL processing and data quality

Recap – physical design process

Describe the high level physical design process

Discuss the development of standards for the DW/BI system

Develop a physical data model and database



Notes/tips on physical design & ETL for project

- SQL Server Management Studio settings:
  - Tools -> Options -> Designers:
  - Remove check for “Prevent saving changes that require table re-creation”
- Database properties (Options):
  - Recovery model: Simple
  - Collation -> select case-sensitive (CS) and accent-sensitive (AS)
  - For example Latin1_General_CS_AS
- Use Unicode strings
  - nchar and nvarchar, as well as Unicode strings in SSIS [DT_WSTR]
- Set ETL input/source field names equal to target field names
  - If input names are equal to target names, mappings are automatic.
- Retain null values
  - Check these options in SSIS
  - Explicitly replace null values with Unknown or defined default values in cleaning
- Naming standards and abbreviations
  - Explicitly define and consistently use entity name abbreviations in staging tables

Thank you