BUILDING THE DATA WAREHOUSE - PART 1

MODULE 10

1

OVERVIEW

Data Warehouse Overview

2

DATA PROCESS

Overview of the Data Process

3

DATABASE

Building the Azure Sql Database

4

STAGING MASTER DATA

Building the Staging Layer Master Data

5

STAGING TRANSACTIONS

Building Staging Layer Transactional Data

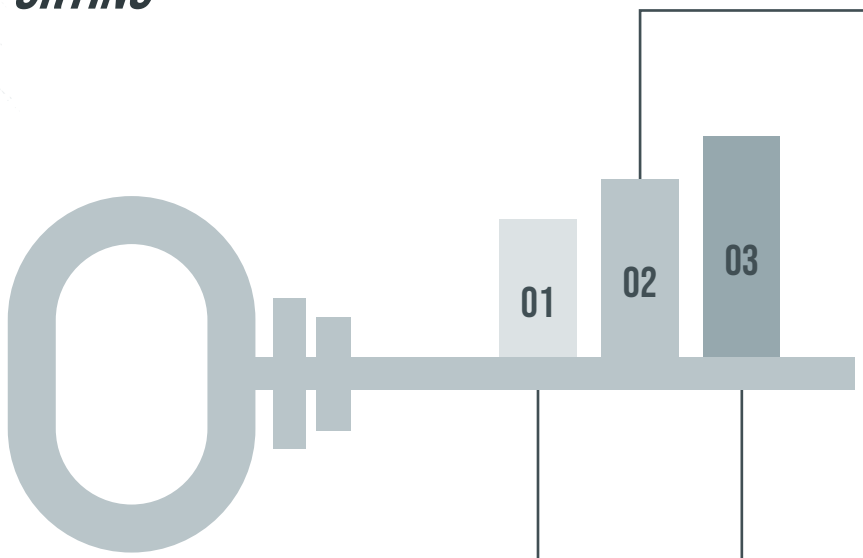
SECTION 1

OVERVIEW

OVERVIEW

Data Warehouse Overview

SINGLE SOURCE OF TRUTH TO STORE DATA FROM MULTIPLE SOURCES FOR HISTORICAL AND ANALYTICAL REPORTING



DATA QUALITY

One version of the truth

Improve data quality from source systems

Provide business friendly naming conventions

DATA INTEGRATION AND ISOLATION

Integrate many sources of data

Isolate from production transaction systems

Protect from source system upgrades

DATA ANALYSIS

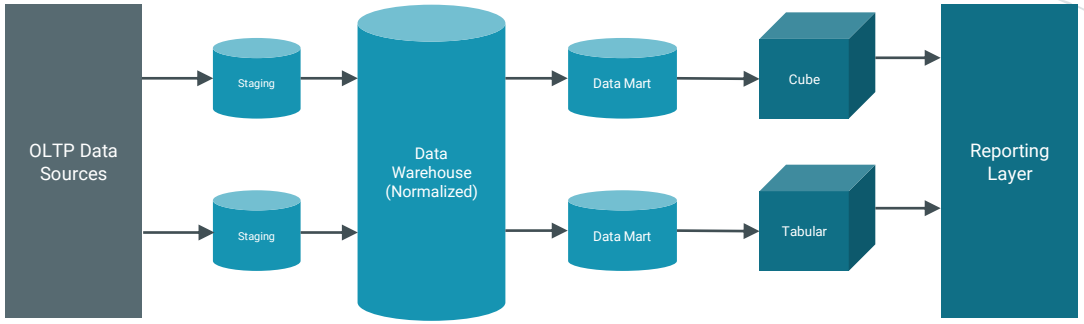
Optimized for read access

Build hierarchies to analyze data

Enable self-serve data analysis for business users

OVERVIEW

Data Warehouse Models



INMON MODEL

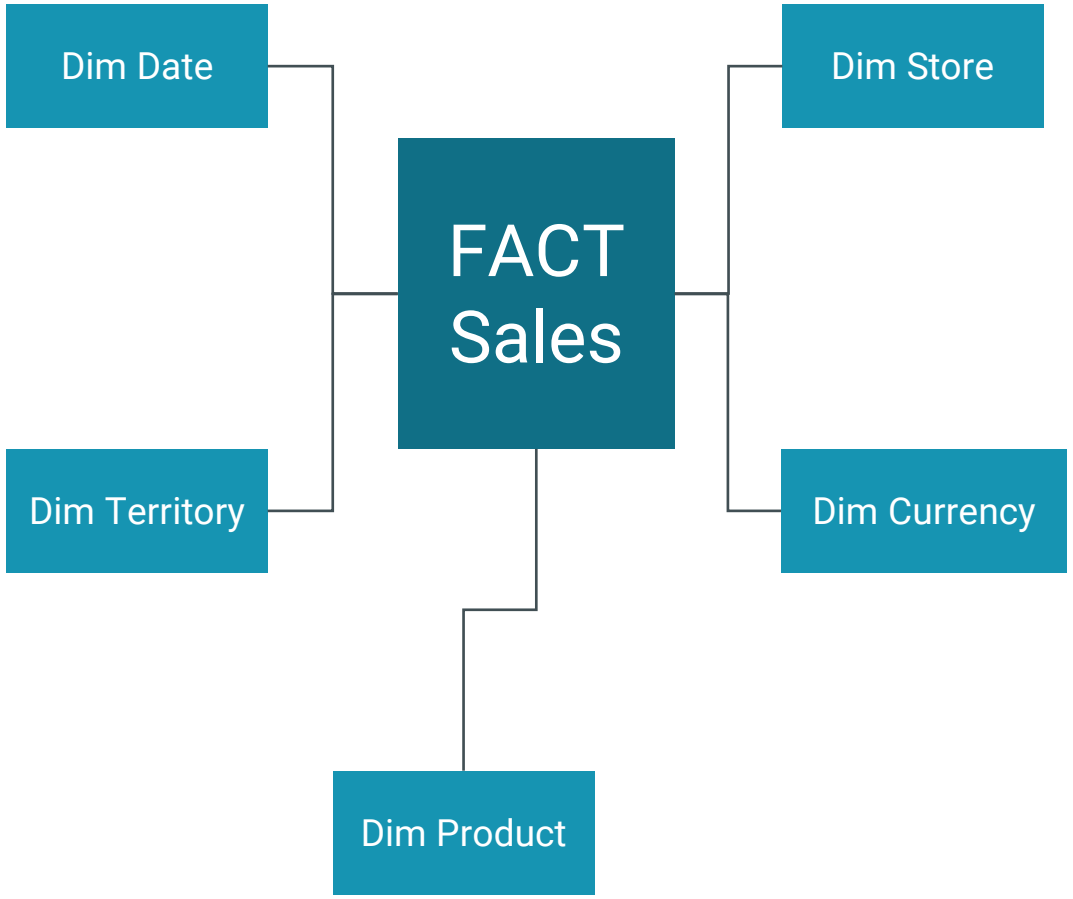
- Enterprise data model (CIF) or enterprise data warehouse (EDW)
- IT Driven, users have passive participation
- Centralized atomic normalized tables
- Dependent data marts that are separate physical subsets of EDW

- Logical data warehouse with subject area - data marts
- Business driven, users have active participation
- Decentralized data marts
- Independent dimensional data marts for reporting/analytics

KIMBALL MODEL

OVERVIEW

Data Warehouse Vino World



STAR SCHEMA DATA MODEL

- Facts and Dimensions, star schema
- Denormalized Fact Sales Table with Measures
- Integrated via Conformed Dimensions providing consistency across data sources
- Slowly changing dimensions with surrogate keys
- Business friendly for direct end-user data access

WHAT QUESTIONS CAN BE ANSWERED?

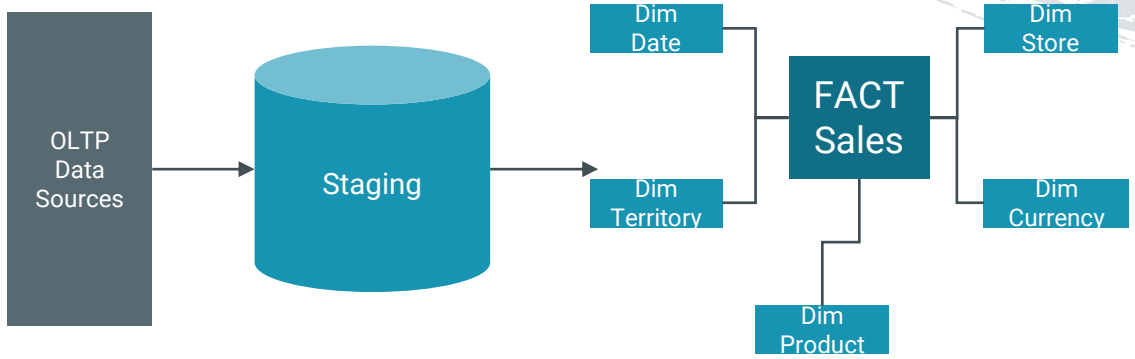
- What are the total sales?
- What is the gross profit?
- What are the total sales by store?
- What are the total sales by product?
- What are the total sales by Territory?
- Which month had the highest sales and for which product?
- Which product is the most profitable?

SECTION 2

DATA PROCESS

DATA PROCESS

Overview of the Data Process



DATA PROCESSING

- Integrate multiple data sources
- Stage the data
- Scrub the data due to data quality issues
- Transform the data and load into the Data Warehouse

EXTRACT, TRANSFORM, AND LOAD (ETL)

- No Staging Tables
- Transform the data while hitting the source system
- Processing done by the ETL Tools

EXTRACT, LOAD, AND TRANSFORM (ELT)

- Uses Staging Tables
- Transform the data while loading into target system
- Processing done by target database engine

ELT is better suited for large volumes of data and for a modern data warehouse architecture

DATA PROCESS

Overview of the ELT Process

1

EXTRACT

- Extract source files to raw zone
    - Transaction files
    - Master data files

2

LOAD

- Load transaction/master files to cleansed zone
- Scrub files for data quality issues
- Harmonize data for consistent format
- Load cleansed data into Staging Layer

3

TRANSFORM

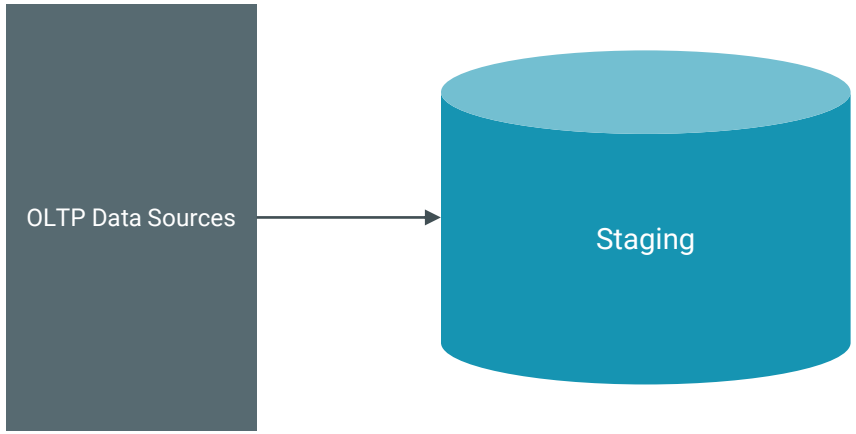
- Transform data to dimensional model
- Load to data warehouse
- Build dimensions
- Build Facts

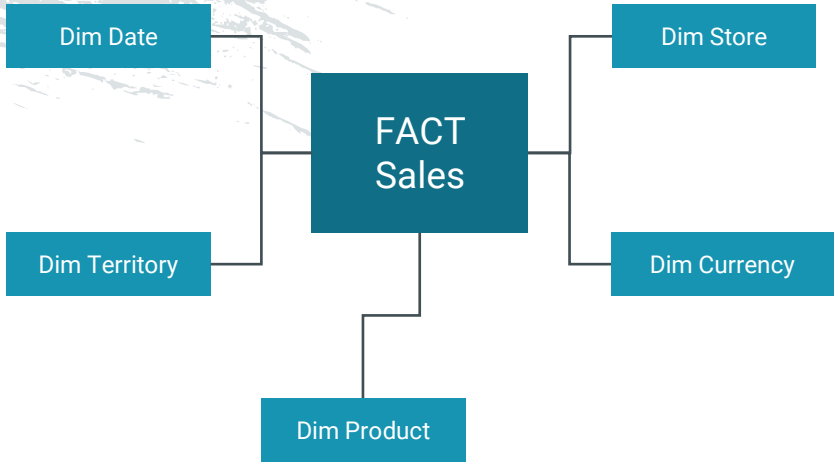
SECTION 3

DATABASE

DATABASE

Overview of Database Objects





STAGING

- stage.Dates
- stage.Currency
- stage.Product
- stage.Store
- Stage.Territory
- stage.Verde_Products
- stage.Arancione_Products
- stage.Celeste_Products
- stage.Verde_Sales
- Stage.Arancione_Sales
- stage.Celeste_Sales
- stage.Sales

DATA WAREHOUSE

- dimDate
- dimCurrency
- dimProduct
- dimStore
- dimTerritory
- factSales

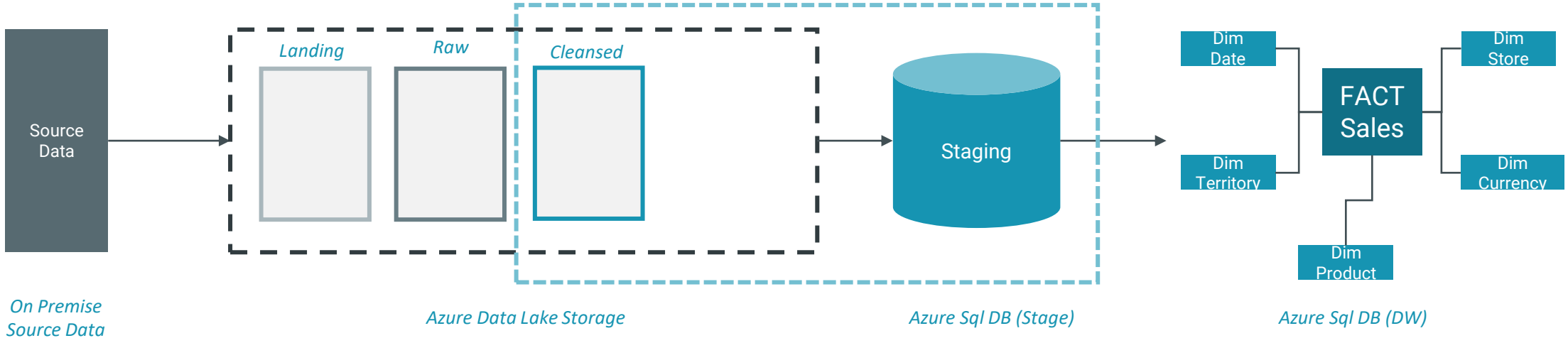
We will build staging tables and data warehouse tables in Azure Sql DB

SECTION 4

STAGING MASTER DATA

STAGING MASTER DATA

Building the Staging Layer Master Data



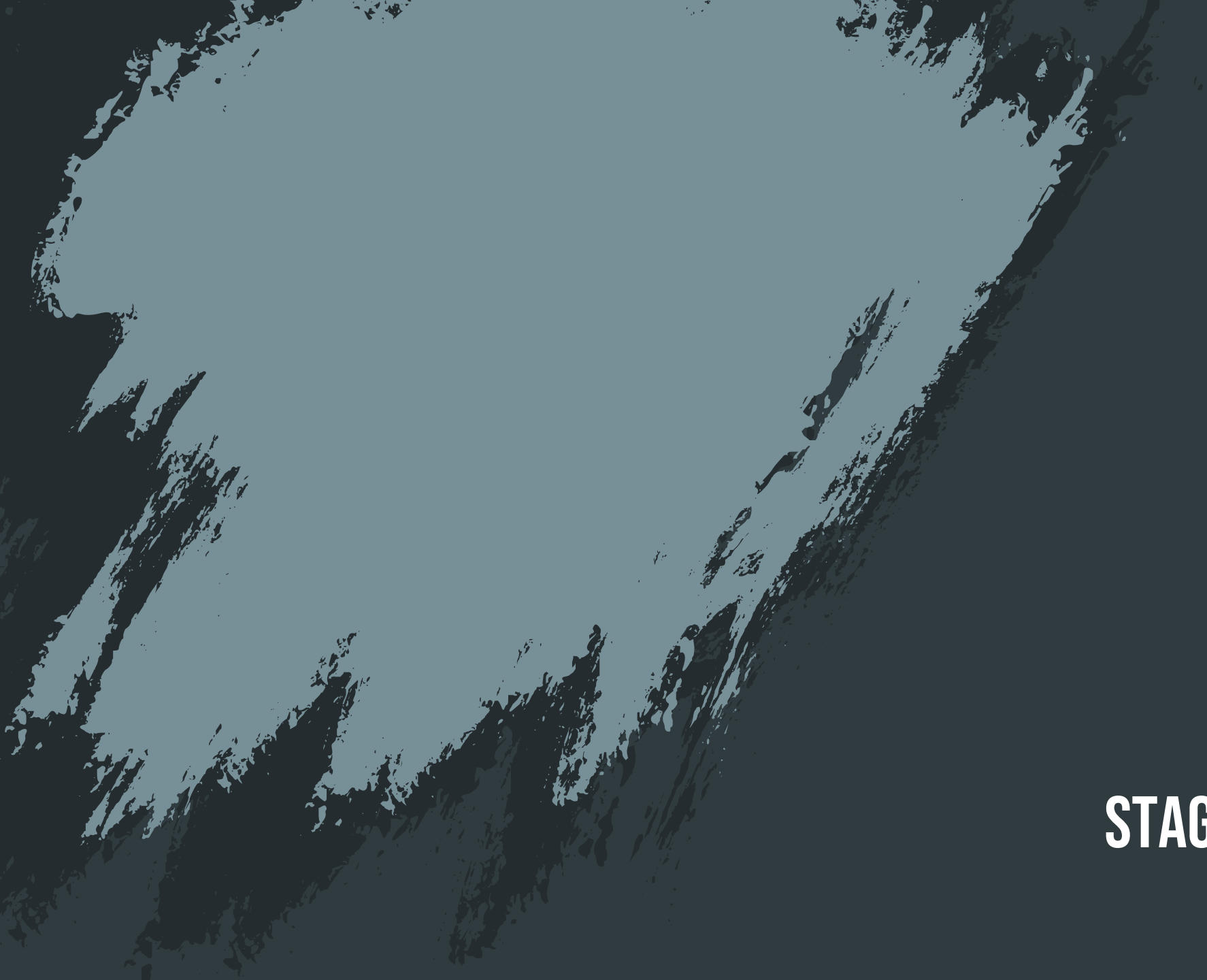
SOURCE FROM ADLS

- Source scrubbed data from cleansed zone
- Master Data

LOAD TO STAGE – AZURE SQL DB

- Load Master data from ADLS cleansed zone
- Load data into stage tables in Azure Sql DB

In this step we will load master data from the cleansed container in Azure Storage to stage tables in Azure Sql DB

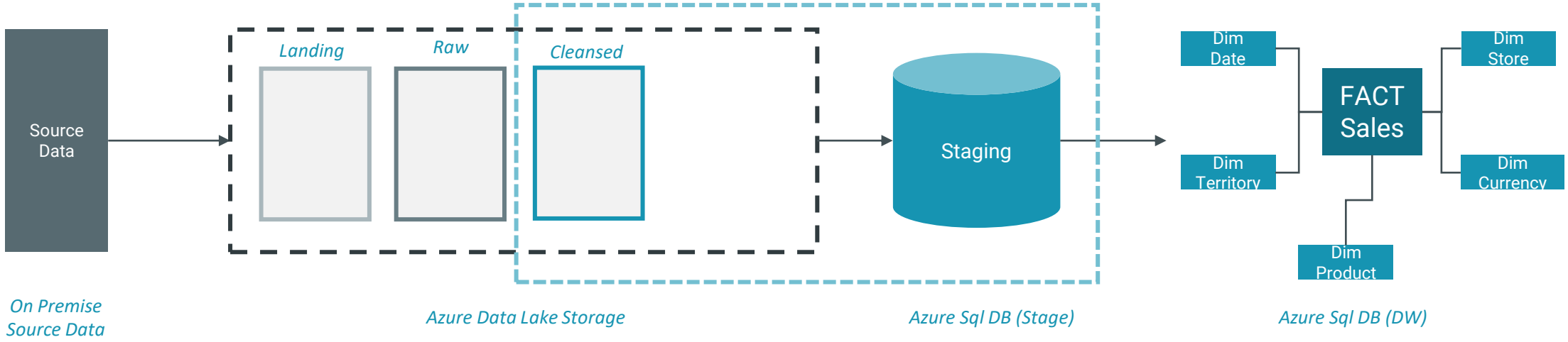


SECTION 5

STAGING TRANSACTIONS

STAGING TRANSACTIONS

Building the Staging Layer Transactions



SOURCE FROM ADLS

- Source scrubbed data from cleansed zone
- Transaction Data

LOAD TO STAGE – AZURE SQL DB

- Load Transaction data from ADLS cleansed zone
- Load data into stage sales table in Azure Sql DB

In this step we will load transaction data from the cleansed container in Azure Storage to stage tables in Azure Sql DB

MODULE SUMMARY

In this module we learnt

OVERVIEW

We got an overview of the Data Warehouse and its benefits.

We learnt about the different Data Warehouse approaches and data processing approaches

INTEGRATION

We learnt about the Star Schema Model we will use for our Data Warehouse

We learnt about the ELT approach that we will use to build the Data Warehouse

HANDS-ON

We learnt how to build the Azure Sql database objects

We then learnt how to stage our master data and transactional data

REFERENCES

Dimensional Modeling – Ralph Kimball

A Dimensional Modeling Manifesto - Kimball Group

ETL/ELT

From Warehouse To Lakehouse – ELT/ETL Design Patterns With Azure Data Services – SQL Of The North

Modern Data Warehouse

The Modern Data Warehouse | James Serra's Blog