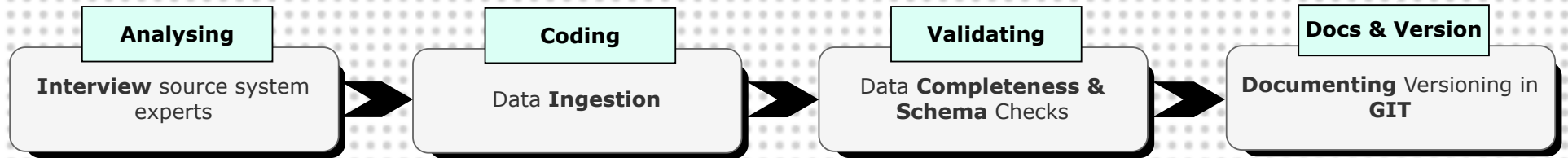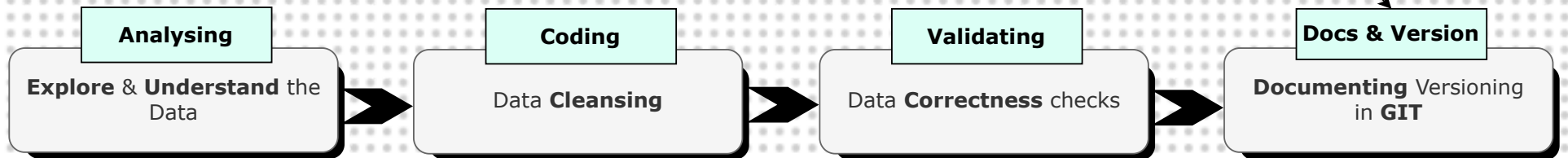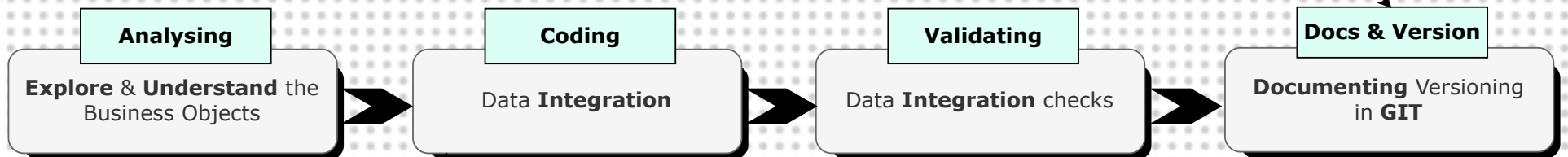| | **Bronze Layer** | **Silver Layer** | **Gold Layer** |
|---|---|---|---|
| **Definition** | Raw, unprocessed data as is from sources | Clean & Standardize Data | Business Ready Data |
| **Objective** | Traceability & debugging | (intermediate Layer) Prepare Data for Analysis | Provide Data to me consumed for reporting and analytics |
| **Objective Type** | Tables | Tables | Views |
| **Load Method** | Full Load (Truncate & Insert) | Full Load (Truncate & Insert) | None |
| **Data Transfomation** | None (as-is) | - Data Cleaning<br>- Data Standardization<br>- Data Normalization<br>- Derived Columns<br>- Data Enrichment | - Data Integration<br>- Data Aggregation<br>- Business Logoc & Rules |
| **Data Modeling** | None (as-is) | None (as-is) | - Star Schema<br>- Aggregated Objects<br>- Flat Tables |
| **Target Audience** | - Data Engineers | - Data Analyst<br>- Data Engineers | - Data Analyst<br>- Business Users |

## Bronze Layer

**Analysing**
**Interview** source system experts

**Coding**
Data **Ingestion**

**Validating**
Data **Completeness & Schema** Checks

**Docs & Version**
**Documenting** Versioning in **GIT**

## Silver Layer

Data Flow
Data Integration

**Analysing**
**Explore** & **Understand** the Data

**Coding**
Data **Cleansing**

**Validating**
Data **Correctness** checks

**Docs & Version**
**Documenting** Versioning in **GIT**

## Gold Layer

Data Model
Data Catalog
Data Flow

**Analysing**
**Explore** & **Understand** the Business Objects

**Coding**
Data **Integration**

**Validating**
Data **Integration** checks

**Docs & Version**
**Documenting** Versioning in **GIT**

**Build** the **Business Object**

**Choose** Type **Dimension** vs **Fact**

**Rename** to **Friendly** Names
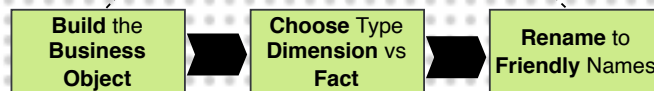
# Source System Interview

## Business Context & Ownership

Who **owns** the **Data?**

What **Business Process** it supports?

System & Data **Documention?**

Data **Model** & Data **Catalog**

## Architecture & Technology Stack

How is data **stored**?
(SQL Server, Oracle, AWS, Azure,...)

What are the **integrsation capabilities**?
(API, Kafka, File Extract, Direct DB,...)

## EXtract & Load

**Incremental** vs. **Full Loads**?

Data **Scope** & **Historical** Needs

What is the expected **size** of the **extracts**?

Are there any data **volume limitations**?

How to **avoid impacting** the source system's **performance**?

**Authentication** and **Authorization**
(tokens, SSH keys, VPN, IP whitelisting, …)