|  |
| --- |
| EPAM Systems, RD Dep. |
| MTN.BI.08 Load and Transformation |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| REVISION HISTORY | | | | | |
| Ver. | Description of Change | Author | Date | Approved | |
| Name | Effective Date |
| 1.0 | Initial status | Hanna Hul | 16-NOV-2017 |  |  |

Contents

[1. Data Modelling Task 3](#_Toc444190110)

[1.1. Detail your diagrams for 3NF and Star layers. 3](#_Toc444190111)

[1.2. Create visual and textual description of layers of your data warehouse. 3](#_Toc444190112)

[2. Data Vault 3](#_Toc444190113)

# Data Modelling Task

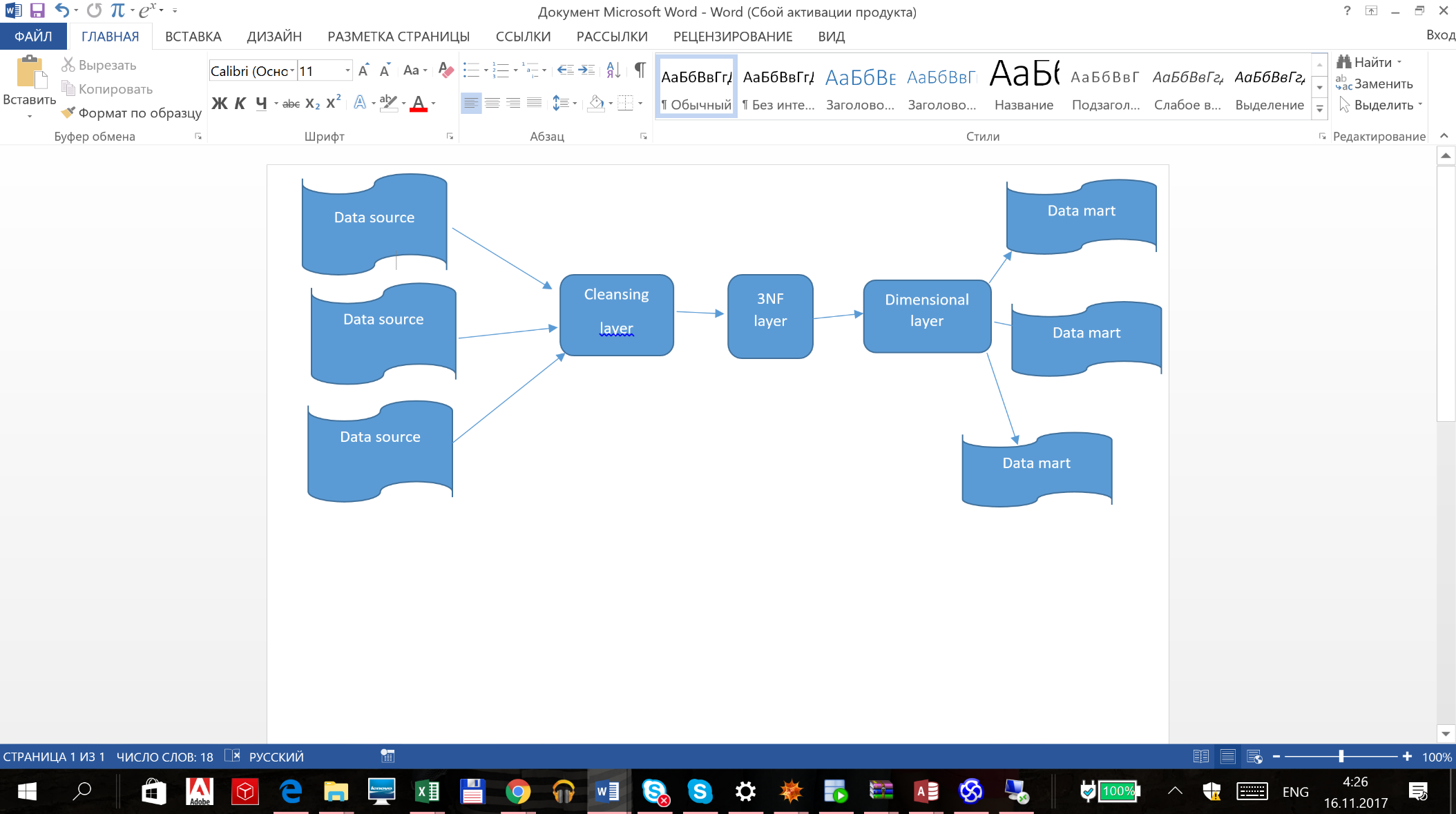
## Detail your diagrams for 3NF and Star/snowflake layers.

Reminder:

* You need to have at least 5 dimensions.
* You need to use at least 2 type SCD2 attributes in different dimensions.

Create visual and textual description of layers of your data warehouse.

* Image (PowerPoint or whatever tool do you prefer) of your data warehouse layers.



1. **Source layer.** A data warehouse system uses heterogeneous sources of data. That data is originally stored to corporate relational databases or legacy databases, or it may come from information systems outside the corporate walls.
2. **Cleansing Layer .**The Cleansing Layer is used for data cleansing, filtering wrong data, replace missing values with singletons and performing transformations like code lookups or currency conversions. As the Staging Area, the Cleansing Area contains only data of the last delivery, and data from different sources is not integrated
3. **Third Normal Form layer**

From staging, the data will transition into the foundation or integration layer via another set of ETL processes. Data begins to take shape and it is not uncommon to have some end-user application access data from this layer especially if they are time sensitive, as data will become available here before it is transformed into the dimension / performance layer. Traditionally this layer is implemented in the Third Normal Form (3NF).

1. **Dimensional layer.** The Dimensional layer consists of dimension and fact tables.It is used as a source for creating data marts.
2. **Data Marts.** You may want to customize your warehouse's architecture for different groups within your organization. You can do this by adding data marts, which are systems designed for a particular line of business.

# Data Vault

* Get back to the task 5 from and refactor sample business schema into the Data Vault model.