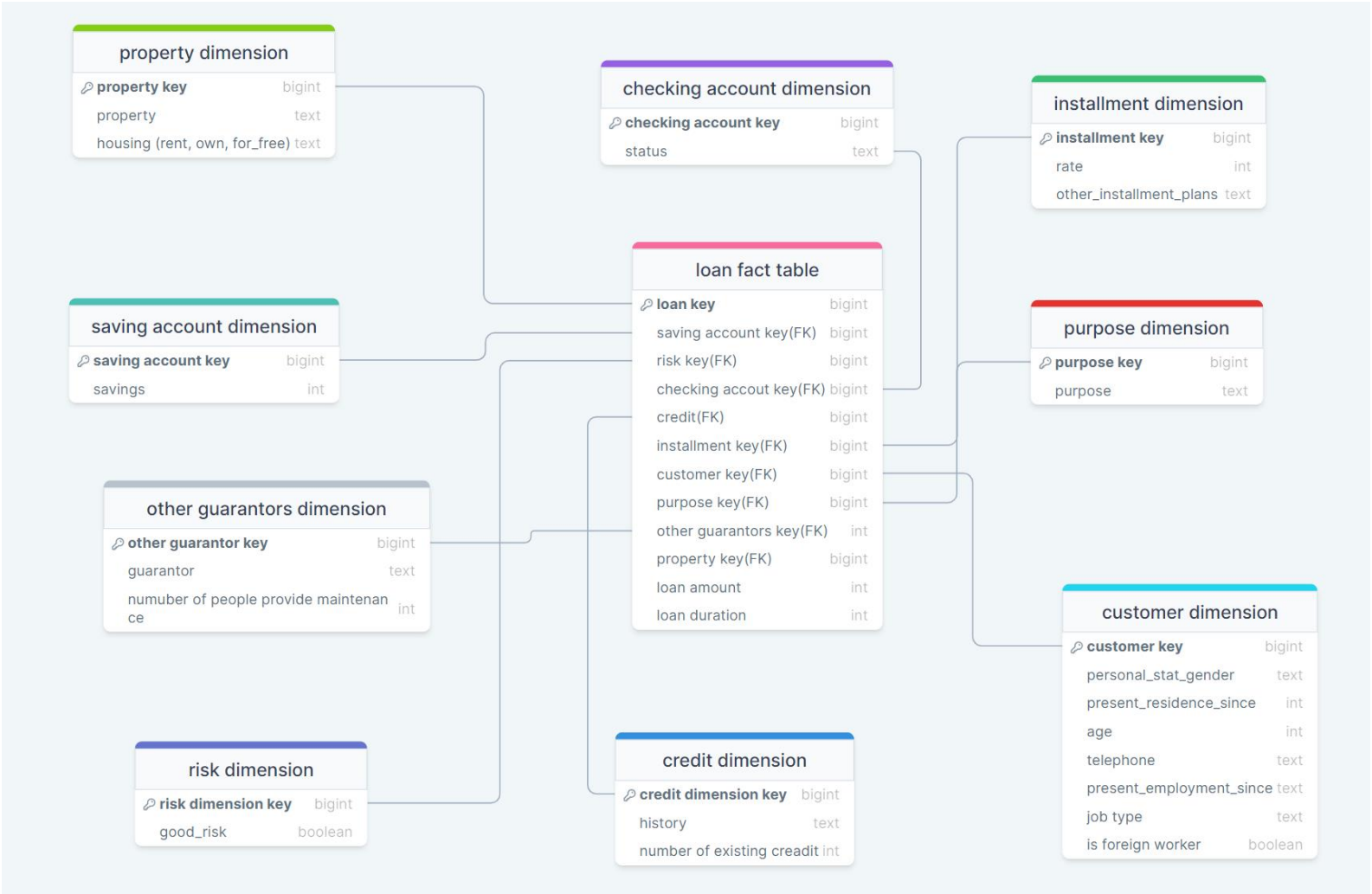
Grain:

the **loan amount** (in DM) and **duration** (in month) in a specific loan made according to the information of **property** (including housing), **savings**, **other guarantors** (including number of people provide maintenance), **loan risk**, **credit history** (including number of existing credits), **installmentall rate** (including other installment plans), **customer** (including, personal stated gender, present resience years, age, telephone, job type, if is foreign worker)



Assumption: the loan decisions were made only by the current information, and won't change over time

Loan fact table:

Loan amount (unit in DM): integer, minimum = 0, sample value = 10000

Loan duration (unit in month): integer, minimum =0, sample value = 24

Property dimension:

Property: enumeration text value (real_estate, life_insurance_or_agreements, car_or_other, unknown_or_no_property), sample value = real_estate

Housing: enumeration text value (rent, own, for_free), sample value = rent

Saving account dimension:

Savings (unit in DM): integer, minimum = 0, sample value = 2000

Risk dimension:

Good_risk: Boolean, sample = true

Checking Account Dimension:

Status: enumeration text value (no_checking_account, below_200, above_200, below_0), sample value = below_0

Installment Dimension:

Rate (in %): Integer, minimum=0, sample=3

Other_installment_plans: enumeration text value (none, bank, stores), sample value = bank

Purpose Dimension:

Purpose: enumeration text value (car_new, car_used, furniture_equipment, radio_tv, domestic_appliance, repairs, education, retraining, business, others), sample value= others

Customer Dimension:

Prsent_Employment: enumeration text value (unemployed, below_1y, below_4y, below_7y, above_7 y), Sample value= unemployed

Personal_stat_gender: enumeration text value (male_single, female_single, male_married_widowed, male_divorced_separated, female_divocred_separated-_married), sample value = male_single

Present_residence_since: integer, unit: year, sample value = 10

Age: integer, minimum = 0, sample value = 35

Telephone: text in telephone format, sample value= 343-456-7890

Job_type: enumeration text value (unemployed_non_resident, unskilled_resident, skilled_official, management_or_self_emp), sample value = unskilled_resident

Is_foreign_worker: Boolean, sample = true

Other gurantors dimension:

Guarantor: enumeration text value (co-applicant, guarantor, none), sample value = guarantor

numuber of people provide maintenance: integer, sample value = 2

Credit dimension:

History: enumeration text value (no_loan_or_paid_duly_other, paid_duly_this_bank, curr_loans_paid_duly, delay_in_past, risky_acc_or_curr_loan_other), sample value = no_loan_or_paid_duly_other

Number of existing credit : integer, sample value = 2

Note: DM = Deutsche Mark

10 design mistakes we avoided/handled:

1. Place text attributes in the fact table

This was avoided by only having numerical metrics in the fact table.

2. Limit verbose descriptions to save space

The field names and descriptions are sufficiently verbose and contain all relevant information.

3. Normalise to save space

We have avoided this pitfall because there is very little duplicated data, for example, we put the work type and status in customer dimension instead of creating a work dimension and attching it to customer dimension.

4. Ignore the need to track changes

Not applicable, as our dataset does not involve tracking data over time, because the loan is made only based on the current information of the applicant.

5. Add new hardware to solve all query performance issues

Not applicable, modern consumer hardware meets our performance needs.

6. Use operational keys as primary keys

Not applicable, our dataset does not use any operational keys.

7. Neglect to declare (and comply with) the grain

The grain has been clearly defined as an application and all tables comply with it.

8. Neglect a detailed design

Our schema is comprehensive, containing all necessary information for database construction.

9. Expect users to query normalized data

Addressed by using views to provide denormalized data for end-user queries.

10. Fail to conform Facts and Dimensions

Not a concern as there is no dimension reuse, we have one fact table here only, eliminating the risk of non-conformity.

Work plan:
Chentao: designed, drew, and modified the model, wrote the attributes domain and sample value, reviewed the document.
Andy: designed the model, wrote the attributes domain and sample value.
Amine: designed the model, wrote the 10 design mistakes we avoided.

Reference:
https://www.kaggle.com/datasets/elsnkazm/german-credit-scoring-data
https://drawsql.app/teams/231/diagrams/1