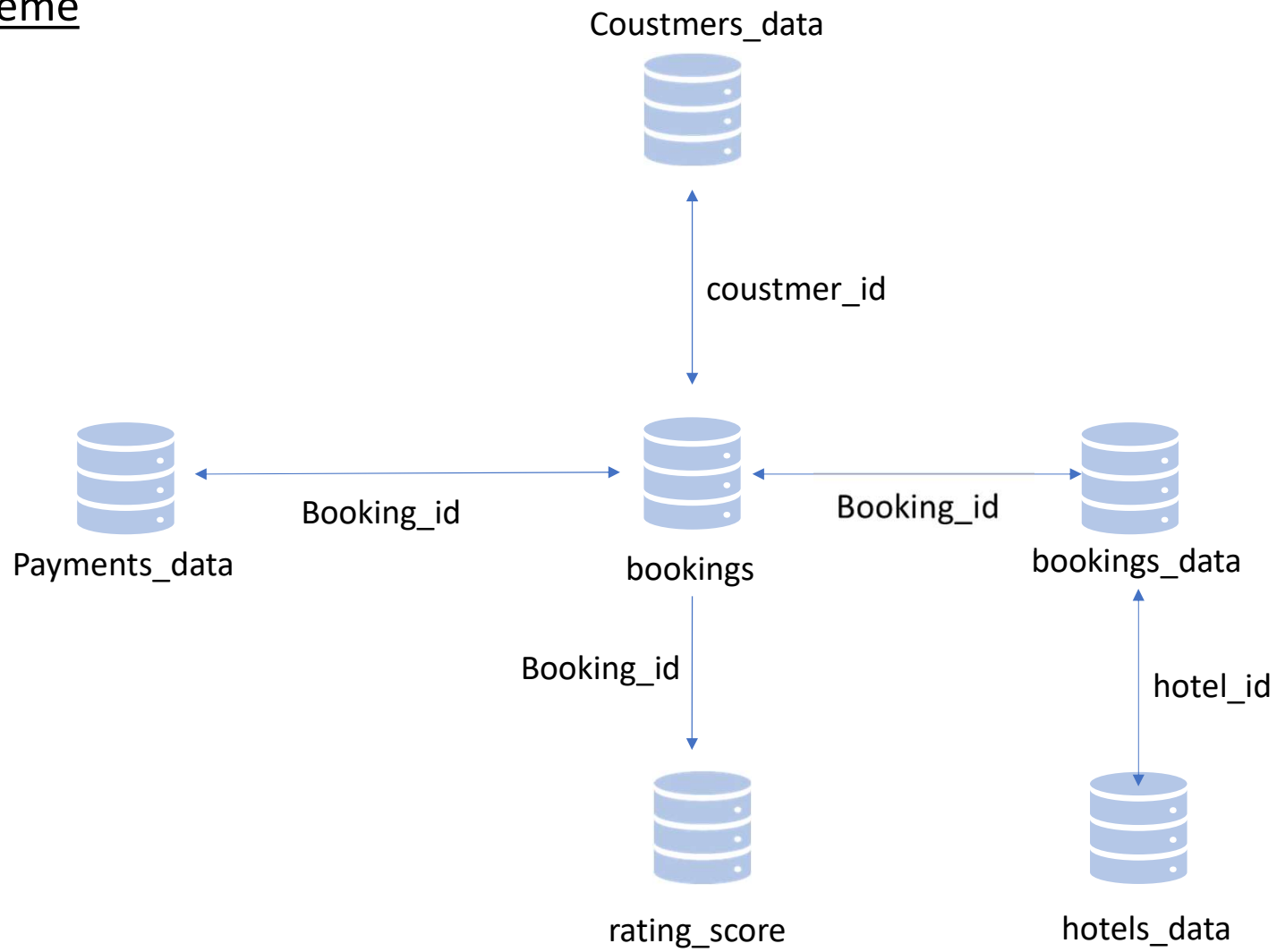


Data Contest - Report

Team Name: CE20B085_BE18B014

Data - Scheme



Features Available: country, booking_status, booking_create_timestamp, booking_approved_at, booking_checkin_customer_date, booking_sequence_id, booking_expiry_date, price, agent_fees, payment_sequential, payment_type, payment_installments, payment_value, hotel_category, hotel_name_length, hotel_description_length, hotel_photos_qty

Correcting Features:

Marking different “country” name with number.

Set *completed* “booking_status” +1 and else status to -1.

Marking different “payment_type” with different number.

Dropping Features : Dropping features which are less correlated with rating score like booking_create_timestamp, booking_approved_at, booking_checkin_customer_date, booking_expiry_date, payment_value.

Adding new features: adding “timediff” as new features which is difference in second between booking_approved_at and booking_create_timestamp

Features used for training: country, booking_status, timediff, booking_sequence_id, price, agent_fees, payment_sequential, payment_type, payment_installments, hotel_category, hotel_name_length, hotel_description_length, hotel_photos_qty

Model used for training and prediction of rating score is **Linear Regression** and then rounding off predicted value to nearest integer.