INFO 4555
FA23
Catherine Wang (cw797), Nancy Zhang (yz548), Shuqian Lyu (sl2335)
Team 6

# Milestone 3

## Requirements

The overarching requirement is to create a BI system that can dynamically perform analysis regarding the lab's usage, financial performance, instrument scheduling, and customer behavior around scheduling for BRC.

List of Requirements:

**Financial Analysis:**

- Actual amount (price * quantity)
- Billed amount (does not include subsidies)

**Usage Analysis:**

- Units of service performed
- Attributes the Director is interested in filtering and aggregating include:
    - Year/Month/Quarter, facility group, facility, service, PI, PI institution, PI department/org (if Cornell), PI college (if Cornell)

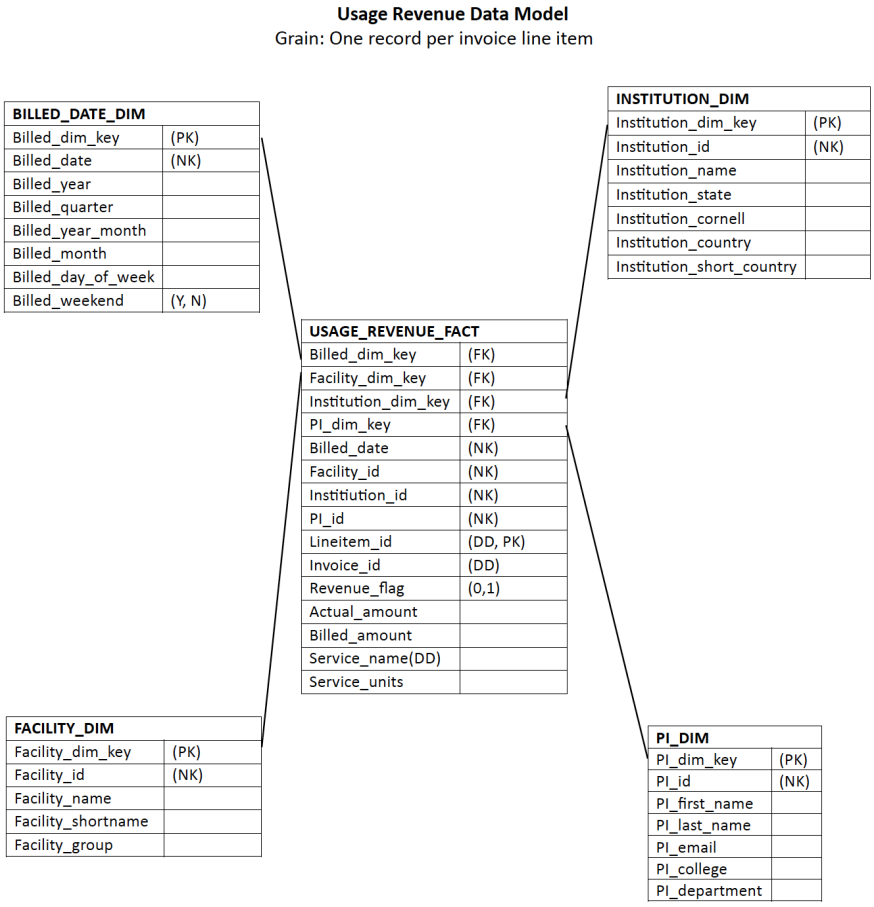
**Customer Scheduling Behavior Analysis (only go back 1 year):**

- Are some customers gaming the system?
    - Customers who double-book and/or double-cancel reservations
    - Customers who scheduled several months ahead of the event and/or reserved large blocks of time
    - Customers who (habitually) cancel reservations 24 hours before the event
- How far in advance are reservations made?
    - Advance Booking Period of Reservations (in days)
        - Reservation date (year, month, date) – reservation booking date (year, month, date)
- How many changes are made to reservations?

    - The number of changes made to reservations can be calculated by counting how many “EDIT” (and “DELETE” if the reservation is canceled) actions are made for a single reservation.
- How far in advance are reservation schedules changed?
    - Advance Notice of Changes in Reservation Schedules (in days)
        - Reservation Date (year, month, date) – Reservation Change Date (year, month, date)
- How far in advance are reservations canceled?
    - Advance Notice of Cancellation in Reservations (in days)
        - Reservation Date (year, month, date) - Reservation Cancellation Date (year, month, date)
- Number of reservations by, customer, college, etc. Along with percent cancelations, and average lead time to cancelation.
    - Number of Reservations by Each Customer
    - Number of Reservations by Each College
    - Percent of Cancellations (%)
        - Total Number of Reservations / Total Number of Reservation Cancellations
    - Average Lead Time to Cancellation (in days)
        - Total of Advance Notice of Cancellation in Reservations / Number of Reservation Cancellations

# Data Model

## Usage Revenue Data Model




# Data Dictionary

## "Usage Revenue Data Model" Dictionary

| Table Name | Column Name | Data Type | Definition |
|---|---|---|---|
| Usage_Revenue_Fact | billed_date | date | The date on which the invoice is billed to its customer |
| Usage_Revenue_Fact | pi_id | number | Uniquely identify a principal investigator |
| Usage_Revenue_Fact | facility_id | number | Uniquely identify a facility |
| Usage_Revenue_Fact | lineitem_id | number | A degenerate dimension for each line item on an invoice |

| | | | |
|---|---|---|---|
| Usage_Revenue_Fact | invoice_id | number | Uniquely identify an invoice |
| Usage_Revenue_Fact | actual_amount | number | The actual amount of revenue earned |
| Usage_Revenue_Fact | billed_amount | number | The amount billed to the customer, with subsidy excluded |
| Usage_Revenue_Fact | service_name | text | The name of the service performed |
| Usage_Revenue_Fact | service_units | number | The number of units of service performed |
| Billed_Date_Dim | billed_dim_key | number | Uniquely identify a bill date |
| Billed_Date_Dim | billed_date | date | The bill date in date format |
| Billed_Date_Dim | billed_year | number | The year of the bill date |
| Billed_Date_Dim | billed_quarter | number | The quarter of the bill date |
| Billed_Date_Dim | billed_yearmonth | number | The year & month of the bill date |
| Billed_Date_Dim | billed_month | text | The month of the bill date |
| Billed_Date_Dim | billed_day_of_week | text | The day of week of the bill date identified with a 3-letter abbreviation (E.g., MON) |
| Billed_Date_Dim | billed_day_of_week_no | number | The day of week of the bill date identified with an integer from 1 to 7. |
| Billed_Date_Dim | billed_weekend | boolean | Y if the bill date is a weekend, N otherwise |
| Facility_Dim | facility_id | number | Uniquely identify a facility; correspond to facility_id |
| Facility_Dim | facility_name | text | A descriptive name for the facility |
| Facility_Dim | facility_shortname | text | A shorter version of the descriptive name for the facility |
| PI_Dim | id | number | Uniquely identify a principal investigator; correspond to pi_id |
| PI_Dim | pi_first_name | text | The first name of the principal investigator |
| PI_Dim | pi_last_name | text | The last name of the principal investigator |
| PI_Dim | pi_email | text | The email address of the principal investigator |
| PI_Dim | pi_college | text | The college to which the principal investigator belongs |
| PI_Dim | pi_department | text | The department to which the principal investigator belongs |
| Institution_Dim | Institution_id | number | Uniquely identify an institution |

| | | | |
|---|---|---|---|
| Institution_Dim | Institution_name | text | The name of the institution |
| Institution_Dim | Institution_state | text | The two-letter abbreviation of the state where the institution is located |
| Institution_Dim | Institution_cornell | boolean | Y if Cornell, N if another institution |
| Institution_Dim | Institution_country | text | Name of the country |
| Institution_Dim | Institution_short_country | text | The two-letter abbreviation of the country |

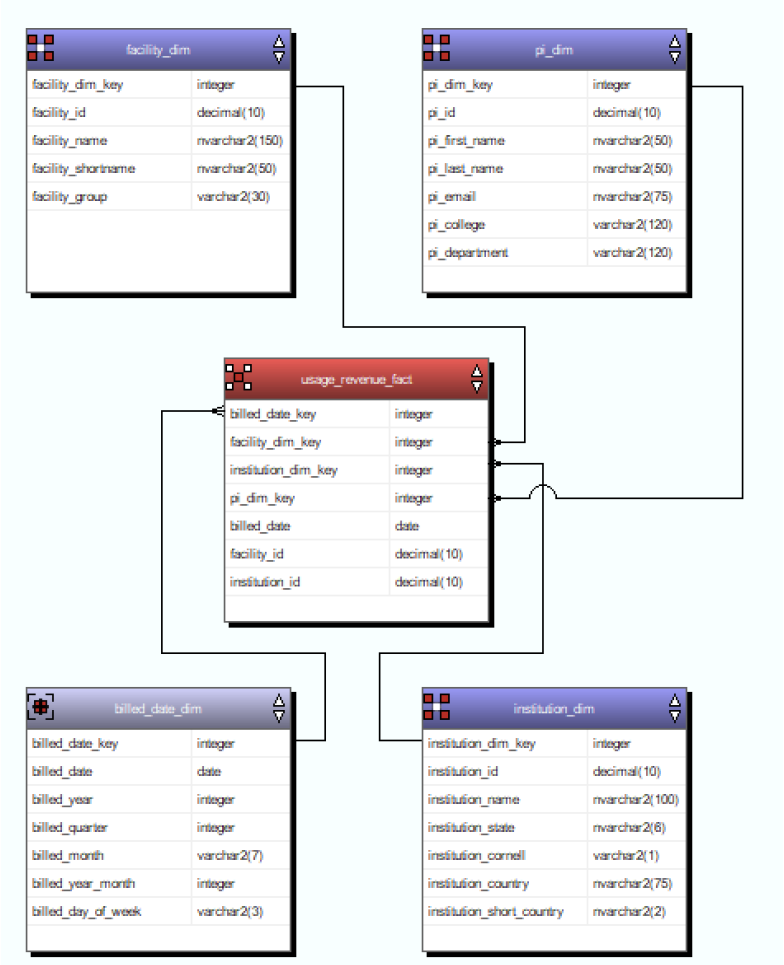
## ETL Documentation

### "Usage Revenue Data Model" Mapping

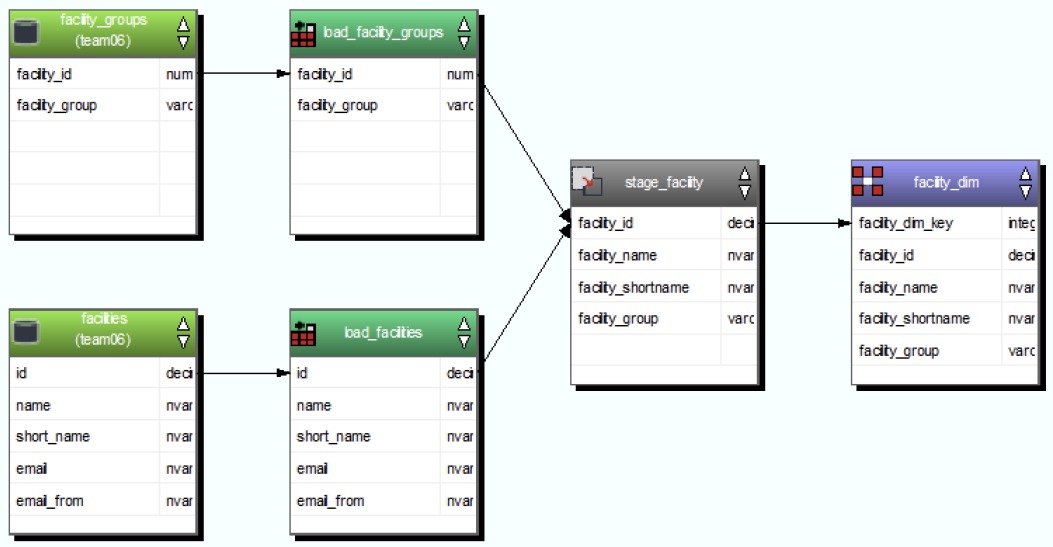
| Source Table | Source Column | Target Table | Target Column | Transformation Logic | Notes |
|---|---|---|---|---|---|
| Invoice | billdate | Usage_Revenue_Fact | billed_date | | NK |
| Invoice | pi_id | Usage_Revenue_Fact | pi_id | | NK |
| Lineitem | facility_id | Usage_Revenue_Fact | facility_id | | NK |
| Lineitem | Institution_id | Usage_Revenue_Fact | institution_id | | NK |
| Lineitem | id | Usage_Revenue_Fact | lineitem_id | | DD, PK |
| Lineitem | invoice_id | Usage_Revenue_Fact | invoice_id | | DD |
| Lineitem | line_total_raw | Usage_Revenue_Fact | actual_amount | | |
| Lineitem | line_total | Usage_Revenue_Fact | billed_amount | | |
| Lineitem | name | Usage_Revenue_Fact | service_name | | DD |
| Lineitem | quantity | Usage_Revenue_Fact | service_units | | |
| Dim_Date | dim_date_key | Billed_Date_Dim | billed_dim_key | | |
| Dim_Date | calendar_date | Billed_Date_Dim | billed_date | | |
| Dim_Date | cal_year | Billed_Date_Dim | billed_year | | |
| Dim_Date | cal_quarter | Billed_Date_Dim | billed_quarter | | |
| Dim_Date | cal_month | Billed_Date_Dim | billed_yearmonth | | |
| Dim_Date | cal_month_name | Billed_Date_Dim | billed_month | | |

| | | | | | |
|---|---|---|---|---|---|
| Dim_Date | cal_day_in_week | Billed_Date_Dim | billed_day_of_week | | |
| Dim_Date | cal_day_week_no | Billed_Date_Dim | billed_day_of_week_no | | |
| Dim_Date | week_end_flag | Billed_Date_Dim | billed_weekend | | |
| Facilities | id | Facility_Dim | facility_id | | NK |
| Facilities | name | Facility_Dim | facility_name | | |
| Facilities | shortname | Facility_Dim | facility_shortname | | |
| Not from source data | Not from source data | Facility_Dim | facility_group | Based on the facility id groupings provided in the project doc | |
| PIs | id | PI_Dim | pi_id | | NK |
| PIs | first_name | PI_Dim | pi_first_name | | |
| PIs | last_name | PI_Dim | pi_last_name | | |
| PIs | email | PI_Dim | pi_email | | |
| CU_Person | primary_college org_name | PI_Dim | pi_college | | |
| CU_Person | primary_org_name | PI_Dim | pi_department | | |
| Institutions | id | Institution_Dim | Institution_id | | NK |
| Institutions | name | Institution_Dim | Institution_name | | |
| Institutions | state | Institution_Dim | Institution_state | | |
| Institutions | cornell | Institution_Dim | Institution_cornell | If 1 then “Y”, else “N” | |
| Countries | name | Institution_Dim | Institution_country | | |
| Countries | iso | Institution_Dim | Institution_short_country | | |

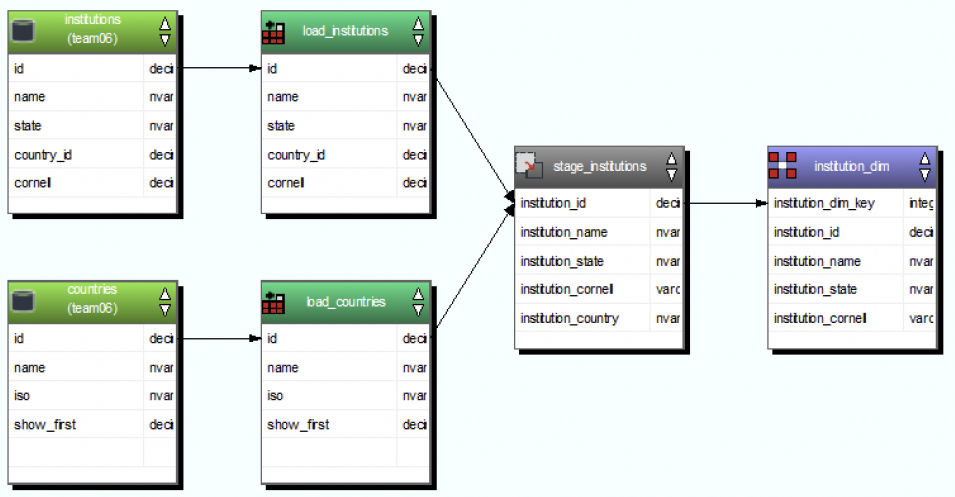
## Usage_Revenue_Fact Diagram



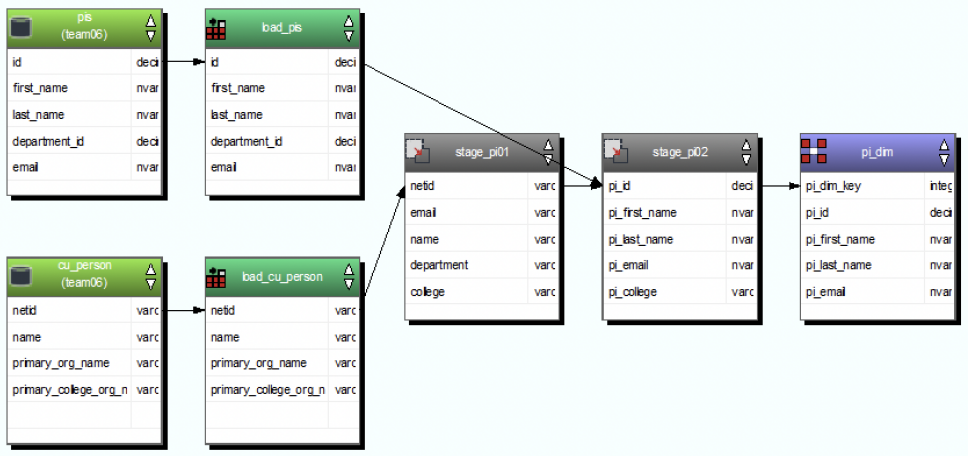

## Facility_Dim Diagram



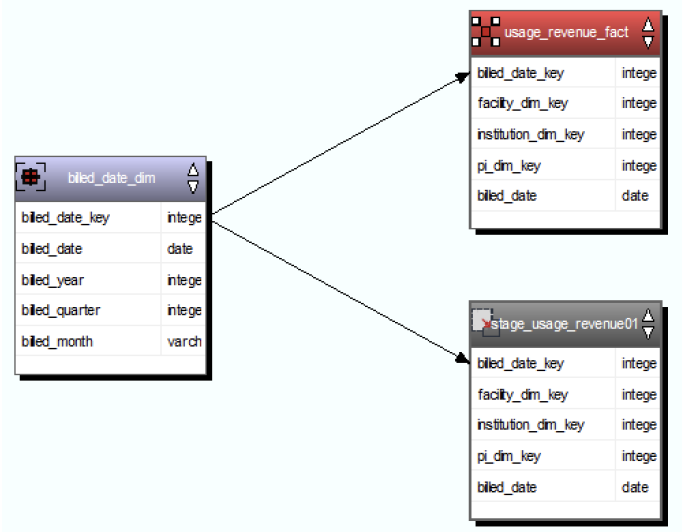
## Institution_Dim Diagram




## PI_Dim Diagram




## Billed_Date_Dim Diagram



## Assumptions & Questions

### Assumptions

- The primary_college_org_name column in CU_Person corresponds to the college of the PI while the primary_org_name column corresponds to the department of the PI.
- Some columns in the delivered dim_date table are irrelevant to our project requirements and are dropped.
- A time dimension is unnecessary for the financial and usage analysis, and it needs to be created only for scheduling analysis purposes.

### Questions

- In the source table dim_date, what is the purpose of including both billed_day_of_week and billed_day_of_week_no (E.g., different formats potentially for convenience)?
- What exactly are the Nulls/NAs/Unknowns in PI_college?
- What is the most informative date range for analysis purposes? Should the range differ to answer different project questions?