

ABC Adventure – Data Warehouse Design

Business Process

The Data Warehouse is designed for the children enrollment business process. This process is described in the document *Specification of business processes* in ABC Adventure network.

Relational Database schema

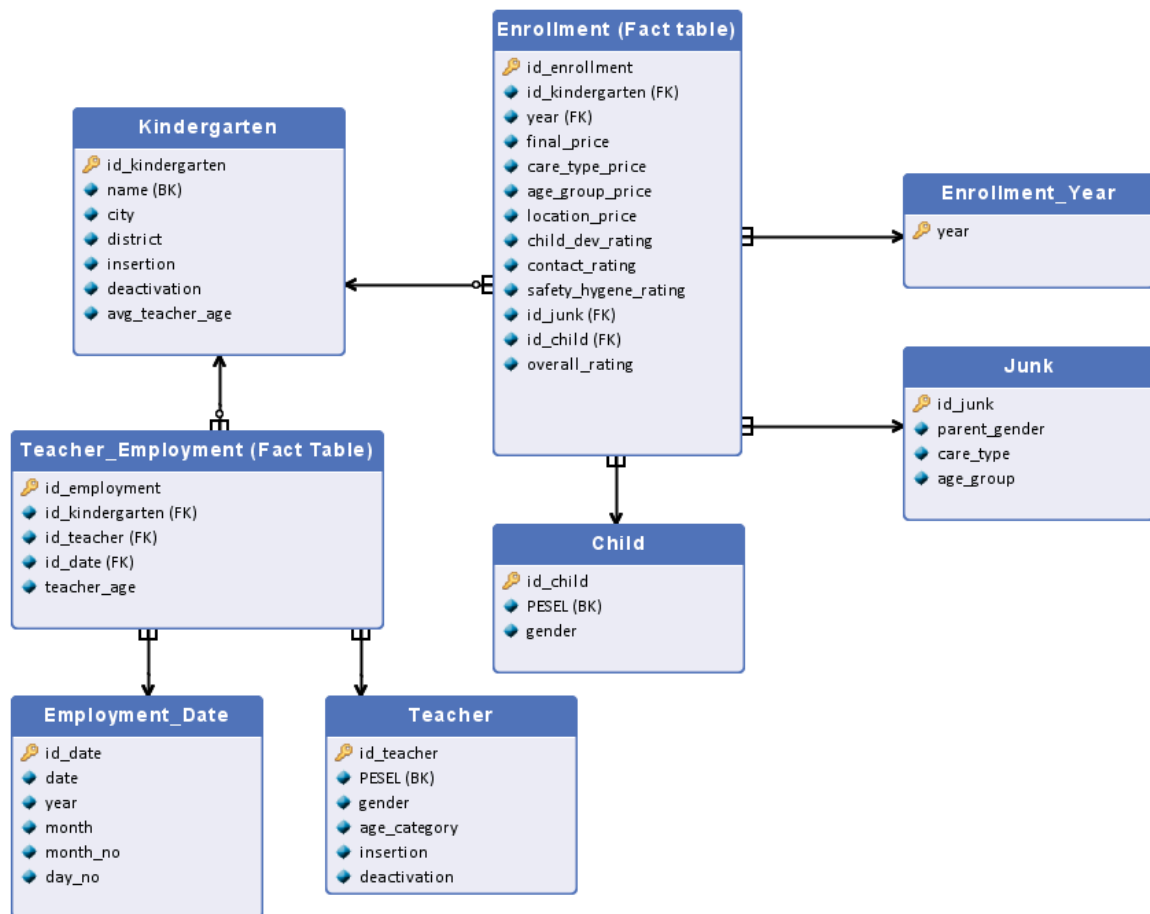


TABLE NAME	ATTRIBUTE	ATTRIBUTE TYPE	DESCRIPTION
Enrollment (Fact Table)	One tuple describes one fact of enrolling a child to a specific kindergarten.		
	id_enrollment	Numeric	Surrogate key
	id_kindergarten	Numeric	Foreign key

			kindergarten
	year	Numeric	Foreign key year of enrollment
	final_price	Money	(Total) price paid by the customer for enrolling their child
	care_type_price	Money	Partial price of final_price, based on the care type of the enrollment
	age_group_price	Money	Partial price of final_price, based on the age group of the enrolled child
	location_price	Money	Partial price of final_price, based on the location of the kindergarten to which the child is enrolled
	child_dev_rating	Numeric	Integer number representing child development rating from 1 (worst) to 10 (best)
	contact_rating	Numeric	Integer number representing contact with facility rating from 1 (worst) to 10 (best)
	safety_hygiene_rating	Numeric	Integer number representing safety & hygiene rating from 1 (worst) to 10 (best)
	id_junk	Numeric	Foreign key junk
	id_child	Numeric	Foreign key child concerned within the enrollment
	overall_rating	Numeric	float, arithmetic mean of the three types of ratings
Teacher_Employment (Fact Table)	One tuple describes one employment of a teacher by a kindergarten.		
	id_employment	Numeric	Surrogate key
	id_kindergarten	Numeric	Foreign key kindergarten

	id_teacher	Numeric	Foreign key teacher employed
	id_date	Numeric	Foreign key date of employment (date of creation of the tuple)
	teacher_age	Numeric	Age of the teacher (w.r.t. the date of creation of the tuple)
Kindergarten (Dimension Table)	One tuple describes one kindergarten institution (under a specific location)		
	id_kindergarten	Numeric	Surrogate key
	name	Varchar(60)	BK Name of the kindergarten
	city	Varchar(16)	City, where the kindergarten is located
	district	Varchar(50)	District, where the kindergarten is located
	avg_teacher_age	Varchar(13)	Average age of teachers of this kindergarten (as a category). Allowed values: under 25, from 26 to 32, from 33 to 40, from 40 to 49, from 50 to 58, more than 58
	insertion	Date	Date of insertion of the row (SCD 2), for the sake of change in average teachers age (if needed)
	deactivation	Date	Date of deactivation of the row (SCD 2), for the sake of change in average teachers age (if needed)
Teacher (Dimension Table)	One tuple describes one teacher enrolled by the kindergarten, to which the child was enrolled		
	id_teacher	Numeric	Surrogate key
	PESEL (BK)	11 digits	BK Personal identification number

	gender	Varchar(6)	Teacher's gender. Allowed values: male, female
	age_category	Varchar(13)	Teacher's age (as a category). Allowed values: under 25, from 26 to 32, from 33 to 40, from 40 to 49, from 50 to 58, more than 58
	insertion	Date	Date of insertion of the row (SCD 2), for the sake of change in teacher age category (if needed)
	deactivation	Date	Date of deactivation of the row (SCD 2), for the sake of change in teacher age category (if needed)
Employment_Date (Dimension Table)	One tuple describes one day.		
	id_date	Numeric	Surrogate key
	date	Date	Date
	year	4 digits	Year
	month	Varchar(10)	Month. Allowed values: January, February, March, April, May, June, July, August, September, October, November, December.
	month_no	Numeric	Month's numeric value. Allowed values: 1-12.
	day_no	Numeric	Day's numeric value. Allowed values: 1-31.
Child (Dimension Table)	One tuple describes one child concerned by the enrollment.		
	id_child	Numeric	Surrogate key
	PESEL (BK)	11 digits	BK Personal identification number

	gender	Varchar(6)	Child's gender. Allowed values: male, female
Junk (Dimension Table)	One tuple describes one of each following dimension attributes.		
	id_junk	Numeric	Surrogate key
	parent_gender	Varchar(6)	teacher's gender. Allowed values: male, female
	care_type	Varchar(10)	BK Name of the care type. Allowed values: full-day, half-day, extended, drop-in.
	age_group	Varchar(16)	BK Name of the age group. Allowed values: toddler, junior, preschooler.
Enrollment_Year (Dimension Table)	One tuple describes one year		
	year	4 digits	PK, BK Year

Dimensional model

Fact definitions

Fact 1 Enrollment fact: Enrollment of a child in kindergarten by a parent for the next school year. The details of the record include: the type of care (and its price), the age group (and its price) and the location of the kindergarten, specifying the city and district (and the price depending on the city and district). The fee is paid at the beginning of the school year, it is the price for the entire year.

Fact table: Enrollment (fact table)

Granularity:

- a specified enrollment
- a specified kindergarten
- a specified care type
- a specified age group
- a specified child
- a specified location (city, district)

Measures and aggregation functions:

- **total child development rating:**
`sum(child_dev_rating)`
- **average child development rating:**
`sum(child_dev_rating) / count(id_enrollment)`
- **total contact rating:**
`sum(contact_rating)`
- **average contact rating:**
`sum(contact_rating) / count(id_enrollment)`
- **total safety and hygiene rating:**
`sum(safety_hygiene_rating)`
- **average safety and hygiene rating:**
`sum(safety_hygiene_rating) / count(id_enrollment)`
- **total number of enrollments:**
`count(id_enrollment)`
- **total revenue:**
`sum(final_price)`
- **average final price:**
`sum(final_price) / count(id_enrollment)`
- **total care type revenue:**
`sum(care_type_price)`
- **average care type price:**
`sum(care_type_price) / count(enrollment_id)`
- **total age group revenue:**
`sum(age_group_price)`
- **average age group price:**
`sum(age_group_price) / count(enrollment_id)`
- **total location price:**
`sum(location_price)`
- **average location price:**
`sum(location_price) / count(enrollment_id)`
- **total overall rating:**
`sum(overall_rating)`
- **average overall rating:**

$\text{sum(overall_rating)} / \text{count(enrollment_id)}$

Fact 2 Teacher_Employment fact: Employment of a teacher by a kindergarten facility. Each kindergarten employs 3 teachers. The details of the record include: date of employment (based on tuple creation date), age of the teacher w.r.t. employment date, teacher's PESEL, gender and age category, and kindergarten the teacher is employed by.

Fact table: Teacher_Employment (fact table)

Granularity:

- a specified employment
- a specified kindergarten (with a specified location (city, district))
- a specified teacher
- a specified date
- a specified child

Measures and aggregation functions:

- **total teacher age:**
 sum(teacher_age)
- **total number of employed teachers:**
 $\text{count(id_employment)}$
- **average teacher age:**
 $\text{sum(teacher_age)} / \text{count(id_employment)}$

Dimension definitions

Dimensions for Fact 1 Enrollment fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
Junk	Junk	Dimension
Parent's gender	Junk.parent_gender	Dimension attribute
Care type	Junk.care_type	Dimension attribute
Age group	Junk.age_group	Dimension attribute
Enrollment Year	Enrollment_Year	Dimension
Year	Enrollment_Year.year	Dimension attribute
Kindergarten	Kindergarten	Dimension

Name	Kindergarten.name	Dimension attribute
City	Kindergarten.city	Dimension attribute
District	Kindergarten.district	Dimension attribute
Kindergarten location hierarchy	<ul style="list-style-type: none"> • Kindergarten.city •• Kindergarten.district 	Hierarchical dimension
Average teacher age	Kindergarten.avg_teacher_age	Dimension attribute
Child	Child	Dimension
Child's PESEL	Child.PESEL	Dimension attribute
Child's gender	Child.gender	Dimension attribute

Dimensions for Fact 2 Teacher_Employment fact:

DIMENSION/DIMENSION ATTRIBUTE	TABLE/COLUMN	TYPE
Teacher	Teacher	Dimension
Teacher's PESEL	Teacher.PESEL	Dimension attribute
Teacher's gender	Teacher.gender	Dimension attribute
Teacher's age category	Teacher.age_category	Dimension attribute
Kindergarten	Kindergarten	Dimension
Name	Kindergarten.name	Dimension attribute
City	Kindergarten.city	Dimension attribute
District	Kindergarten.district	Dimension attribute
Kindergarten location hierarchy	<ul style="list-style-type: none"> • Kindergarten.city •• Kindergarten.district 	Hierarchical dimension
Average teacher age	Kindergarten.avg_teacher_age	Dimension attribute
Employment date	Employment_Date	Dimension
Year	Employment_Date.year	Dimension attribute
Month	Employment_Date.month	Dimension attribute
Month number	Employment_Date.month_no	Dimension attribute

Day number	Employment_Date.day_no	Dimension attribute
Employment date hierarchy	<ul style="list-style-type: none"> • Employment_Date.year •• Employment_Date.month ••• Employment_Date.day_no 	Hierarchical dimension

Checking the feasibility of queries based on the multidimensional model

1. Which district has the highest average price of the offers?
Measure: average final price
Dimension: Kindergarten (dimension attributes: Kindergarten.district)
2. Does the contact rating affect the average price of the offers?
Measure: average final price
Measure: average contact rating
3. What are the average prices for each care type?
Measure: average care type price
Dimension: Junk (dimension attributes: Junk.care_type)
4. What are the most popular care types in selected districts?
Measure: total number of enrollments
Dimension: Junk (dimension attributes: Junk.care_type)
Dimension: Kindergarten (dimension attributes: Kindergarten.district)
5. Does the average age of the teachers affect the average price of the offer?
Measure: average final price
Measure: average teacher age
Dimension: Kindergarten (dimension attributes: Kindergarten.avg_teacher_age)
Dimension: Kindergarten (dimension attributes: Kindergarten.name)
6. How many children are enrolled each year?
Measure: total number of enrollments
Dimension: Enrollment_Year (dimension attributes: Enrollment_Year.year)
7. Which age group is least commonly enrolled in kindergartens where contact with employees is below average and what is the average age group price in those facilities?
Measure: total number of enrollments
Measure: average contact rating
Measure: average age group price
Dimension: Junk (dimension attributes: Junk.care_type)
Dimension: Junk (dimension attributes: Junk.age_group)

Dimension: Kindergarten (dimension attributes: Kindergarten.name)

8. Are there more children enrolled in kindergartens with higher (than the average) child development ratings each year?

Measure: total number of enrollments

Measure: average child development rating

Dimension: Kindergarten (dimension attributes: Kindergarten.name)

Dimension: Enrollment_Year (dimension attribute: Enrollment_Year.year)

9. What is the proportion of children enrolled in each age group and what is the average price of the specific age group?

Measure: total number of enrollments

Measure: average age group price

Dimension: Junk (dimension attributes: Junk.age_group)

10. Compare the number of enrollments in 2022 of the best 10 and the worst 10 kindergartens, in terms of overall grade.

Measure: total number of enrollments

Measure: average overall grade

Dimension: Enrollment_Year (dimension attributes: Enrollment_Year.year)

Checking if there is data in the data sources needed to fill the data warehouse

TABLE NAME	COLUMN	SOURCE
Enrollment (Fact Table)	One tuple describes one fact of enrolling a child to a specific kindergarten.	
	id_enrollment	Enrollment ID. Surrogate key - generated by data warehouse.
	id_kindergarten	Kindergarten ID. Foreign key from dimension table. Based on dimension's surrogate key.
	year	Year. Foreign key from dimension table. Based on dimension's surrogate key.
	final_price	Price paid by the customer - sum of care_type_price, age_group_price and location_price from Enrollment.
	care_type_price	Partial price based on the care type. Taken from CEO Excel, columns I, J, K, L - each column stores price for a

		specific care type.
	age_group_price	Partial price based on the age group. Taken from CEO Excel, columns F,G,H - each column stores price for a specific age group.
	location_price	Partial price based on the location of the kindergarten. Taken from CEO Excel, column E.
	child_dev_rating	Rating of child development, taken from ABC Adventure Database - table Ratings, column child_dev_rating.
	contact_rating	Rating of contact with facility, taken from ABC Adventure Database - table Ratings, column contact_rating.
	safety_hygiene_rating	Rating of safety and hygiene, taken from ABC Adventure Database - table Ratings, column safety_hygiene_rating.
	id_junk	Junk ID. Surrogate key - generated by data warehouse.
	id_child	Child ID. Foreign key from dimension table. Based on dimension's surrogate key.
	overall_rating	Overall rating - an arithmetic mean of child_dev_rating, contact_rating and safety_hygiene_rating.
Teacher_Employment (Fact Table)	One tuple describes one employment of a teacher by a kindergarten.	
	id_employment	Employment ID. Surrogate key - generated by data warehouse.
	id_kindergarten	Kindergarten ID. Foreign key from dimension table. Based on dimension's surrogate key.
	id_teacher	Teacher ID. Foreign key from dimension table. Based on dimension's surrogate key.
	id_date	Date ID. Foreign key from dimension table. Based on dimension's surrogate key.

	teacher_age	Integer number representation of age of the teacher, calculated from subtraction of teacher's birth year (ABC Adventure database -> Teachers -> birth_year) from the year taken from the date included in Date dimension table.
Kindergarten (Dimension Table)	One tuple describes one kindergarten institution (under a specific location)	
	id_kindergarten	Kindergarten ID. Surrogate key - generated by data warehouse.
	name	Name of the kindergarten. Taken from the ABC Adventure database - Kindergartens table, column name
	city	City, where the kindergarten is located. Taken from the ABC Adventure database - Cities table, column name
	district	District, where the kindergarten is located. Taken from the ABC Adventure database - Districts table, column name
	avg_teacher_age	Average age of the teachers in the kindergarten, calculated as a mean of teacher_age measure values (from Teacher_Employment fact table) and then assigned to one of the following categories: under 25, from 26 to 32, from 33 to 40, from 40 to 49, from 50 to 58, more than 58.
	insertion	Date of insertion - generated by data warehouse.
	deactivation	Date of deactivation - generated by data warehouse.
Teacher (Dimension Table)	One tuple describes one teacher enrolled by the kindergarten, to which the child was enrolled	
	id_teacher	Teacher ID. Surrogate key - generated by data warehouse.
	PESEL	BK, taken from the ABC Adventure database - Teachers

		table, column PESEL
	gender	Teacher's gender taken from the ABC Adventure database - Teachers table, column gender. Allowed values: male, female
	age_category	Teacher's age (as a category). Value from Teacher_Employment fact table transformed into one of the following categories: under 25, from 26 to 32, from 33 to 40, from 40 to 49, from 50 to 58, more than 58
	insertion	Date of insertion - generated by data warehouse.
	deactivation	Date of deactivation - generated by data warehouse.
Employment_Date (Dimension Table)	One tuple describes one day.	
	id_date	Date ID. Surrogate key - generated by data warehouse.
	date	Date of the generation of Teacher_Employment tuple, generated by data warehouse.
	year	Year taken from Date.date (described above)
	month	Month name determined by the month number taken from the date.
	month_no	Month number taken from Date.date (described above)
	day_no	Day number taken from Date.date (described above)
Child (Dimension Table)	One tuple describes one child concerned by the enrollment.	
	id_child	Child ID. Surrogate key - generated by data warehouse.
	PESEL	BK, taken from the ABC Adventure database - Children table, column PESEL
	gender	Child's gender taken from the ABC Adventure database -

		Children table, column gender. Allowed values: male, female
Junk (Dimension Table)	One tuple describes one of each following dimension attributes.	
	id_junk	Junk ID. Surrogate key - generated by data warehouse.
	parent_gender	Parent's gender taken from the ABC Adventure database - Parents table, column gender.
	care_type	BK Name of the care type, taken from ABC Adventure database - Care_types table, column name. Allowed values: full-day, half-day, extended, drop-in.
	age_group	BK Name of the age group, taken from ABC Adventure database - Enrollments table, column age_group. Allowed values: toddler, junior, preschooler.
Enrollment_Year (Dimension Table)	One tuple describes one year	
	year	Year in which the child was enrolled, taken from ABC Adventure database - Enrollments table, column year.