

# Система поддержки выдачи и приема заданий

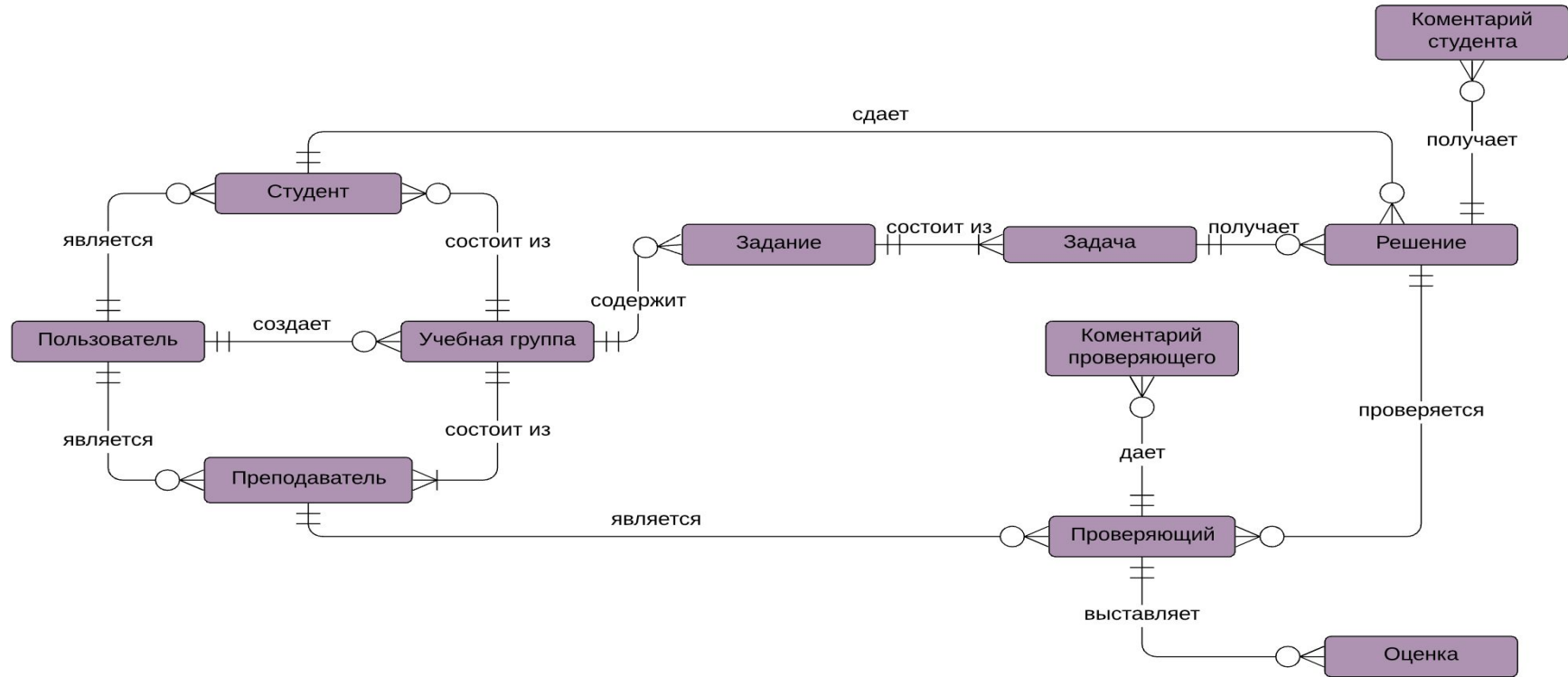
Команда Альфа

курс “Практикум на ЭВМ”  
2019

# Состав команды и распределение ролей

- Холодов Александр Валерьевич
  - Концептуальная модель
  - BCNF модель
  - DataVault
  - Презентация
- Крыжановская Светлана Юрьевна
  - Концептуальная модель
  - BCNF модель
  - Запросы загрузки
  - Презентация
- Коптелов Дмитрий Андреевич
  - Концептуальная модель
  - BCNF модель
  - Генерация данных
  - Запросы загрузки
- Устюжанин Александр Олегович
  - Концептуальная модель
  - Запросы от бизнеса
- Лукьянов Павел Александрович
  - Концептуальная модель
  - Модель DataVault
  - Запросы загрузки
  - Запросы от бизнеса

# Концептуальная модель



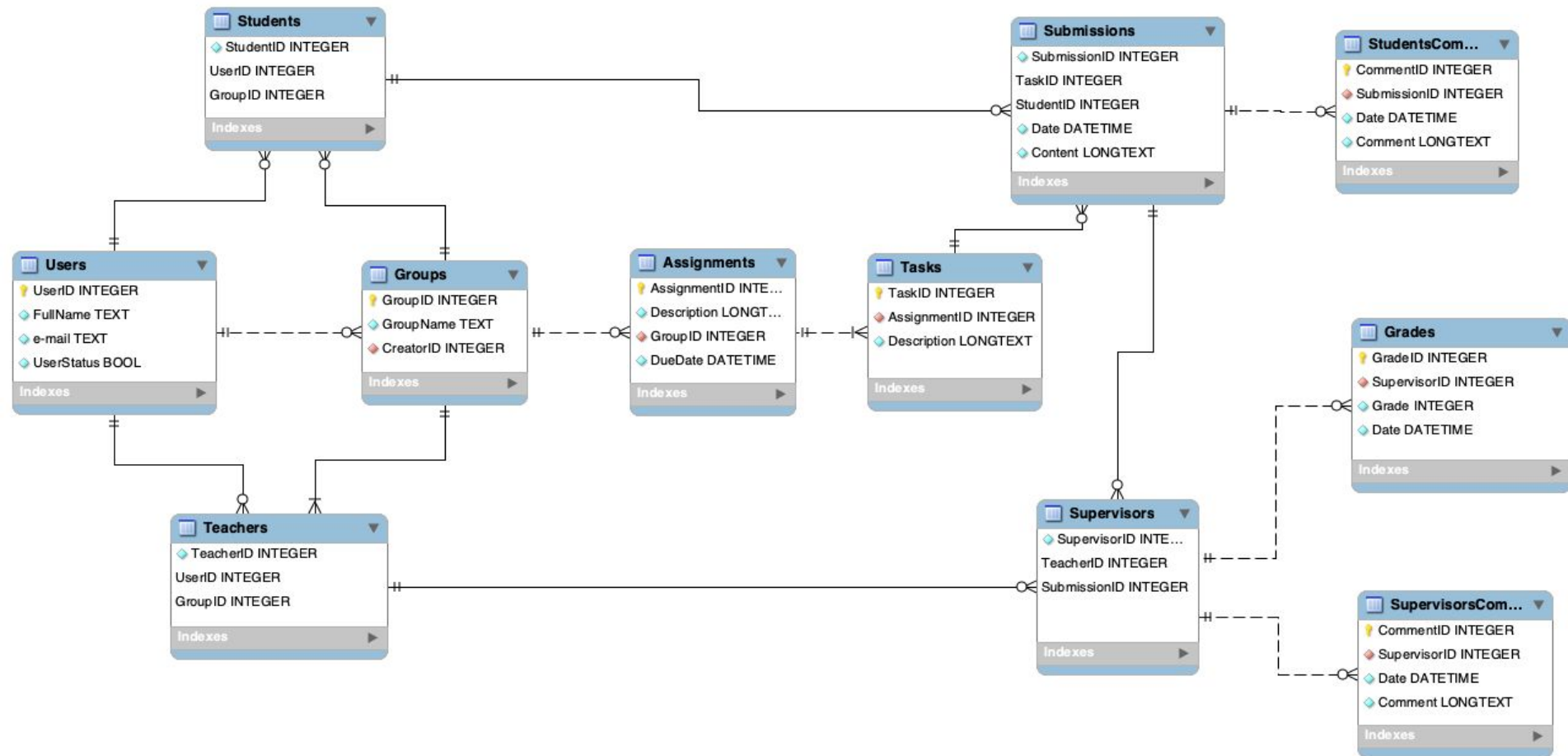
# Сгенерированные данные

- База пользователей из 100000 человек
- 10000 учебных групп
- В каждой группе 10 - 60 человек
- 10 - 20% участников группы - преподаватели в ней
- В учебных группах 3 - 7 заданий
- В задании 1 - 4 задач
- Каждую задачу решают 30 - 80 % студентов данной группы
- Каждое решение проверяют до 3 преподавателей данной группы
- Каждой преподаватель выставляет до 2 оценок
- Студенты и преподаватели оставляют до 2 комментариев по решению

# Сгенерированные данные

Пользователи	100000
Группы	10000
Студенты	306450
Учителя	43132
Задания	49854
Задачи	124718
Решения	1988148
Проверяющие	2859692
Оценки	2862543
Комментарии от студента	281466
Комментарии от проверяющих	402708

# BCNF: логическая модель



# BCNF: протокол запуска запросов загрузки

✓	10	01:26:30	INSERT INTO logic.Users SELECT * FROM data.Users	100000 row(s) affected Records: 100000 Duplicates: 0 Warnings: 0	0.579 sec
✓	11	01:26:30	INSERT INTO logic.Groups SELECT * FROM data.Groups	10000 row(s) affected Records: 10000 Duplicates: 0 Warnings: 0	0.062 sec
✓	12	01:26:30	INSERT INTO logic.Students SELECT * FROM data.Students	306450 row(s) affected Records: 306450 Duplicates: 0 Warnings: 0	1.641 sec
✓	13	01:26:32	INSERT INTO logic.Teachers SELECT * FROM data.Teachers	43132 row(s) affected Records: 43132 Duplicates: 0 Warnings: 0	0.234 sec
✓	14	01:26:32	INSERT INTO logic.Assignments SELECT * FROM data.Assignments	49854 row(s) affected Records: 49854 Duplicates: 0 Warnings: 0	0.297 sec
✓	15	01:26:33	INSERT INTO logic.Tasks SELECT * FROM data.Tasks	124718 row(s) affected Records: 124718 Duplicates: 0 Warnings: 0	0.703 sec
✓	16	01:26:33	INSERT INTO logic.Submissions SELECT * FROM data.Submissions	1988148 row(s) affected Records: 1988148 Duplicates: 0 Warnings: 0	12.938 sec
✓	17	01:26:46	INSERT INTO logic.Supervisors SELECT * FROM data.Supervisors	2859692 row(s) affected Records: 2859692 Duplicates: 0 Warnings: 0	14.672 sec
✓	18	01:27:01	INSERT INTO logic.Grades SELECT * FROM data.Grades	2862543 row(s) affected Records: 2862543 Duplicates: 0 Warnings: 0	12.438 sec
✓	19	01:27:13	INSERT INTO logic.StudentsComments SELECT * FROM data.StudentsComments	281466 row(s) affected Records: 281466 Duplicates: 0 Warnings: 0	11.797 sec
✓	20	01:27:25	INSERT INTO logic.SupervisorsComments SELECT * FROM data.SupervisorsComments	402708 row(s) affected Records: 402708 Duplicates: 0 Warnings: 0	17.016 sec

Суммарное время загрузки: 72 сек

## BCNF: протокол запуска запросов от бизнеса

Запрос: Топ-5 студентов за последние 30 дней по средней оценке за любой один учебный курс

	StudentID	GroupID	FullName	MeanGrade
▶	50134	1604	ЧЕРНЫХ ПЁТР ИЛЬДАРОВИЧ	10.0000
	10532	330	УШАКОВ ГЕННАДИЙ МАНСУРОВИЧ	10.0000
	273301	8919	ЛУКИН СЕРГЕЙ РАВИЛЕВИЧ	10.0000
	195331	6361	ТОКАРЕВА ЭЛИНА ВАЛЕРИЕВНА	10.0000
	284182	9267	ЖУРАВЛЕВА ИННА НИКОЛАЕВНА	10.0000

176	03:54:33	CREATE TABLE ST...	1235712 row(s) affected Records: 1235712 Duplicates: 0 Wami...	28.828 sec
177	03:55:14	SELECT StudentID, ...	5 row(s) returned	2.484 sec / 0.000 sec

Суммарное время выполнения: 31.3 сек



## BCNF: протокол запуска запросов от бизнеса

Запрос: Топ-5 преподавателей за прошлый календарный месяц по количеству проверенных решений по всем учебным курсам

	Fullname	TASK_VOLUME
►	ФЕДОРОВА ИНГА ВАЛЕРЬЕВНА	265
	ИЛЬИН ВЛАДИСЛАВ АРКАДЬЕВИЧ	261
	ПРОХОРОВ СТЕПАН АНАТОЛЬЕВИЧ	255
	МЕЛЬНИКОВ ВИКТОР ВАЛЕНТИНОВИЧ	249
	ЮДИНА ЗАЛИНА ПАВЛОВНА	239

2	04:08:23	CREATE TABLE SUP...	43132 row(s) affected Records: 43132 Duplicates: 0 Warnings: 0	11.219 sec
3	04:08:36	SELECT Fullname, TA...	5 row(s) returned	0.015 sec / 0.000 sec

Суммарное время выполнения: 11.2 сек

## BCNF: протокол запуска запросов от бизнеса

Запрос: Доля сданных решений из всех возможных по заданиям с уже прошедшими сроками по каждому учебному курсу

	groupId	assignmentId	procent_of_submitted_decisions
	6	28	0.50000
	7	29	0.47101
	7	30	0.58696
	7	31	0.38587
	7	32	0.28261
	7	33	0.58696
	8	34	0.63492
	8	35	0.45238
	8	36	0.61905
	9	37	0.44444
	9	38	0.47222
	9	39	0.36111
	9	40	0.27778
	9	41	0.55556
	9	42	0.27778
	9	43	0.33333

35	05:08:43	CREATE TABLE SUBMITTED_IN...	5.719 sec
36	05:08:49	SELECT * FROM SUBMITTED_IN... 1000 row(s) returned	0.000 sec / 0.000 sec

Суммарное время выполнения: 5.7 сек

## BCNF: протокол запуска запросов от бизнеса

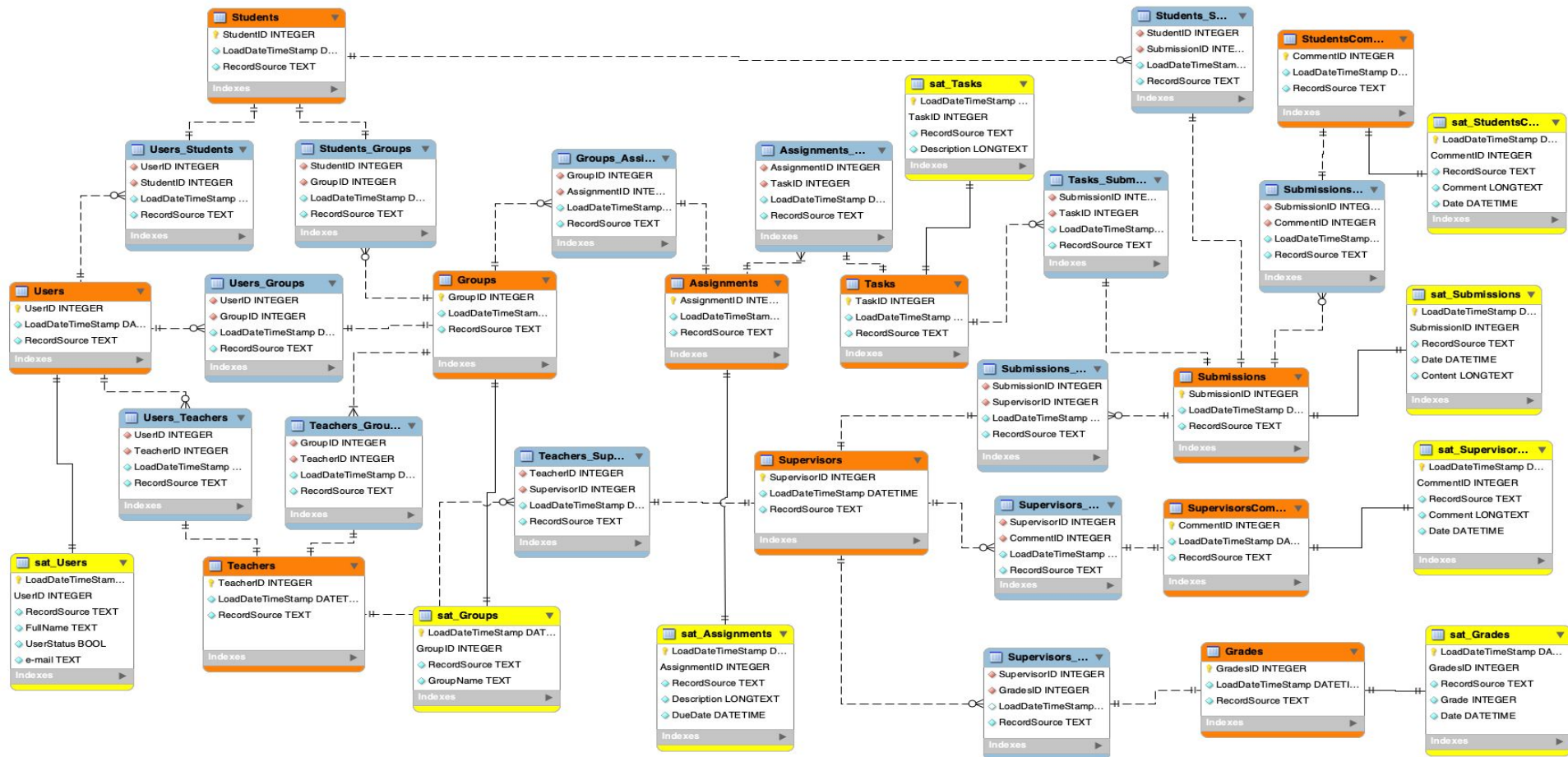
Запрос: Структура студентов по учебным курсам (т.е. учебные курсы числом студентов, отсортированные по убыванию)

	GroupID	count_students
▶	9494	54
	9571	54
	9592	54
	9734	54
	9763	54
	9821	54
	88	54
	129	54
	142	54
	180	54
	265	54
	334	54

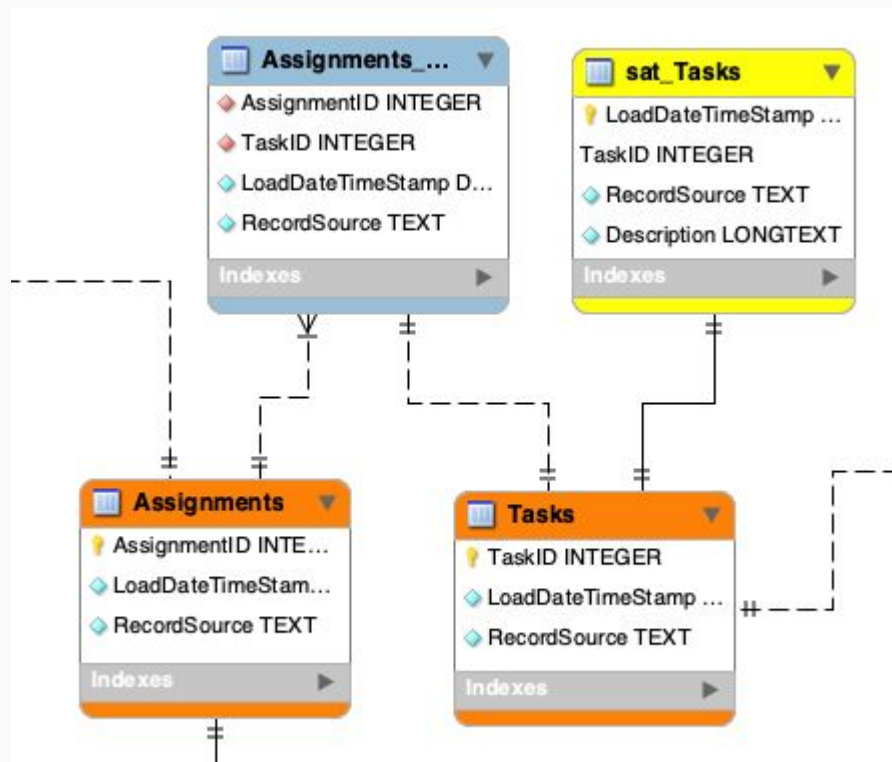
20	04:56:39	CREATE TABLE TASK4_INFO AS SE...	10000 row(s) affected Records: 10000 Duplicates: 0 Wa...	0.500 sec
21	04:56:39	SELECT * FROM TASK4_INFO LIMIT ...	1000 row(s) returned	0.000 sec / 0.000 sec

Суммарное время выполнения: 0.5 сек

# DataVault: логическая модель



# DataVault: логическая модель



Hub

Link

Satellite

LoadDateTimeStamp - метка времени загрузки

RecordSource - метка источника загрузки



# DataVault: протокол запуска запросов загрузки

119	02:55:23	INSERT INTO data_vault.Users SELECT Userid, current_timestamp() as LoadDateTimeStamp, 'data.Users' as RecordSource FROM d...	100000 row(s) affected Records: 100000 Duplicates: 0 Warnings: 0	0.532 sec
120	02:55:24	INSERT INTO data_vault.Groups SELECT GroupId, current_timestamp() as LoadDateTimeStamp, 'data.Groups' as RecordSource FR...	10000 row(s) affected Records: 10000 Duplicates: 0 Warnings: 0	0.063 sec
121	02:55:24	INSERT INTO data_vault.Students SELECT StudentId, current_timestamp() as LoadDateTimeStamp, 'data.Students' as RecordSource...	306450 row(s) affected Records: 306450 Duplicates: 0 Warnings: 0	2.250 sec
122	02:55:26	INSERT INTO data_vault.Teachers SELECT TeacherId, current_timestamp() as LoadDateTimeStamp, 'data.Teachers' as RecordSour...	43132 row(s) affected Records: 43132 Duplicates: 0 Warnings: 0	0.312 sec
123	02:55:26	INSERT INTO data_vault.Submissions SELECT SubmissionId, current_timestamp() as LoadDateTimeStamp, 'data.Submissions' as Re...	1988148 row(s) affected Records: 1988148 Duplicates: 0 Warnings: 0	15.234 sec
124	02:55:41	INSERT INTO data_vault.Assignments SELECT AssignmentId, current_timestamp() as LoadDateTimeStamp, 'data.Assignments' as Re...	49854 row(s) affected Records: 49854 Duplicates: 0 Warnings: 0	0.281 sec
125	02:55:42	INSERT INTO data_vault.Tasks SELECT TaskId, current_timestamp() as LoadDateTimeStamp, 'data.Tasks' as RecordSource FROM ...	124718 row(s) affected Records: 124718 Duplicates: 0 Warnings: 0	0.703 sec
126	02:55:42	INSERT INTO data_vault.Grades SELECT GradeId, current_timestamp() as LoadDateTimeStamp, 'data.Grades' as RecordSource FR...	2862543 row(s) affected Records: 2862543 Duplicates: 0 Warnings: 0	14.453 sec
127	02:55:57	INSERT INTO data_vault.SupervisorsComments SELECT CommentId, current_timestamp() as LoadDateTimeStamp, 'data.Supervisors...	402708 row(s) affected Records: 402708 Duplicates: 0 Warnings: 0	3.735 sec
128	02:56:01	INSERT INTO data_vault.StudentsComments SELECT CommentId, current_timestamp() as LoadDateTimeStamp, 'data.StudentsComm...	281466 row(s) affected Records: 281466 Duplicates: 0 Warnings: 0	2.890 sec
129	02:56:03	INSERT INTO data_vault.Supervisors SELECT SupervisorId, current_timestamp() as LoadDateTimeStamp, 'data.Supervisors' as Reco...	2859692 row(s) affected Records: 2859692 Duplicates: 0 Warnings: 0	20.172 sec
130	02:56:24	INSERT INTO data_vault.Users_Students select UserId, StudentId, current_timestamp() as LoadDateTimeStamp, 'data.Students' f...	306450 row(s) affected Records: 306450 Duplicates: 0 Warnings: 0	4.594 sec
131	02:56:28	INSERT INTO data_vault.Students_Groups select StudentId, GroupId, current_timestamp() as LoadDateTimeStamp, 'data.Studen...	306450 row(s) affected Records: 306450 Duplicates: 0 Warnings: 0	4.984 sec
132	02:56:33	INSERT INTO data_vault.Users_Groups select CreatorId, GroupId, current_timestamp() as LoadDateTimeStamp, 'data.Groups' fro...	10000 row(s) affected Records: 10000 Duplicates: 0 Warnings: 0	0.328 sec
133	02:56:34	INSERT INTO data_vault.Users_Teachers select UserId, TeacherId, current_timestamp() as LoadDateTimeStamp, 'data.Teacher...	43132 row(s) affected Records: 43132 Duplicates: 0 Warnings: 0	0.703 sec
134	02:56:34	INSERT INTO data_vault.Teachers_Groups select GroupId, TeacherId, current_timestamp() as LoadDateTimeStamp, 'data.Teach...	43132 row(s) affected Records: 43132 Duplicates: 0 Warnings: 0	0.641 sec
135	02:56:35	INSERT INTO data_vault.Groups_Assignments select GroupId, AssignmentId, current_timestamp() as LoadDateTimeStamp, 'data....	49854 row(s) affected Records: 49854 Duplicates: 0 Warnings: 0	0.625 sec
136	02:56:36	INSERT INTO data_vault.Assignments_Tasks select AssignmentId, TaskId, current_timestamp() as LoadDateTimeStamp, 'data.T...	124718 row(s) affected Records: 124718 Duplicates: 0 Warnings: 0	1.640 sec
137	02:56:37	INSERT INTO data_vault.Tasks_Submissions select SubmissionId, TaskId, current_timestamp() as LoadDateTimeStamp, 'data.Su...	1988148 row(s) affected Records: 1988148 Duplicates: 0 Warnings: 0	29.063 sec
138	02:57:06	INSERT INTO data_vault.Teachers_Supervisors select TeacherId, SupervisorId, current_timestamp() as LoadDateTimeStamp, 'da...	2859692 row(s) affected Records: 2859692 Duplicates: 0 Warnings: 0	42.515 sec
139	02:57:49	INSERT INTO data_vault.Submissions_Supervisors select SubmissionId, SupervisorId, current_timestamp() as LoadDateTimeStamp, ...	2859692 row(s) affected Records: 2859692 Duplicates: 0 Warnings: 0	53.140 sec
140	02:58:42	INSERT INTO data_vault.Students_Submissions select StudentId, SubmissionId, current_timestamp() as LoadDateTimeStamp, 'da...	1988148 row(s) affected Records: 1988148 Duplicates: 0 Warnings: 0	34.156 sec
141	02:59:16	INSERT INTO data_vault.Supervisors_SupervisorsComments select SupervisorId, CommentId, current_timestamp() as LoadDateTimeS...	402708 row(s) affected Records: 402708 Duplicates: 0 Warnings: 0	12.641 sec
142	02:59:29	INSERT INTO data_vault.Submissions_StudentsComments select SubmissionId, CommentId, current_timestamp() as LoadDateTimeS...	281466 row(s) affected Records: 281466 Duplicates: 0 Warnings: 0	7.390 sec
143	02:59:36	INSERT INTO data_vault.Supervisors_Grades select SupervisorId, GradeId, current_timestamp() as LoadDateTimeStamp, 'data.G...	2862543 row(s) affected Records: 2862543 Duplicates: 0 Warnings: 0	41.360 sec
144	03:00:18	INSERT INTO data_vault.sat_Users SELECT current_timestamp() as LoadDateTimeStamp, Userid, 'data.Users' as RecordSource, Full...	100000 row(s) affected Records: 100000 Duplicates: 0 Warnings: 0	1.671 sec
145	03:00:19	INSERT INTO data_vault.sat_Groups SELECT current_timestamp() as LoadDateTimeStamp, GroupId, 'data.Groups' as RecordSource...	10000 row(s) affected Records: 10000 Duplicates: 0 Warnings: 0	0.172 sec
146	03:00:19	INSERT INTO data_vault.sat_Assignments SELECT current_timestamp() as LoadDateTimeStamp, AssignmentId, 'data.Assignments' a...	49854 row(s) affected Records: 49854 Duplicates: 0 Warnings: 0	0.703 sec
147	03:00:20	INSERT INTO data_vault.sat_Submissions SELECT current_timestamp() as LoadDateTimeStamp, SubmissionId, 'data.Submissions' as...	1988148 row(s) affected Records: 1988148 Duplicates: 0 Warnings: 0	33.094 sec
148	03:00:53	INSERT INTO data_vault.sat_Tasks SELECT current_timestamp() as LoadDateTimeStamp, TaskId, 'data.Tasks' as RecordSource, D...	124718 row(s) affected Records: 124718 Duplicates: 0 Warnings: 0	1.391 sec
149	03:00:55	INSERT INTO data_vault.sat_Grades SELECT current_timestamp() as LoadDateTimeStamp, GradeId, 'data.Grades' as RecordSource...	2862543 row(s) affected Records: 2862543 Duplicates: 0 Warnings: 0	31.797 sec
150	03:01:26	INSERT INTO data_vault.sat_StudentsComments SELECT current_timestamp() as LoadDateTimeStamp, CommentId, 'data.StudentsC...	281466 row(s) affected Records: 281466 Duplicates: 0 Warnings: 0	12.203 sec
151	03:01:39	INSERT INTO data_vault.sat_SupervisorComments SELECT current_timestamp() as LoadDateTimeStamp, CommentId, 'data.Superi...	402708 row(s) affected Records: 402708 Duplicates: 0 Warnings: 0	21.187 sec

Суммарное время загрузки: ~ 8 мин.

# DataVault: протокол запуска запросов от бизнеса

Запрос: Доля сданных решений из всех возможных по заданиям с уже прошедшими сроками по каждому учебному курсу

	groupId	assignmentId	procent_of_submitted_decisions
	6	28	0.50000
	7	29	0.47101
	7	30	0.58696
	7	31	0.38587
	7	32	0.28261
	7	33	0.58696
	8	34	0.63492
	8	35	0.45238
	8	36	0.61905
	9	37	0.44444
	9	38	0.47222
	9	39	0.36111
	9	40	0.27778
	9	41	0.55556
	9	42	0.27778
	9	43	0.33333

26	04:59:44	CREATE TABLE TASK...	26.688 sec
27	05:00:11	SELECT * FROM TASK.... 1000 row(s) returned	0.000 sec / 0.000 sec

Суммарное время выполнения: 26.7 сек

# DataVault: протокол запуска запросов от бизнеса

Запрос: Структура студентов по учебным курсам (т.е. учебные курсы числом студентов, отсортированные по убыванию)

	GroupID	count_students
▶	9494	54
	9571	54
	9592	54
	9734	54
	9763	54
	9821	54
	88	54
	129	54
	142	54
	180	54
	265	54
	334	54

38	05:11:12	CREATE TABLE TASK4_I...	10000 row(s) affected Records: 10000 Dupli...	5.547 sec
39	05:11:17	SELECT * FROM TASK4_I...	1000 row(s) returned	0.000 sec / 0.000 sec

Суммарное время выполнения: 5.547 сек



## Сравнение моделей BCNF и DataVault

- Запросы загрузки выполняются значительно дольше для модели DataVault
- Для выполненных запросов результаты для обеих моделей совпадают
- DataVault неэффективен для аналитических задач
  - Запросы сложнее реализуются, чем для модели BCNF
  - Время выполнения сильно больше в сравнении с моделью BCNF

# Заключение

- Было произведено моделирование предметной области в соответствии с 2 логическими моделями данных
- Были созданы схемы хранилища данных
- Реализованы процедуры загрузки данных в хранилища
- Созданы витрины под бизнес-запросы
- Произведено качественное сравнение двух моделей