

ЗАДАНИЕ 8. КОМАНДА А

**ИНФОРМАЦИОННЫЕ МОДЕЛИ ДЛЯ СИСТЕМЫ
ПОДДЕРЖКИ ВЫДАЧИ И ПРИЕМА ЗАДАНИЙ**

КОРОЛЕВ НИКОЛАЙ СЕРГЕЕВИЧ

ЕРЕМЕЕВ МАКСИМ АЛЕКСЕЕВИЧ

ПОПОВ ДМИТРИЙ ОЛЕГОВИЧ

НАХОДНОВ МАКСИМ СЕРГЕЕВИЧ

ЧЕРНЫШЁВ АЛЕКСАНДР ВЛАДИСЛАВОВИЧ

МГУ ВМК. ПРАКТИКУМ

МОСКВА, 2019

КОРОЛЕВ НИКОЛАЙ

- РЕАЛИЗАЦИЯ МОДЕЛИ DATA VAULT
- РЕАЛИЗАЦИЯ МОДЕЛИ BCNF
- ЗАГРУЗКА СГЕНЕРИРОВАННЫХ ДАННЫХ

ЕРЕМЕЕВ МАКСИМ

- РЕАЛИЗАЦИЯ МОДЕЛИ DATA VAULT
- РЕАЛИЗАЦИЯ МОДЕЛИ BCNF
- ПРЕЗЕНТАЦИЯ, ДОКЛАД

ПОПОВ ДМИТРИЙ

- ГЕНЕРАЦИЯ ДАННЫХ
- РЕАЛИЗАЦИЯ МОДЕЛИ BCNF
- РЕАЛИЗАЦИЯ БИЗНЕС-ЗАПРОСОВ

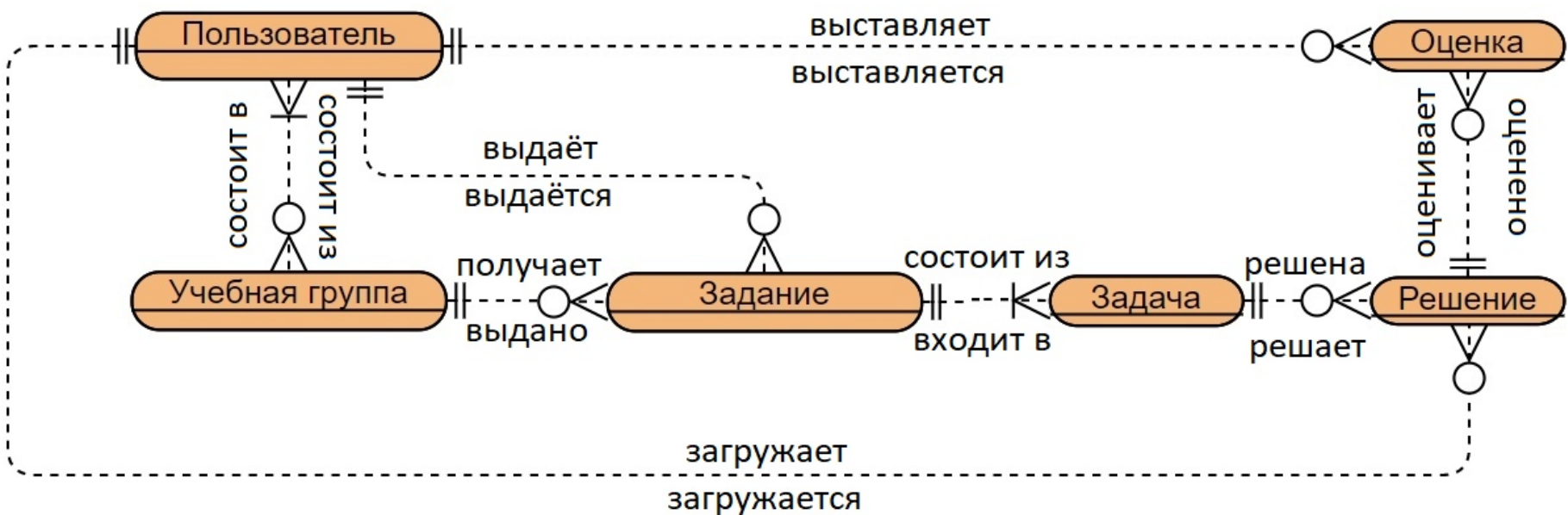
НАХОДНОВ МАКСИМ

- РЕАЛИЗАЦИЯ БИЗНЕС-ЗАПРОСОВ ДЛЯ BCNF
- РЕАЛИЗАЦИЯ БИЗНЕС-ЗАПРОСОВ ДЛЯ DATA VAULT
- РЕАЛИЗАЦИЯ МОДЕЛИ DATA VAULT

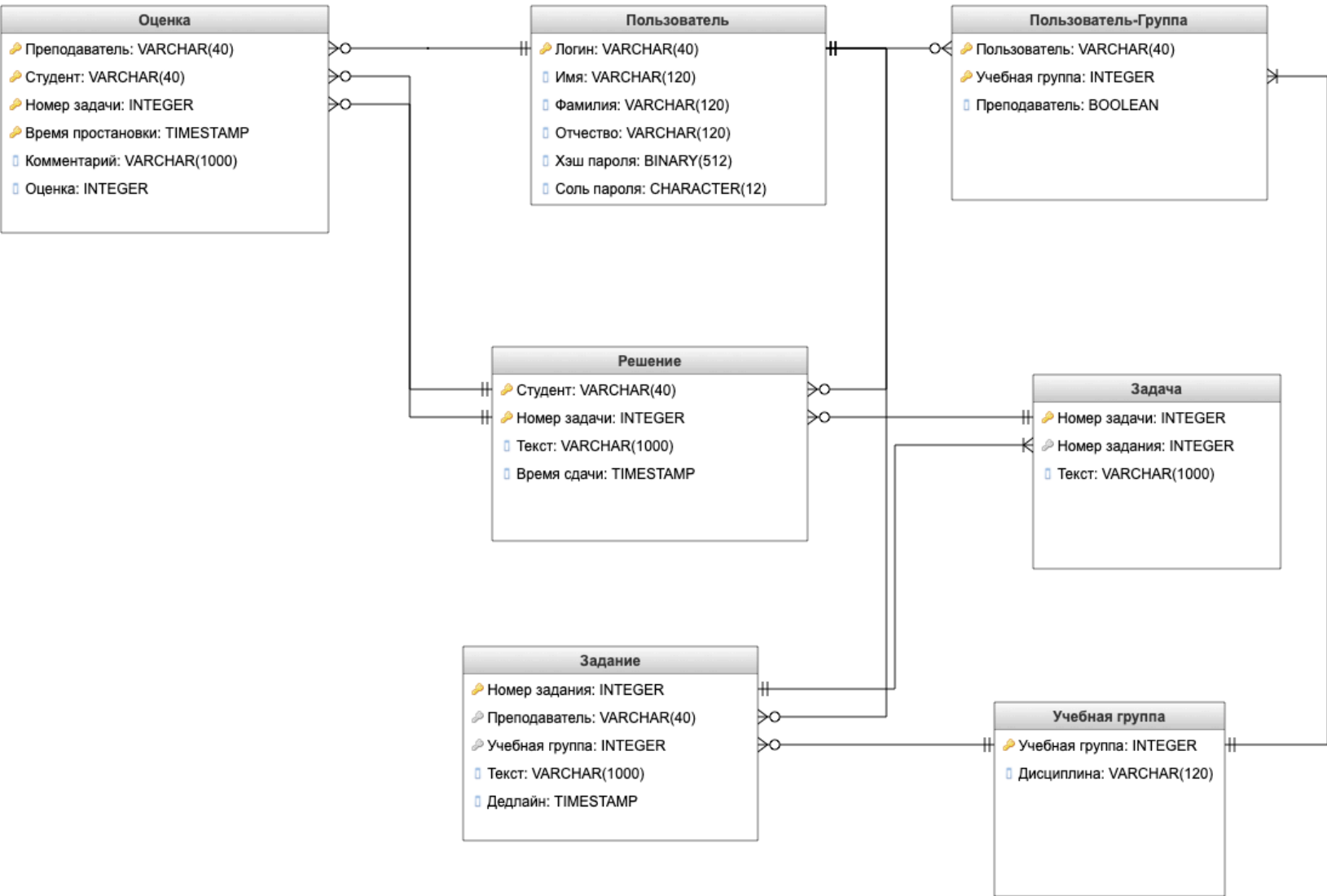
ЧЕРНЫШЁВ АЛЕКСАНДР

- ЗАГРУЗКА СГЕНЕРИРОВАННЫХ ДАННЫХ
- РЕАЛИЗАЦИЯ БИЗНЕС-ЗАПРОСОВ ДЛЯ BCNF
- РЕАЛИЗАЦИЯ БИЗНЕС-ЗАПРОСОВ ДЛЯ DATA VAULT

Концептуальная



Логическая



Физическая

login	first_name	last_name	middle_name	password_hash	password_salt	active	super_user
a00yuY43g	Liam	Hall	William	f08319f87e0f6dd819ceb259487ffe5f695e9cf8602cd151650f9af2a... b5807aaa34d5		1	0
a01OEyRU6	Loui	Jones	William	f7452534a98c9e99c9b040a55aff809a38da5732d47c7ca4727f95... e97024b121fb		1	0
A02T9zWuX	William	Smith	▼ William	07d0a5263d18c8ff33637c7c028e8425cebee338376ca3dbbd483... 8dc00895c7b2		1	0
A05103f3V	William	Brown	Elias	803489f5d716838228edd8f48c0f95825d7cebb00241e5197ceb0f... 86b7d897ee33		1	0
A05Gp3dLn	William	Jones	Ludvig	cbab49d15ffb07b51bf1ac677f43a0c0d8f24ce31f0d1cd1e58afa5f... 421979b3c990		1	0
a05jm7jQB	William	Smith	William	24040327317ca1681aacf777eec795dcf4cd6318fb2f407646ae9e... 6f53e72502d7		1	0
A06KL3TW0	Lucas	Jones	Noah	9e829484b131a1d3284833e620c72d26c2b14b6487f19aa41d70... cf2e0e55790c		1	0

ПРИМЕР СГЕНЕРИРОВАННЫХ ДАННЫХ

Tutors (9012 rows)

	group_key	tutor_login
1	0	NDqS5FSP5
2	1	qzfxz65Tg
3	2	oyhh4ZzFK
4	2	t5gAt58il
5	2	YXchjZqwm

Users (100 000 rows)

	login	name	last_name	middle_name	password_hash	password_salt	active	super_user
1	A00oYQUa5	Noah	Porter	Adrian	a70a5195f7d9a0877c56ec6443ed65fc8bb4dd4d5df33dbea046d7a09...	c85aceb0de2c	1	0
2	a017ROngQ	William	Porter	Noah	96fd1b5f39be5fdc0ae518c8cc662a7f17b3034ee4d5f69d232dfa76555...	503798ba2cdf	1	0
3	a01CaqshB	Lucas	Smith	William	3d504a775da4d66e4d972adf4d1cad4bc2b6eb8660926360de4f54b82...	1e6c3146f3b1	1	0
4	A029bCDtt	Olle	Williams	Lias	d2e95fb4222b607bf3c3e18480081f413e5cf9390428f56e50803f381a5...	2303416b5a87	1	0
5	A02cmzIxd	Liam	Hall	Elias	d01a5f906b77aadd3e53b8d8c03461e6b0262e5296597fd2bbd191903...	2ccfdb475da	1	0

ПРИМЕР СГЕНЕРИРОВАННЫХ ДАННЫХ

Students (17 722 rows)

	group_key	student_login
1	0	A3n896BpM
2	0	cZTPDLuQX
3	0	l9ybMG2AC
4	0	TVg7mobzA
5	0	UY7upp4Ea

Tasks (18 184 rows)

	assignment_id	text	task_id
1	0	wZ4r KxKI(hl?WEjmFuJkRSK(mOe0UXWkrxNX.pnzZUKZnt?hJ?qhqLGb2XW9.pOdU.8!FMiZw?XZ.mv67)ptbzFQ6hRYDr7k1)jpsee7yPuubH?2V7iG4AP...	0
2	0	w!GgTbZ7CdY(zW!9gu1ObxOoMQ!lpJDv)ahoLiIVQ,.7(haF63zdnPHdndw9lu4q96440UoSmeLt QT5uBcGo6tzxvEKGq6My61?M)cobSyqOaHSSo4TBNJ...	1
3	0	JdksKTmyJCU?NPeGqeSi.uQRt0thZLG?mvjUqh5P!brA?lcCUlvOe6?ZPq?qJS5R,X 1UUQSJ.bUBgrTeitVxQ??rQvS4aq8seK)JqZpG.vvJx5bQbB ZD2bX...	2
4	0	.tOfBT!IN4HUNFv5VPiGLOSZpv2x(!n.x4op1Thv2HuVsEtIH,AJ9A0nvmMOquuQZ(y.xQ1?YMwGYzJqnnTGqv1WL.2o4f3Thqj)wyX4MniPQRLjdNL52oeQ,...	3
5	1	pKNnRhHwgi6xCY!z2N?7zD LUNrxiXV.Rz.,8f4f1amNiu1HqJXnEtYXM1UXlh4AoOC484Dqf.ZK93hR8nr7I7TDk11egPO5QY6Ym4iVBK0xBxkoa583D(.njf...	4

ПРИМЕР СГЕНЕРИРОВАННЫХ ДАННЫХ

Assignments (6041 rows)

	author	group_key	text	deadline	assignment_id
1	NDqS5FSP5	0	hZ,reD7Vfe5WHhHxdWDoPcv1O)4sEi)J2(aqZ2uyd!nmuZKUWqZ.lbY88d10e2,FjZuLMOpIPt.SKnK7S7jhO!G...	2019-02-02 06:32:37	0
2	NDqS5FSP5	0	Q?fW96!iSSiggyKSSyv,FPgxSdkmV)3(tgi,dod1tzkWsnSITd(MJtpMvkNISHf(CQGuS?3MF0?5mYDy.3ayAw3...	2018-06-11 10:08:01	1
3	NDqS5FSP5	0	EMs2FnDW(2LREkG?GfzDHq)jM!ZX57wXedPSluXa bFY0LsGtqCVWqil!jDXfICysXS 7Vhyc OTBP5N5pm32...	2018-08-30 22:43:06	2
4	NDqS5FSP5	0	ekp4qcuLfQvgRQFo4PrSrA7T)8od.op8SYn944 Tu!FZ8a94y7W4CXi01CRSuxZdhHDd2s3BehKM4MXHNP2...	2018-10-11 10:29:04	3
5	qzfxz65Tg	1	Zm539INMm4qR8BbM.xlGN3Tg)KrmL9bl2x(h2LSt20)qTPYi3H518ra71T9N7st68QFIBk?CLqPA!cmIxUTkNR...	2018-09-17 02:42:28	4

Grades (16 452 rows)

	grader	student	task_id	date	comment	grade	grade_id
1	a0hhfW8iY	JxenDugoJ	106	2018-02-28 04:23:44	ksSF8.yyyKo5BbB,3kLKyi4TeHJDRUKtrny1GfOxYQEKCtdhDU)7EJa.2l?S!UveR,?yC	0	7663
2	a0hhfW8iY	yKEMg4yo9	101	2019-08-17 18:36:35	Gz90AGK68tmOkGgi3K!mXE(lSHHAf5oX5? a449!N56W?HAH5rbSa4p,,k?tAeUGp	2	10731
3	A0rxCtO2V	rbUj0XmOS	11473	2018-12-05 07:27:49	Gl?uj5g?.h,s,RwTEJsEx j7CgUBjY4pA2B0!(,rTE 7XIWmQKIF9uixeoZbIFUZA9.!yAYGNchHplj r 4		9611
4	A0rxCtO2V	rbUj0XmOS	11479	2015-07-06 18:01:56	izEkR!vJT47la0Vb9FO17Ma!U4CrnFadnqm ?TylhW1clmT4GCHM0u0suCE9Gq(9E6(t	7	11397
5	A0rxCtO2V	y2Cboo7nU	11480	2019-11-13 12:11:49	wm2813OKjjcX2Jxj,VVPwGsy3G OTeTuE!8e5AYAAa5mVX7Rz!3L4VZx(lvchRHk1(H,.azqDuxw...	6	3931

ПРИМЕР СГЕНЕРИРОВАННЫХ ДАННЫХ

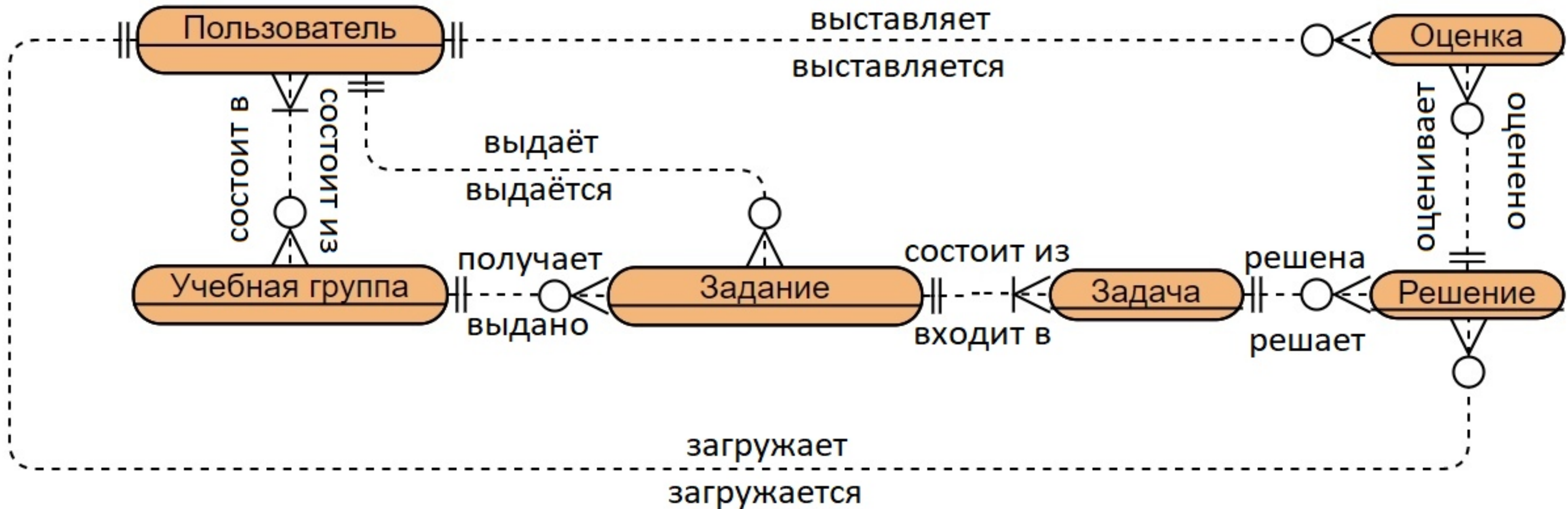
Groups (3000 rows)

	group_key	discipline
1	0	IzaECNUZk
2	1	O5v1XGTCa
3	2	rM7R2PYOB
4	3	NVvxlQmyy
5	4	Hn8w98rwX

Solutions (65 809 rows)

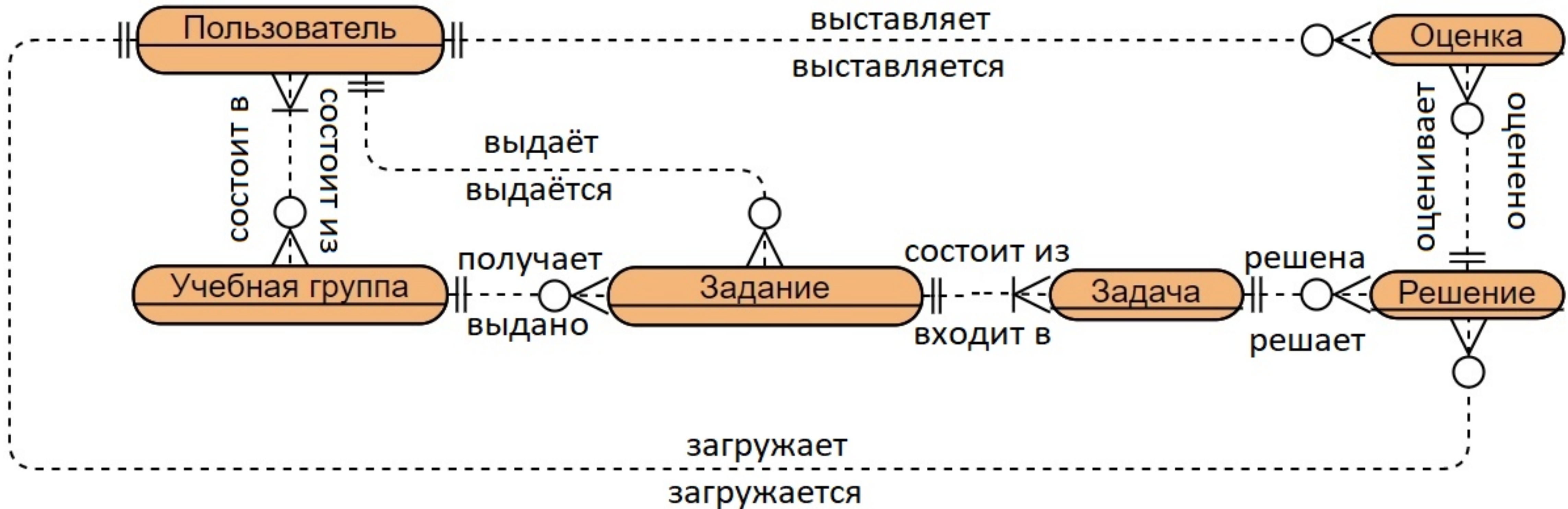
	assignment_id	task_id	student	date	text	solution_id
1	0	0	A3n896BpM	2017-05-04 00:42:42	K(C31CfTag8Fqr8w3hB?ob0Ay(7te9VGfw919JYvIEWJx9n,qzwZbRu6.WuvF SnmEtM0CXkbw...	2
2	0	0	cZTPDLuQX	2017-08-07 12:45:02	qya!QCK)J25NzLUNWCQyZzm4k7EgOHK(l8.YbrbLI8bXoc!9R5MR.Jh	1
3	0	0	TVg7mobzA	2017-12-31 00:02:14	yQcWfHhQ7DZ1Ht3)4BIKIBi1vqSR9VHBOwE8J	0
4	0	1	A3n896BpM	2018-08-24 08:46:42	FbYmb5slEjQm!kziwCVgBQVmCU?xJ8Kz651vFvYyut8gZ0cwm9kU7X43NdtvabQgM.fH0Awsy...	4
5	0	1	cZTPDLuQX	2017-06-08 11:05:05	J,AonPfG(XPXUO7ewXGuwe81A(ECderjJQ2eV7N!?mbt!hxUBdBT)f81d5ekCeK71Wbft Nw2L...	3

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ



Добавлены названия связей

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ



6 сущностей, 8 связей

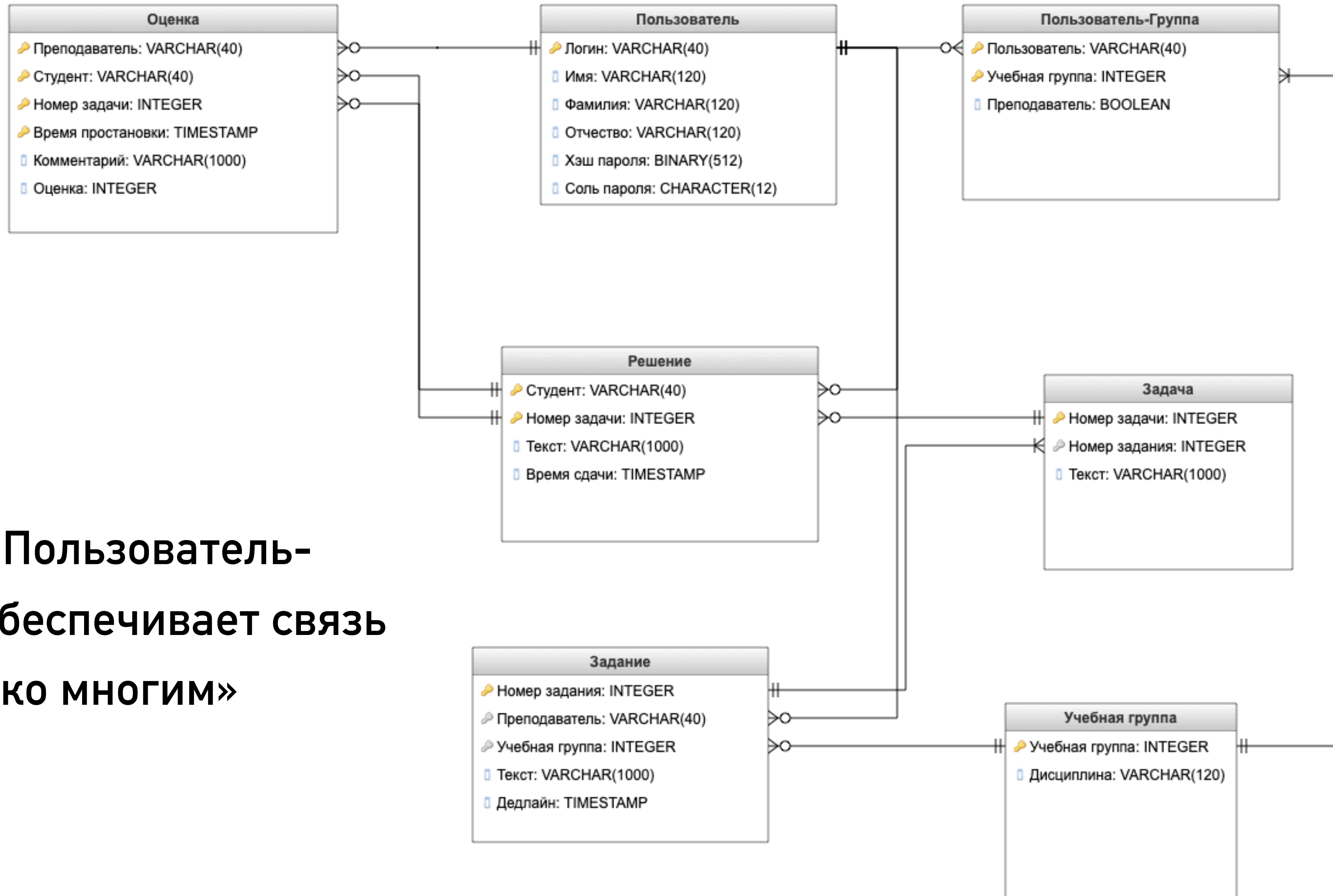
Пользователь - Учебная группа — единственная связь “многие ко многим”

Пользователь — единая сущность для студентов и преподавателей

ЛОГИЧЕСКАЯ МОДЕЛЬ. BCNF

7 таблиц

Таблица Пользователь-
Группа обеспечивает связь
«многие ко многим»



	Assignment	Grades	Groups	Solutions	Tasks	Users	User-Group
Размер	6041	16 452	3000	65 809	18 184	100 000	26 734

На заполнение таблиц: 39 секунд

Единый запрос загрузки на 114 строк

Пример запроса в задаче 4. Структура студентов по учебным курсам

```
SELECT
    group_id,
    COUNT(*) AS students_amount
FROM
    s_gr1.bcnf_user_group AS ug
WHERE
    ug.tutor = 0
GROUP BY
    ug.group_id
ORDER BY
    students_amount DESC
```

group_id	students_amount
13	12
15	12
25	12
31	12
44	12
46	12
77	12
88	12
94	12
96	12
101	12
103	12
112	12
114	12
120	12
122	12

0.4 сек

Топ 5 студентов за последние 30 дней по средней оценке

student
Ah2cEeIY2
ahBvFSUJr
AmA8ZIFJR
apZJANIEo
bZkpCGGWF

0.9 сек

Топ 5 преподавателей за прошлый календарный месяц

login
CmYfp1TWH
EW5FdnTyR
Ds4ID4oFZ
bmUSr5VVd
d07d5WLSC

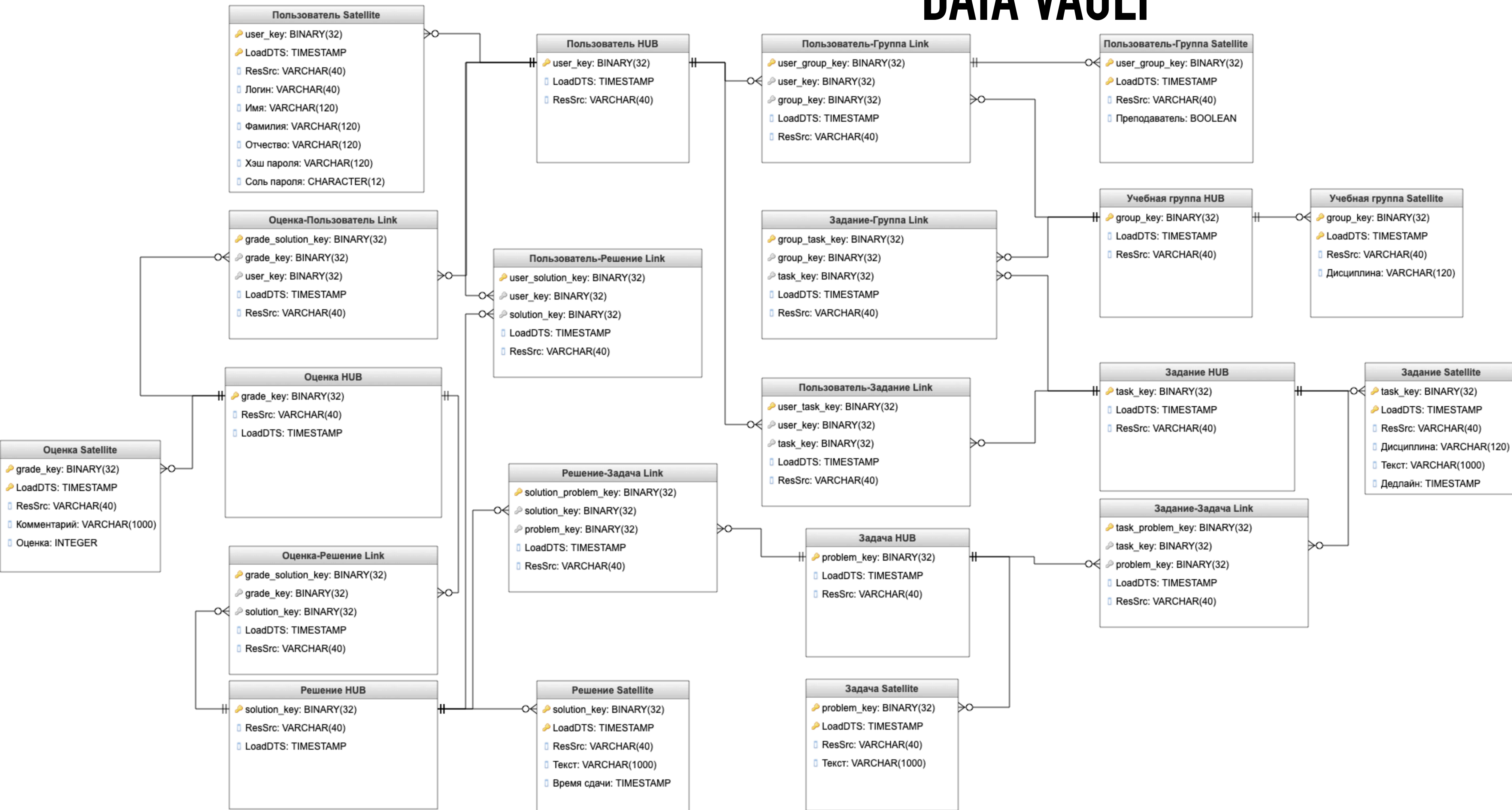
0.6 сек

Доля сданных решений из всех возможных по просроченным заданием

group_id	part_of_solved
0	0.6
1	0.6
2	0.5
3	0.6
4	0.66666667
6	0.6
8	0.66666667
10	0.5
11	0.6
12	0
13	0.66666667

0.5 сек

DATA VAULT



	Assignment- HUB	Grades- HUB	Groups- HUB	Solutions- HUB	Tasks- HUB	Users- HUB
Размер	6041	16 452	3000	65 809	18 184	100 000

	Assignment- Satellite	Grades- Satellite	Groups- Satellite	Solutions- Satellite	Tasks- Satellite	Users- Satellite	Users-Group- Satellite
Размер	6041	16 452	3000	65 809	18 184	100 000	26 734

	Assignment- Group-LINK	Grade- Solution- LINK	Grade- User- LINK	Solution- Task-LINK	Task- Assignment- LINK	User- Assignment- LINK	User- Group- LINK	User- Solution- LINK
Размер	6041	16 452	16 452	65 809	18 184	6041	6041	65 809

На заполнение таблиц: 3 минуты 24 секунды

Единый запрос загрузки на **490** строк!


Group-HUB

	group_key	LoadDTS	ResSrc	
1	1	2019-12-18 04:01:55	junk	
2	2	2019-12-18 04:01:55	junk	
3	3	2019-12-18 04:01:55	junk	
4	4	2019-12-18 04:01:55	junk	
5	5	2019-12-18 04:01:55	junk	

Task-Satellite

	LoadDTS	task_key	ResSrc	text
1	2019-12-18 04:01:11	1	junk	wZ4r KxKI(hl?WEjmFuJkRSK(mOe0UXWkrxNX.pnzZUKZnt?hJ?qhqLGb2XW9.pOdU.8!FMiZw?XZ.mv67)ptbzFQ6hRYDr7k1))jpsee7y...
2	2019-12-18 04:01:11	2	junk	w!GgTbZ7CdY(zW!9gu1ObxOoMQ!!pJDv)ahoLiIVQ,.7(haF63zdnPHdndw9lu4q96440UoSmeLt QT5uBcGo6tzxEKGq6My61?M)cobS...
3	2019-12-18 04:01:11	3	junk	JdksKTmyJCU?NPeGqeSi.uQRt0thZLG?mvjUqh5P!brA?lcCUIvOe6?ZPq?qJS5R,X 1UUQSJ.bUBgrTeitVxQ??rQvS4aq8seK)JqZpG.v...
4	2019-12-18 04:01:11	4	junk	.tOfBT!!N4HUNFv5VPiGLOSZpv2x(!n.x4op1Thv2HuVsEtIH,AJ9A0nvmMOquuQZ(y.xQ1?YMwGYzJqnnTGqv1WL.2o4f3Thqj)wyX4Mn...
5	2019-12-18 04:01:11	5	junk	pKNnRhHwgi6xCY!z2N?7zD LUNrxiXV.Rz.,8f4f1amNiu1HqJXnEtYXM1UXlh4AoOC484Dqf.ZK93hR8nr7I7TDk11egPO5QY6Ym4iVB...

Task-Assignment-LINK

	task_assignment_key	LoadDTS	task_key	assignment_key	ResSrc	
	1	2019-12-18 04:01:58	1	1	junk	
	2	2019-12-18 04:01:58	2	1	junk	
	3	2019-12-18 04:01:58	3	1	junk	
	4	2019-12-18 04:01:58	4	1	junk	
	5	2019-12-18 04:01:58	5	2	junk	

Пример запроса в задаче 4. Структура студентов по учебным курсам

```
SELECT
    ugl.group_key,
    COUNT(*) AS students_amount
FROM
    s_gr1.datavault_user_group_link AS ugl
INNER JOIN
    s_gr1.datavault_user_group_sattelite AS ugs
ON
    ugl.user_group_key = ugs.user_group_key
WHERE
    ugs.tutor = 0
GROUP BY
    ugl.group_key
ORDER BY
    students_amount DESC
```

group_id	students_amount
13	12
15	12
25	12
31	12
44	12
46	12
77	12
88	12
94	12
96	12
101	12
103	12
112	12
114	12
120	12
122	12

0.9 сек

Топ 5 студентов за последние 30 дней по средней оценке

student	mean_grade
zqYNzw8eZ	9
z2wp1BR8l	9
yIX7YJxjS	9
yaQnXENjf	9
XwfhVVQcy	9

1.2 сек

Топ 5 преподавателей за прошлый календарный месяц

login
CmYfp1TWH
EW5FdnTyR
Ds4ID4oFZ
bmUSr5VVd
d07d5WLSC

0.8 сек

Доля сданных решений из всех возможных по просроченным заданием

group_id	part_of_solved
0	0.6
1	0.6
2	0.5
3	0.6
4	0.66666667
6	0.6
8	0.66666667
10	0.5
11	0.6
12	0
13	0.66666667

0.5 сек



Данные в BCNF загружаются в три раза быстрее, чем в Data Vault

Бизнес-запросы можно писать для обеих моделей, скорость работы запросов примерно одинаковая

Data Vault менее наглядна, ее сложнее проектировать на логическом и физическом уровнях

ИНТЕРЕСНЫЕ ФАКТЫ

В ходе экспериментов данные были загружены на сервер более 25 раз (> 75 минут ожидания загрузки)

На редактирование кода для Data Vault ушло более трех часов

Моделирование данных - действительно непростая работа

РЕЗУЛЬТАТЫ

Система для работы с данными
доступна уже сегодня!

Физические модели полностью
отражает устройство логических
моделей

Можете работать или в модели
BCNF, или в модели Data Vault

На синтетических данных
проведено сравнение двух
моделей, BCNF лучше для данной
задачи