

Глава 13. Загрузка данных из Exchange в Power Query

Это продолжение перевода книги Кен Пульс и Мигель Эскобар. Язык M для Power Query. Главы не являются независимыми, поэтому рекомендую читать последовательно.

[Предыдущая глава](#) [Содержание](#) [Следующая глава](#)

В электронной почте и календаре есть море информации, которая может быть полезна для бизнеса, но к ней довольно трудно добраться. Power Query облегчает извлечение такой информации. Существует пять основных аспектов, к которым можно подключиться в базе данных Exchange: почта, календарь, контакты, задачи, приглашения на собрания. Рассмотрим несколько сценариев и способов использования сервера Exchange.

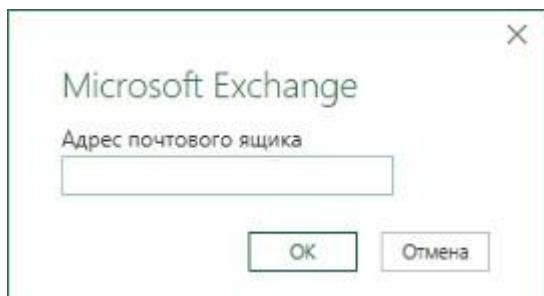


Рис. 13.1. Окно подключения к Microsoft Exchange

Статистическая обработка данных опроса. С помощью Power Query вы можете создать опрос с таблицей в теле письма и отправить его аудитории. Затем вы можете использовать Power Query для чтения ответов, извлечения информации из отдельных таблиц и объединения таблиц в источник данных. Вам также не нужно беспокоиться, если кто-то отправит ответ позже. Вы просто кликните Обновить.

Агрегирование бюджетов. Вы можете отправить книги Excel по почте и попросить сотрудников заполнить информацию по своему подразделению. Power Query найдет письма с определенной темой и/или вложением, а затем откроет книги Excel и объединит их в единый набор данных.

Дашборд алертов. Если вы по электронной почте получаете различные уведомления от серверов и оборудования, вы можете использовать Power Query для агрегирования сообщений, построения диаграмм объема и серьезности проблем и создания информационной панели (дашборда).

Подключение к Microsoft Exchange

Этот пример вы сможете реализовать, если ваша рабочая почта реализована с помощью Microsoft Exchange. Создайте новую книгу Excel. Пройдите по меню *Данные* → *Получить данные* → *Из других источников* → *Из Microsoft Exchange* (рис. 13.1). Введите свой корпоративный адрес с указанием домена. Нажмите *Ok*. В окне авторизации введите пароль, нажмите *Подключение*.

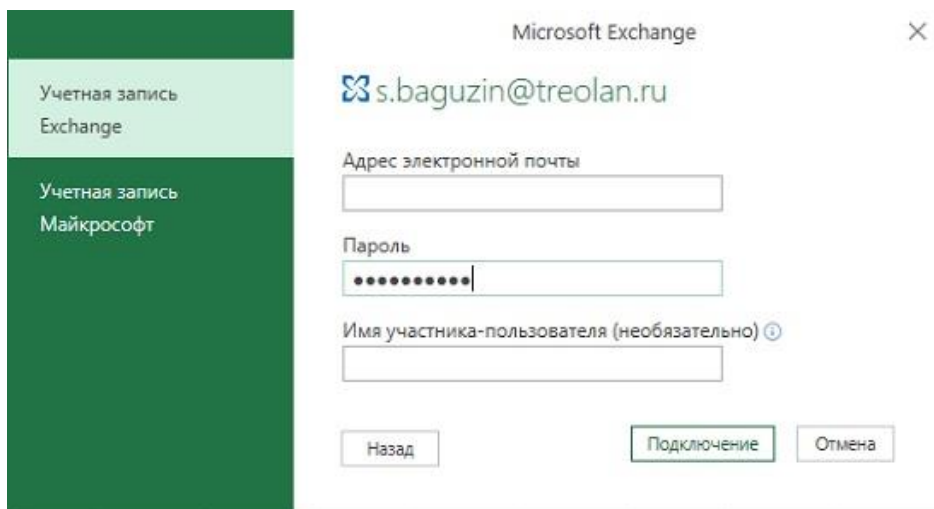


Рис. 13.2. Окно авторизации доступа к Microsoft Exchange (поскольку пример основан на моей корпоративной почте, я удалил всю конфиденциальную информацию)

Появится окно службы автообнаружения Microsoft Exchange. Кликните *Разрешить*.

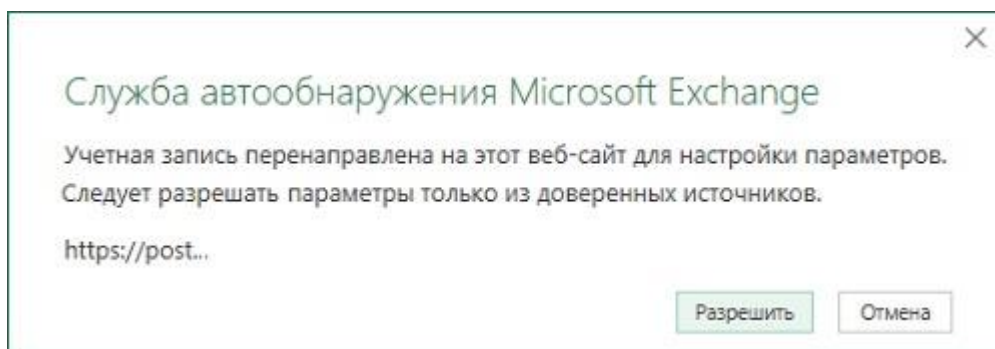


Рис. 13.3. Окно службы автообнаружения Microsoft Exchange

Как правило, служба автообнаружения уже включена и настроена сотрудниками ИТ-отдела вашей компании. Благодаря этому вы можете удаленно получать доступ к своей электронной почте с мобильных телефонов.

После небольшой задержки, вызванной подключением Power Query к серверу Exchange, вы попадете в окно *Навигатор*. Выберите *Mail* → *Изменить*.

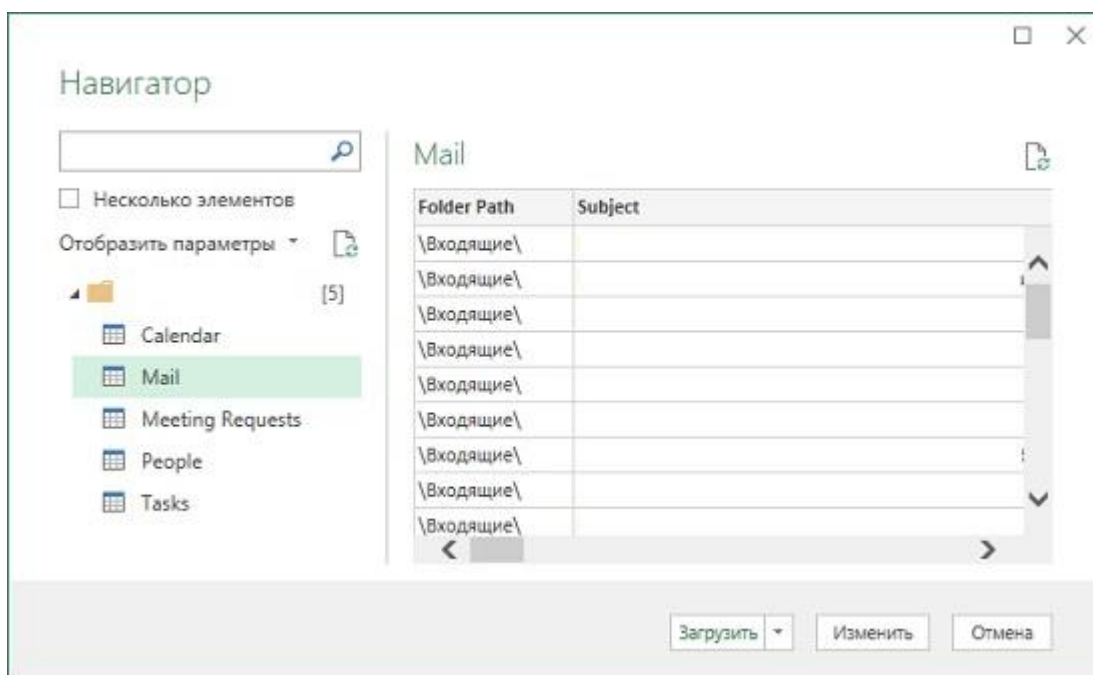


Рис. 13.4. Окно *Навигатор*

Не нажимайте кнопку *Загрузить* после выбора таблицы *Mail*. Ожидание будет очень долгим, т.к. Power Query будет загружать весь ваш почтовый ящик!

	Folder Path	Subject	Sender	DisplayTo	DisplayCc
1	\Входящие\		Record	null	null
2	\Входящие\		Record	null	null

Рис. 13.5. Окно редактора Power Query с таблицей писем

Чтобы уменьшить количество писем для анализа, вы перенаправили нужные письма в папку *_in*. Теперь вам нужно отфильтровать столбец путь к папке *Folder Path*. Интересно, что при попытке установить фильтр нужная папка не отображается:

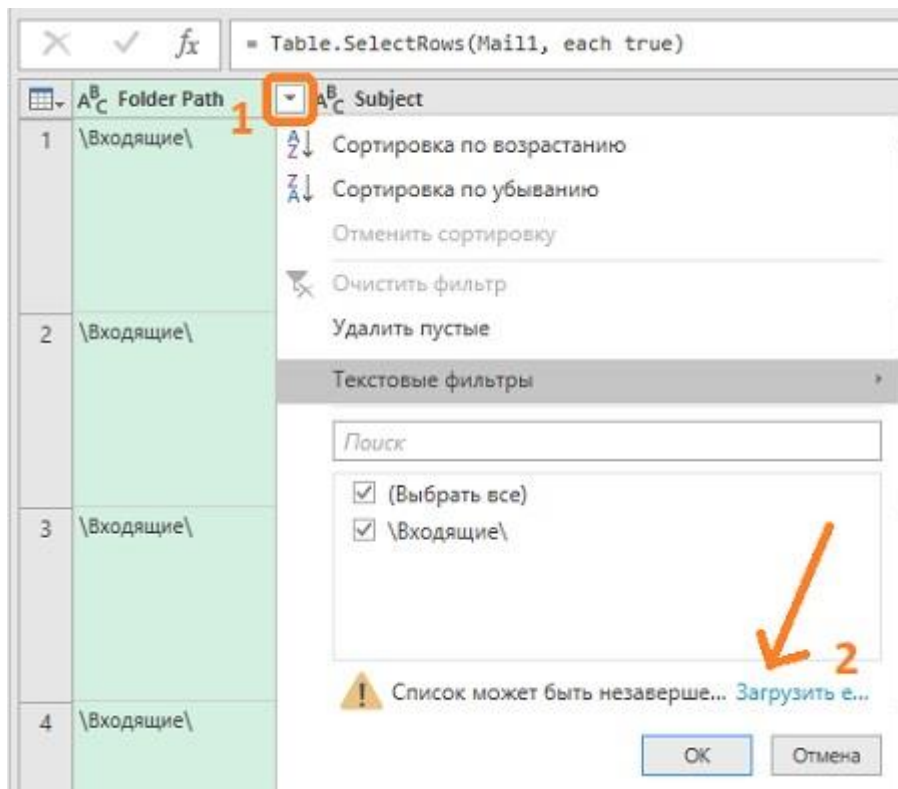


Рис. 13.6. Нужная папка не отражается в списке доступных

Кликните *Загрузить еще...* в правом нижнем углу. По умолчанию этот список сокращается до ограниченного объема данных, отображаемых в предварительном просмотре, но при нажатии кнопки отображаются все доступные опции. Когда список появляется в полном объеме, выбираете нужную папку, нажмите *Ok*. Список писем отфильтруется.

Таблица содержит около 20 столбцов, включая такие полезные поля, как *To*, *CC*, *Sender*, *Attachments* и многое другое. Для целей нашего исследования требуются, только поля *DateTimeSent* (время отправки) и *Body* (тело письма). Выбираете эти два столбца и кликаете *Удалить другие столбцы*:

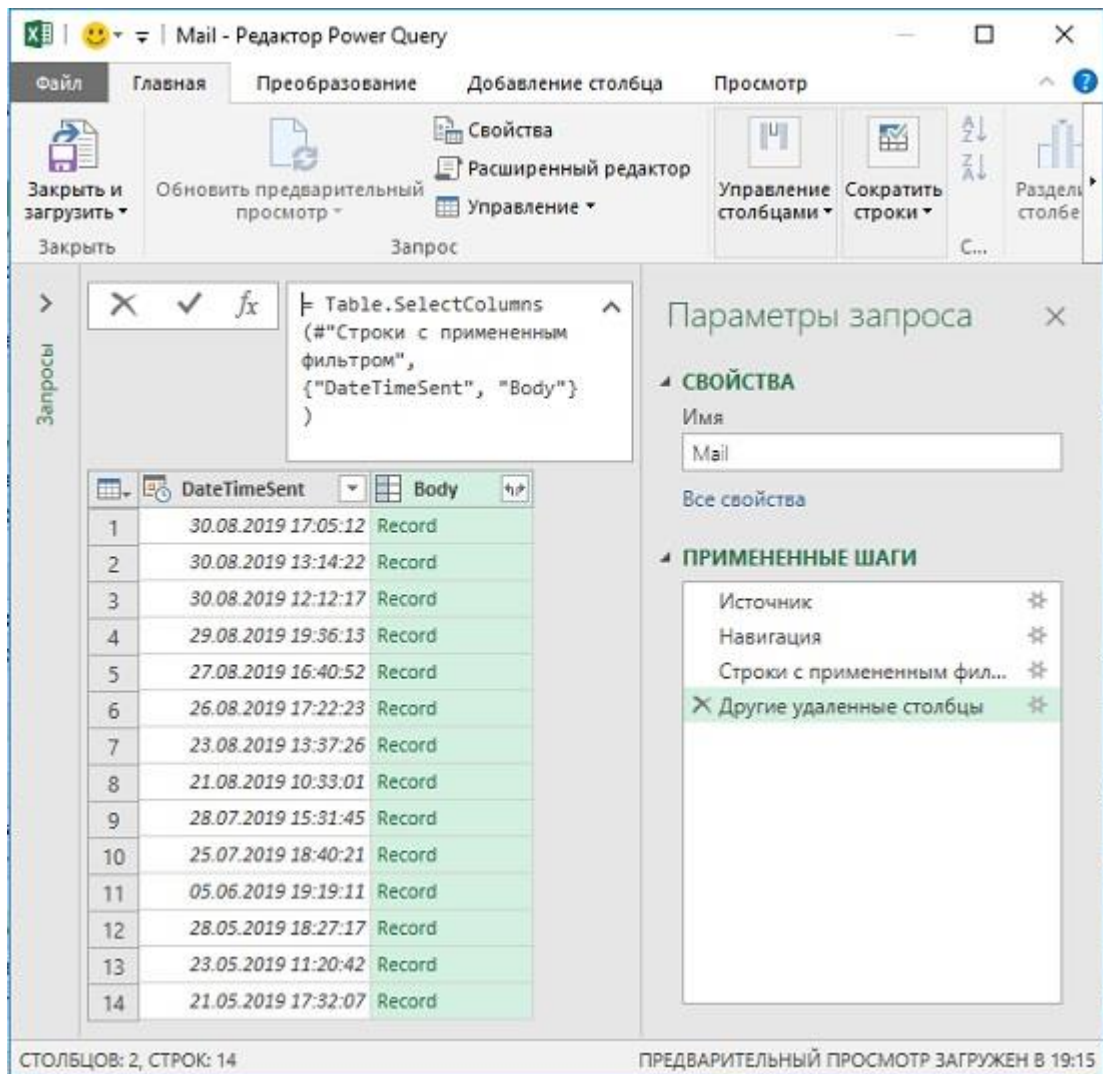


Рис. 13.7. Осталось два столбца

Столбец *Body* содержит полный текст письма, и его можно рассматривать как отдельные ячейки таблицы. Поскольку столбец *Body* содержит стрелку расширения в правом верхнем углу, вы можете развернуть эти записи в столбец. Оказывается, что за текстом скрывается два столбца:

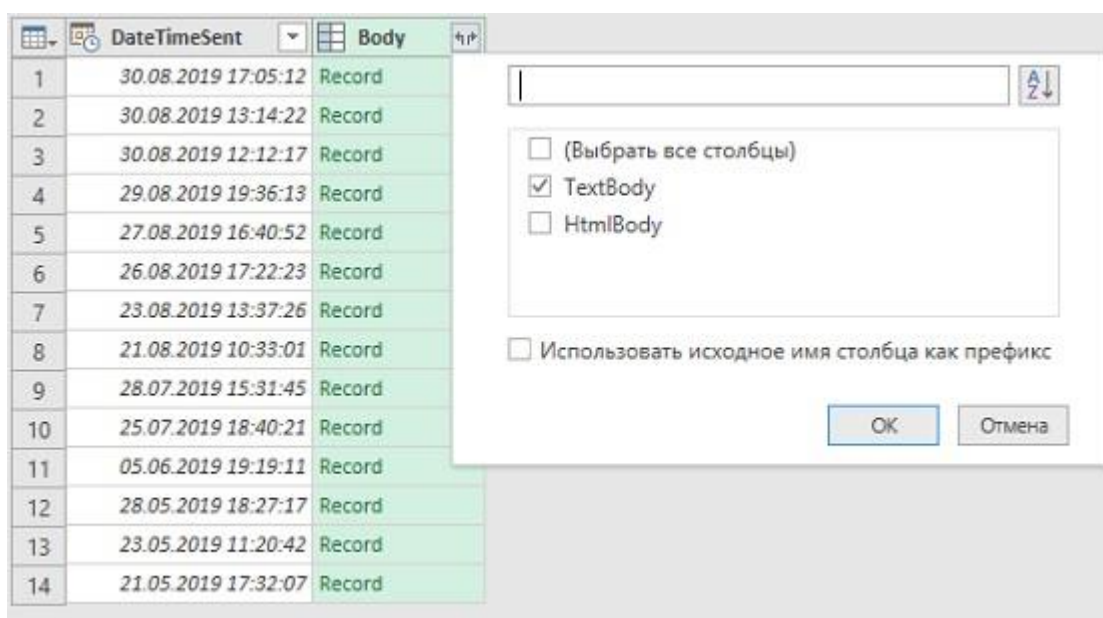


Рис. 13.8. Столбец *Body* может быть развернут в два столбца

Поскольку работа с текстом намного проще, чем с HTML, оставьте только текстовое поле. Обработайте столбец *TextBody*: выделите его и кликните правой кнопкой мыши *Преобразование* – > *Усечь* и еще раз *Преобразование* –> *Очистить*:

	DateTimeSent	TextBody
1	8/3/2015 3:04:35 AM	Hi Ken Puls,This email is to notify you that you have made a sale.You have earned \$4.50 in
2	8/2/2015 12:16:35 PM	Hi Ken Puls,This email is to notify you that you have made a sale.You have earned \$3.00 in
3	7/27/2015 10:42:59 AM	Hi Ken Puls,This email is to notify you that you have made a sale.You have earned \$11.70 in
4	7/4/2015 4:16:50 AM	Hi Ken Puls,This email is to notify you that you have made a sale.You have earned \$4.50 in
5	6/7/2015 8:23:11 AM	Hi Ken Puls,This email is to notify you that you have made a sale.You have earned \$3.00 in
6	6/1/2015 2:49:28 AM	Hi Ken Puls,This email is to notify you that you have made a sale.You have earned \$4.50 in

Рис. 13.9. Обработанные отобранные письма (здесь и далее я использую оригинальные примеры из книги, поскольку лишь они содержат нужную для изложения специфику)

Сейчас содержание писем может быть сокращено, чтобы получить интересующую нас информацию о суммах. Один из замечательных интерфейсов Power Query – окно предварительного просмотра, которое появляется внизу при выборе любой ячейки. Вы можете не только увидеть данные, но и выбрать их в этом окне и скопировать. Вместо того, чтобы разделить столбец *TextBody* по знаку \$ можно выполнить следующие действия:

- Скопируйте текст из окна предварительного просмотра, начиная с Hi и заканчивая \$
- Щелкните правой кнопкой мыши столбец *TextBody* –> *Замена значений...*
- Вставьте скопированную текстовую строку в поле *Значение для поиска*
- Оставьте пустым поле *Заменить на*
- Нажимать ОК

	DateTimeSent	TextBody
1	8/3/2015 3:04:35 AM	4.50 in commission for this sale.This will be added to your total commissions for
2	8/2/2015 12:16:35 PM	3.00 in commission for this sale.This will be added to your total commissions for
3	7/27/2015 10:42:59 AM	11.70 in commission for this sale.This will be added to your total commissions for
4	7/4/2015 4:16:50 AM	4.50 in commission for this sale.This will be added to your total commissions for
5	6/7/2015 8:23:11 AM	3.00 in commission for this sale.This will be added to your total commissions for
6	6/1/2015 2:49:28 AM	4.50 in commission for this sale.This will be added to your total commissions for

Рис. 13.10. Столбец *TextBody* теперь начинается с суммы

Теперь можно разделить столбец *TextBody* по разделителю – пробелу, и удалить правый столбец. Установите тип данных – десятичное число, переименуйте столбец, и вы останетесь со следующими данными:

	Date	Commission
1	8/3/2015	4.5
2	8/2/2015	3
3	7/27/2015	11.7
4	7/4/2015	4.5
5	6/7/2015	3
6	6/1/2015	4.5

Рис. 13.11. Данные, полученные из электронной почты, очищенные и готовые к использованию

Главная –> *Закреть и загрузить*. Постройте сводную диаграмму:

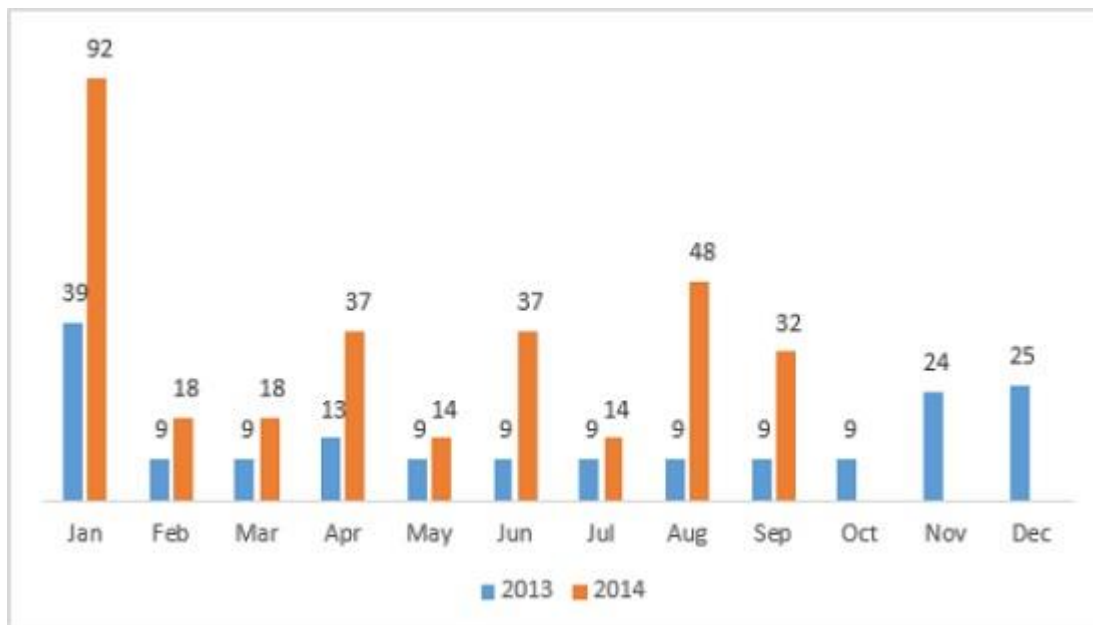


Рис. 13.12. Диаграмма прибыли, построенная на данных текста электронной почты

Рекомендации для источников Exchange

При построении решения на основе службы Microsoft Exchange необходимо учитывать два нюанса: *переносимость* и *производительность*. Планируете ли вы передать решение другому пользователю? Поскольку решение основано на вашей учетной записи электронной почты, использовать его можете только вы. Это решение не переносимо, так как основано на электронных письмах вашего почтового ящика.

Другим важным соображением является производительность. Поскольку почтовые ящики слабо структурированы, и охватывают годы, извлечение данных из Exchange может быть очень медленным. Решение этой проблемы заключается в использовании правил для перенаправления ключевых сообщений электронной почты во вложенные папки. Когда вложенные папки содержат только соответствующие данные, извлечение, преобразование и загрузка готового решения будет происходить гораздо быстрее.