

# КХЕН: высокоскоростная предикативная аналитика

*Краткий обзор решений компании КХЕН, получивших в начале этого года мощный импульс развития за счет новых опций, а также интеграции с Sybase IQ и с решением Sybase in-memory Event Stream Processing, что позволило максимально приблизить аналитику бизнес-процессов к реальному времени.*



Лозовая Екатерина Александровна — директор ЗАО «Терн», к.ф.-м.н.

## Введение

Одной из ключевых задач современного бизнеса становится оперативное принятие правильных бизнес-решений в условиях стремительно возрастающего объема информации и данных. При этом если еще 5 лет назад предикативная аналитика требовалась в основном для принятия тактических и стратегических решений (период предоставления таких отчетов составлял неделю/месяц/год), то в настоящее время она становится неотъемлемой компонентой производственных систем, использующих ее в режиме, максимально приближенном к реальному времени. Для этого совершенствуются методы и алгоритмы прогнозирования, активно развиваются специализированные высокопроизводительные (с мультипоточковой обработкой) устройства для хранения данных (DW), унифицированные OLAP/OLTP-платформы, специализированные решения для работы с неструктурированными массивами данных (нетабличного уровня на базе, например, hadoop-кластеров и распределенных файловых систем), специализированное ПО для углубленной (advanced) аналитики (с погружением обработки на уровень DW) и др.

## КХЕН «подрывает» рынок предикативной аналитики<sup>1)</sup>

КХЕН — это венчурная компания, имеющая 14-летнюю историю развития своих решений. Занимая на рынке небольшую

долю из-за того, что развивает в основном решения только для предикативного анализа, она добилась в этом направлении больших результатов.

Компания КХЕН разработала ряд уникальных приемов с целью максимально автоматизировать процесс анализа данных и сделать его более доступным, с точки зрения необходимой квалификации пользователя. При этом разработчикам КХЕН удалось не только полностью автоматизировать процесс моделирования, но и добиться того, что он практически всегда дает хорошие результаты.

Математический аппарат, который заложен в КХЕН, в ходе анализа строит несколько конкурирующих моделей. Но этот процесс осуществляется не случайным образом (перебором разных методов моделирования), а путем изучения различных наборов моделей с опорой на Теорию минимизации структурного риска В.Вапника (Structured Risk Minimization). Создатели КХЕН разработали механизм сравнения моделей, с тем чтобы добиться наилучшего соотношения между их точностью и надежностью, и уже эту оптимальную модель представить в качестве результата анализа пользователю.

Ключевым моментом процесса прогнозирования является интерпретируемость результатов. Все компоненты КХЕН были сконструированы таким образом, чтобы представлять конечным пользователям содержательные результаты. При этом речь здесь не идет о представлении данных в виде красочных трехмерных диаграмм. Имеется в виду содержательное наполнение, которое отображается затем в виде графиков, например, понятие вклада переменных, важности категорий, индикаторов качества и надежности.

Сами по себе методы описательного и предсказательного анализа бесполезны. Чтобы извлечь прибыль из модели прогнозирования оттока клиентов, необходимо внедрить эту модель в операционную среду компании и на основе прогнозной оценки предпринимать те или иные действия. Поскольку перед использованием моделей их необходимо натренировать (адаптировать к текущей ситуации), дескриптивный анализ и прогнозирование — это всего лишь часть процесса. Недостаточно обеспечить себя инструментами, необходимо обеспечить себя компонентами, которые будут интегрированы

в операционную среду. В результате появляются следующие требования:

- четкий и лаконичный API;
- возможность интеграции в любой пользовательский интерфейс;
- возможность внедрения моделей в операционную среду компании. Выполнение этого требования дает возможность не только производить моделирование в режиме on-line, но и экспортировать построенные модели, используя другие программные языки, например, Java, SQL, PMML и др. Встраивание модели КХЕН в виде программного кода в рабочую базу данных позволяет производить анализ и получать прогнозную оценку в регулярном режиме.

Процесс прогнозирования не заканчивается построением модели. Данные меняются со временем, и необходимо периодически производить мониторинг эффективности моделей с целью принятия решения об их корректировке или выставлении меток в операционной среде. Компания КХЕН включила управление конфигурацией модели в API. Это означает сигнализацию об автоматическом выявлении отклонений на входных распределениях или во взаимосвязях входов-выходов. Очевидно, в последнем случае необходимо использование надежных методов, потому что статистические отклонения в производительности модели не должны являться следствием техники моделирования, но должны идентифицировать различия в данных, которые требуется моделировать. Процедуры, заложенные в КХЕН, будут автоматически выполнять следующие операции:

- выявлять отклонения в операционной среде;
- запускать переобучение моделей;
- использовать эти модели в режиме реального времени или в процессе пакетной обработки;
- строить операционные пользовательские интерфейсы, которые будут использовать все возможности по построению моделей.

В качестве других основных преимуществ решений КХЕН можно отметить такие как:

- широкая поддержка DB-платформ для DW: Teradata, Oracle, Sybase IQ, Vertica,

<sup>1)</sup> Из отчета Gartner: "Vendor Rating: KXEN", 18 July 2012

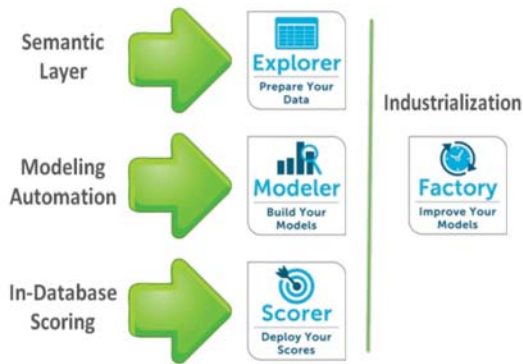


Рис. 1. Семейство решений KXEN.

SQL Server, IBM DB2 и Netezza (углубленная интеграция с рядом этих платформ – Teradata и Sybase IQ – позволяет добиться скорости увеличения обработки запросов в 100 и даже в 1000 раз);

- **простота развертывания.** Платформа KXEN может быть установлена на отдельном стоящем компьютере с удаленным доступом к данным или в составе клиент-серверной конфигурации;
- **высокая производительность.** KXEN написан на языке C++ и поддерживает API в нескольких форматах: C++ загружаемая библиотека, CORBA-сервер на всех поддерживаемых платформах, компонента объектной модели (Component Object Model – COM) библиотеки и распределенная компонента объектной модели (Distributed Component Object Model – DCOM) сервера для Windows;
- **широкая поддержка серверных платформ и ОС:** 32-bit и 64-bit Windows OS, Sun SPARC и Intel, IBM AIX, HP-UX IA-64, Red Hat ES 3 и ES 5.

Gartner в недавно представленном отчете (“Vendor Rating: KXEN”, июль 2012 г.) дала положительную оценку KXEN в следующих категориях: стратегия, маркетинг, технология/методология, ценовая структура и поддержка/управление аккаунтами.

KXEN-платформа InfiniteInsight состоит из трех основных продуктов (рис. 1) для поддержки процессов исследования данных (data mining):

- **Explorer** – исследование, препарирование и подготовка основных данных для анализа;
- **Modeler** – автоматизация создания сложных предикативных моделей, которые включают большое число разнообразных типов данных, несколько тысяч переменных (обычно представленных в типовых бизнес-применениях), а также источники больших данных;
- **Scorer** – помощь в развертывании модели с возможностью экспорта моделей в серию форматов: SQL, SAS, Java, C, Predictive Model Markup Language и др.;

а также трех дополнительных – для поддержки специфических процедур: Social (for network analysis), Genius (marketing optimization), Factory (model management).

KXEN позиционирует свою предикативную аналитику для четырех секторов:

- **телеком.** Примеры решаемых задач: 1) интеллектуальное управление жизненным циклом клиента и повышение его лояльности, 2) разработка тарифо-

в с учетом поведения клиентов, 3) анализ и проработка маркетинговых компаний, 4) анализ и прогнозирование нагрузок на оборудование и линии связи. Общее число инсталляций – более 40 ведущих телеком-провайдеров (свыше 300 млн пользователей), включая Vodafone, Cox, Belgacom, U.S. Cellular и Rogers;

- **финансовый сектор** – более 45 мировых самых больших банков и страховых компаний стандартизовали свою предикативную аналитику на InfiniteInsight®. В этот список входят: Bank of America, Wells Fargo, ING, Discover и Barclays с общим числом управления более 200 млн коммерческих взаимоотношений;
- **электронная торговля (e-business).** Примерами имплементаций при анализе больших данных петабайтного уровня могут служить: Allegro, Allocine, CBS Interactive, HMV, RealNetworks, Shutterfly, Skyrock.com and Tipp24;
- **розничная торговля.** Решаемые задачи: 1) анализ и прогнозирование спроса, 2) оптимизация склада, 3) сопутствующие и перекрестные продажи и др. Примеры использования: Lowe’s, Sears, True Value, Eldorado, Allegro, Darty, Shutterfly, Overstock.com и MonotaRO (общий ежегодный оборот этих компаний составляет более \$500 млрд).

Последнее обновление линейки решений KXEN – InfiniteInsight® версии 6 – стало доступно в марте 2012 г. Основная направленность обновлений – возможность работы с большими данными. Также стал доступен новый продукт InfiniteInsight® Genius, разработанный специально для маркетологов.

Среди основных нововведений следующие опции:

- **Supercommunity Detection (InfiniteInsight® Social)** – возможность обнаружения суперсообществ на основе использования связей между индивидуумами или узлами для идентификации общности между ними. В шестой версии значительно расширены возможности по обнаружению суперсообществ и подобществ на основе подобных интересов и/или поведения, что может быть использовано для усиления результативности маркетинговых акций;
- **Enhanced Mega-Hub Detection (InfiniteInsight® Social)** – возможность детектировать социальные отношения между десятками миллионов индивидуумов, сотнями миллионов связей и миллиардами транзакций. В новой версии расширены возможности по обнаружению мегахабов (мегацентров) с целью устранения чрезвычайно больших и неправомерно связанных узлов, стирающих связи других социальных сообществ;
- **Bipartite Social Graphs (InfiniteInsight® Social)** – установление двусторонних социальных графов, что позволяет связывать индивидуумов, которые купили одни и те же розничные изделия и изделия, которые куплены одним и тем же индивидуумом. Эта особенность дает возможность усиливать коммерческие предложения для посетителей сайтов в электронной коммерции;

– **Export to Microsoft Powerpoint** – пользователи Powerpoint теперь могут экспортировать ценные операционные и итоговые отчеты, включая характеристики моделей и результаты, из InfiniteInsight® Modeler непосредственно в Microsoft Powerpoint;

- **In-Database Score Deviation** – вычисление отклонений результатов скоринга со средневзвешенным значением;
- **введение InfiniteInsight® Genius** – первого решения для предикативной аналитики с дружелюбным интерфейсом (и без использования какого-либо языка программирования) для маркетологов. Оно позволяет строить предикативные модели за минуты на самой последней информации для оптимизации маркетинговых программ. Решение может конфигурироваться с учетом особенностей различных бизнес-процессов на основе преднастроенных шаблонов.

В полной мере все инновации шестой версии InfiniteInsight при анализе больших данных проявляют себя в интеграции с программно-аппаратными решениями, поддерживающими массивно-параллельную обработку. В настоящее время наибольшую интеграцию решения KXEN имеют с платформами Teradata и Sybase IQ (подразделение SAP).

В январе 2012 г. KXEN объявила об интеграции своих решений с Sybase IQ версией 15<sup>2)</sup>. Прежде всего, это возможность внутрибазовой обработки (на уровне аппаратных средств Sybase IQ) внешних программ C++ (от KXEN). Комбинация этих двух решений позволяет получать беспрецедентную скорость и производительность при построении предикативных моделей и анализа социальных сетей. Возможность извлечения неизвестных паттернов из исторических данных для предсказания будущих событий в режиме реального времени открывает широкие возможности для улучшения бизнес-процессов в самых широких областях – от целевого маркетинга и обнаружения мошенничества до персонализированных продуктовых рекомендаций (в розничной торговле и e-коммерции).

Предикативная аналитика требует большого количества данных для построения основных моделей. Ключевым шагом здесь является установление полного обзора бизнес-проблем, включая данные от разнообразных систем типа CRM, web-данных или даже социальных медиа-платформ. Эти исходные данные обычно называют аналитическим набором дан-

2) Главными нововведениями версии 15.0 стали: новая платформа работы с особыми крупными базами данных (VLDB) и модернизированные алгоритмы загрузки данных и обработки запросов, учитывающие особенности новых многоядерных процессорных архитектур и многоярусных систем хранения данных. В версии 15.1 была введена возможность внутрибазовой аналитической обработки – запуск внешних программ на C++ средствами механизма БД. В версии 15.2 были представлены новые функции поиска в текстовых массивах, федерации запросов и поддержки Web 2.0 на уровне программных интерфейсов (API).

Главным новшеством выпуска 15.3 стала кластерная архитектура массово-параллельной обработки PlexIQ. Наконец, в версии 15.4 была введена поддержка нативного API MapReduce, обеспечивающего внутрибазовую аналитическую обработку больших объемов данных. Благодаря этому, а также поддержке языка предикативного моделирования PMML (Predictive Model Markup Language), интеграции с Hadoop, наличию расширенной библиотеки алгоритмов статистических расчетов и информационной проходимости, использующих все возможности технологии PlexIQ®, разработчики могут быстро и надежно программировать аналитическую обработку внутри базы данных, что дает более чем 10-кратный прирост производительности по сравнению с обычным способом (см. SN № 2/50, 2012).

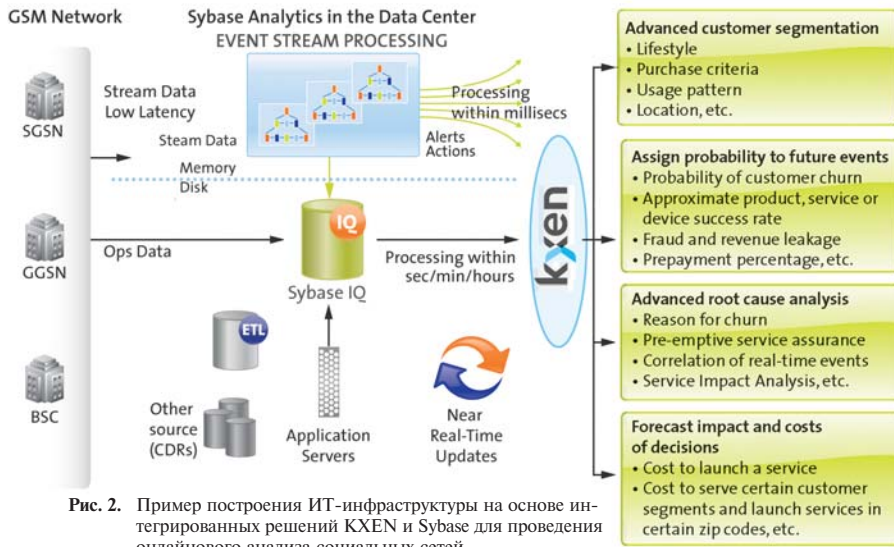


Рис. 2. Пример построения ИТ-инфраструктуры на основе интегрированных решений KXEN и Sybase для проведения онлайн-анализа социальных сетей.

ных, который часто состоит из сотен и тысяч колонок и миллионов строк. И здесь для построения DW нет никаких ограничений. За счет мультипроцессорной обработки на базе интегрированного решения KXEN и Sybase скорость обработки запросов на больших наборах данных (например, с десятками миллионов подписчиков какого-либо телеком-провайдера) удается увеличить в 100 и в 1000 раз в сравнении с традиционными решениями для реляционных БД и более ранними data mining системами.

InfiniteInsight® также жестко интегрирован с решением in-memory Event Stream Processing (ESP) от Sybase, позволяя вместе с вышеназванными особенностями приблизить аналитику бизнес-процессов к реальному времени — интерактивному режиму взаимодействия с клиентом (рис. 2).

## Применение решений KXEN

Из четырех вышеназванных сфер применения предикативной аналитики, телеком и банковский сектор для России являются наиболее актуальными. К названным западным банкам можно добавить и российские: "Промсвязьбанк", "Уральский Банк Реконструкции и Развития", "Севергазбанк", в которых внедрение решений KXEN осуществляла компания "Терн".

На волне общей стагнации экономики банки активно осваивают розничное направление — работу с физическими лицами и малым бизнесом. Количество ежедневно выполняемых банком операций выросло в разы, а современные АБС позволяют по каждой проведенной операции регистрировать и сохранять детальную информацию.

Таким образом, в базах данных банков накапливаются терабайты детальной информации о взаимодействии с клиентом. Эта информация обладает значительным потенциалом для развития бизнеса банка, на основе которой можно отвечать на многие вопросы: каков риск невозврата ссуды заемщиком? какой новый продукт захочет приобрести клиент? какую стратегию выбрать для работы с должником?

В банковской практике KXEN хорошо зарекомендовал себя при решении различных аналитических задач, таких как:

- скоринг кредитования (application scoring);

- адресная работа с клиентами;
- привлечение новых клиентов;
- прогноз состояния выданных ссуд (behavioral scoring);
- прогноз состояния для карточных продуктов;
- работа с плохими долгами (collection scoring);
- выявление мошенничества с пластиковыми картами (fraud detection);
- анализ ключевых показателей деятельности (KPI);
- оптимизация загрузки банкоматов.

### Скоринг кредитования

Кредитный скоринг — процедура балльной оценки кредитоспособности заемщика. Для каждого потенциального заемщика банк рассчитывает рейтинг кредитоспособности — скоринг-балл. При этом используется скоринг-карта — заранее подготовленная процедура, которая присваивает каждой из возможных характеристик заемщика определенный вес.

KXEN позволяет сотрудникам банка строить и обновлять скоринг-карты. Основой для построения моделей выступает кредитная история, накопленная в банке, либо искусственно полученная специальным методом. Используя KXEN для скоринга, банк снижает уровень дефолтов и просрочек по кредитам, упрощает и ускоряет процедуру оформления кредита, увеличивает прибыльность своего розничного портфеля.

### Адресная работа с клиентами

Использование адресной, персонализированной работы с клиентами позволяет банку поднять продажи в розничном сегменте. Целью адресного обращения может быть удержание клиента, продажа дополнительного продукта и другие. В качестве канала обращения банки используют бумажную и электронную почту, смс-сообщения, телефонные звонки и др. По каждому клиенту банка KXEN дает ответы: какое предложение, в какой момент и по какому каналу следует до него донести.

Также при решении указанных задач обычно требуется определить, какие именно клиенты должны быть подвергнуты адресному воздействию. KXEN решает и эту задачу. К примеру, при решении задачи удержания клиентов, KXEN

позволяет определить характеристики клиентов, склонных перейти на обслуживание в другой банк, и в результате выделить "группу риска". С помощью KXEN банк получает максимальный эффект от адресной клиентской работы.

### Привлечение новых клиентов

Налаженная процедура привлечения новых клиентов — одна из ключевых составляющих успешной работы кредитной организации. KXEN дает специалистам возможность получить объективную картину состава клиентской базы банка: ранжирование клиентов по доходности, определение профилей наиболее выгодных из них, сегментирование клиентской базы, и в конечном итоге — создание целевых предложений (продуктов) для разных групп. Используя KXEN, банк повышает эффективность привлечения за счет четкого нацеливания новых предложений на определенные сегменты.

### Прогноз состояния выданных ссуд

Чтобы не допустить возникновения проблем с ликвидностью, банки постоянно контролируют уровень просроченной задолженности. При этом используется информация о текущем состоянии каждой выданной ссуды. Однако простого мониторинга оказывается недостаточно. Если в какой-то момент банк отмечает рост просрочки, то нет возможности немедленно остановить или замедлить рост.

KXEN позволяет спрогнозировать состояние кредитного портфеля на несколько месяцев вперед и принять необходимые меры заранее.

По каждому отдельному клиенту KXEN позволяет получить прогноз вероятности возникновения просрочки или дефолта в ближайший месяц, два, три, а также вероятность досрочного погашения ссуды. При построении моделей используются как анкетные данные, так и информация о "поведении" клиента — время и величина платежей по кредиту, время и величина снятия денег со счета, обращения в сервис-центр и офисы банка.

### Прогноз состояния для карточных продуктов

Прогноз состояния выданных ссуд (behavioral scoring) особенно актуален для карточных программ. В отличие от обычных розничных ссуд, карта выдается однажды и может быть использована для получения множества кредитов. Если банк оценил кредитоспособность заемщика только однажды, при выдаче карты, то он рискует пропустить момент, когда заемщик перейдет в разряд неблагонадежных. Прогноз состояния карточной ссуды в KXEN решает эту проблему.

### Работа с плохими долгами

Классификация плохих долгов по типам дальнейшей работы (списание, суд, в агентство и т.д.), позволяет сократить затраты ручного труда на анализ каждого случая. С помощью моделирования выявляются наиболее значимые факторы, что позволяет с большей вероятностью определять принадлежность долга к определенному типу. В результате увеличивается эффективность работы коллекторской службы банка, снижаются операционные издержки.

### **Выявление мошенничества с пластиковыми картами**

Кредитные и дебетовые карты часто становятся мишенью мошенников. Для снятия средств в свою пользу мошенники используют возможности сети Интернет, специальные технические средства. Эффективное выявление подозрительных транзакций с помощью KXEN позволяет банку решить эту проблему.

Исходными данными при построении моделей выявления мошеннических транзакций служат: демографические данные владельцев карточек, данные расходования и пополнения средств. Модели встраиваются в мониторинговую систему, отслеживающую операции с пластиковыми картами, что позволяет в реальном времени эффективно обнаруживать среди них мошеннические.

Система не только учитывает уже известные мошеннические приемы, но и выявляет ранее неизвестные принципы мошенничества. При появлении подозрительной транзакции сотрудники службы безопасности в течение одной минуты могут связаться с клиентом, принять меры по задержанию мошенника.

### **Анализ ключевых показателей деятельности (KPI)**

Крупнейшие российские банки, стремясь повысить управляемость и эффективность бизнеса, внедряют методологию управления по ключевым показателям (KPI). Как правило, для обслуживания данной методологии в банке развертывается специальная информационная система. Подобные системы позволяют регистрировать, хранить и отслеживать

историю изменения ключевых показателей. Интеграция с KXEN значительно усиливает возможности управления по KPI. KXEN позволяет менеджменту банка оперативно выявлять взаимное влияние различных KPI, определять параметры бизнеса и характеристики бизнес-инициатив, влияющие на наиболее значимые KPI. KXEN увеличивает возврат на инвестиции в систему управления по показателям.

### **Оптимизация загрузки банкоматов**

С расширением сети банкоматов затраты банка на ее обслуживание заметно возрастают. Кроме технического обслуживания, банкоматы требуют регулярной доставки и загрузки наличных денег. Это затратная операция, так как требуется привлечение инкассаторской службы. Однако, имея по каждому банкомату точный прогноз необходимого количества купюр каждого номинала, можно сократить число необходимых загрузок в 1,5-2 раза.

KXEN позволяет наладить процесс так, что для каждого банкомата модель прогноза перестраивается автоматически, каждый день, предсказывая загрузку на следующие несколько дней. Для построения моделей загрузки используются исторические данные снятия наличности, данные о выходных и праздничных днях, прогнозы погоды и др.

### **Только факты**

Результаты недавно опубликованного отчета Aberdeen Group (апрель 2012 г.) свидетельствуют о следующем:

- пользователи KXEN в 5 раз быстрее обновляют свои модели и более чем в 3 раза быстрее строят их, чем другие;

- пользователи KXEN имеют в 2 раза больше шансов в сравнении с другими (69 % против 35 %) почти или полностью автоматизировать процессы подготовки данных;
- 55% пользователей KXEN находят процедуру интеграции моделей в существующие приложения или бизнес-процессы легкой в сравнении с 33% от всей исследованной выборки пользователей.

После внедрения InfiniteInsight® в самом большом банке Австрии – Unicredit – удалось достичь следующих результатов:

- получить дополнительную прибыль в 50 млн евро за один сезон за счет того, что 20% новых проектов банка стало поддерживаться предикативной аналитикой;
- поддерживать минимумом 20 моделей для предикативной аналитики уже через месяц, что прежде могло занять, по крайней мере, четыре месяца.

### **Заключение**

*В июле 2012 г. IDC выпустило первое исследование по региону Россия, посвященное рынку решений BA (business analytics), в соответствии с которым российский рынок BA-решений превысил \$800 млн.*

*В 2012 г. IDC прогнозирует устойчивый рост данного рынка в России. Роль бизнес-аналитики как инструмента корпоративного управления будет возрастать. Среди ключевых факторов, способствующих этому и которые были отмечены в исследовании, названы следующие: экспоненциальный рост объемов информации и необходимость ускорения принятия решений.*

**Компания «Терн» представляет весь спектр бизнес-аналитики**

- Более 300 успешных проектов за 20 лет на ИТ-рынке
- Решения для любых сфер бизнеса
- Использование передовых технологий SAP, IBM, KXEN, PROPHIX
- Консалтинг, обучение и техподдержка сертифицированными специалистами

[www.tern.ru](http://www.tern.ru)