

Сергей Солдатов

Технология Big Data – расширение возможностей АСУ ТП

Традиционные технологии обработки информации уже не справляются с её лавинообразным ростом. На смену им идут специализированные решения и технологии, объединяемые термином “Big Data”. В статье рассказывается о том, что скрывается за данным термином и могут ли быть применены решения Big Data для АСУ ТП.

Технология Big Data

Термин “Big Data” сравнительно молодой, он впервые появился в технической литературе в конце 2000-х годов. С английского Big Data переводится как «большие данные», не частый случай, когда английский термин звучит по-русски так просто. Но Big Data – это не

только большие данные, это совокупность методов, средств и подходов к обработке больших объёмов многообразных данных с большой скоростью прироста. В зарубежной литературе часто все три определяющие характеристики называют «три V»: Volume – объём данных; Velocity – скорость прироста либо

высокоскоростная обработка; Variety – разнообразие типов структурированных и полуструктурированных данных. В настоящее время в связи с активным развитием Интернета вещей, а также средств обработки информации, к характеристикам «три V» добавилась ещё одна – Distributed – распределённая обработка данных.

На практике методы Big Data далеко не всегда идут в связке с большими данными, они могут использоваться и для совсем небольших объёмов информации, причём весьма эффективно.

Источники данных, требующие Big Data

Откуда же берутся все эти данные, которые и большие, и постоянно растут, да ещё и неоднородные и имеют необходимость в распределённой обработке? Достаточно выйти на улицу современного города, и можно сразу увидеть одну из систем, требующих Big Data, – это система «Безопасный город/Умный город» [1, 2]. В подобную систему подключены десятки тысяч камер видеонаблюдения, метеодатчики, датчики загазованности, спутниковой навигации на городском транспорте и экстренных служб. В систему поступают данные от служб коммунального хозяйства, информация по телефонным звонкам жителей города. Очевидно, что

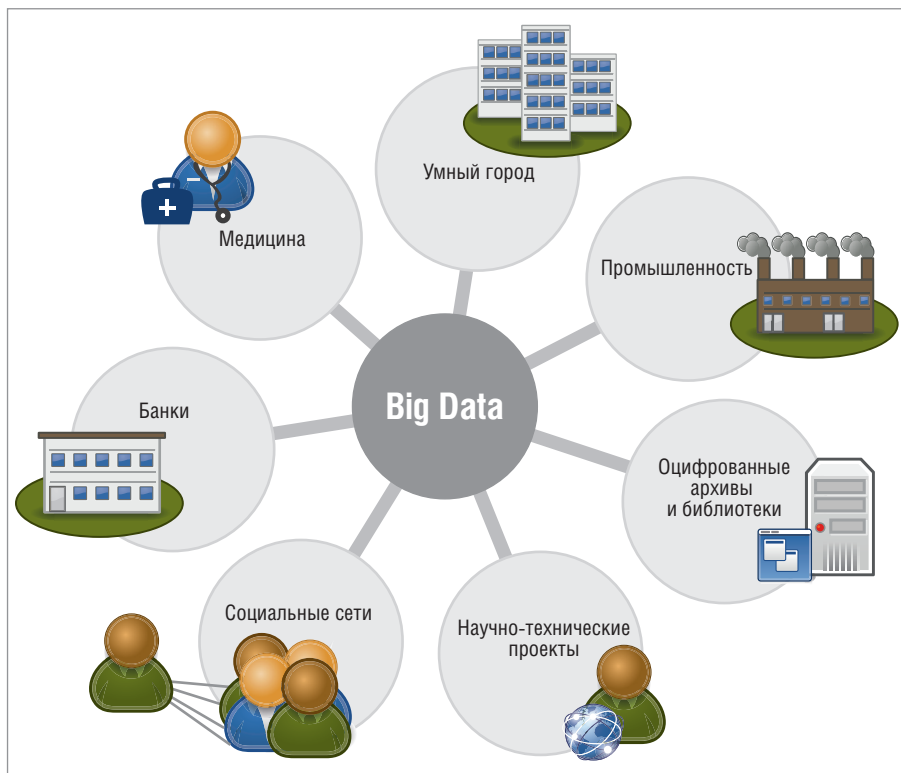


Рис. 1. Примеры областей применения технологий Big Data



Рис. 2. Области применения баз данных SQL и NoSQL

(источник: <https://docs.microsoft.com/ru-ru/azure/documentdb/documentdb-nosql-vs-sql>)

все указанные источники неоднородны, а объёмы информации колоссальны и постоянно растут. Справиться с таким наплывом традиционными методами невозможно.

Другой наглядный пример – банковская сфера. Для многих стало уже привычно оплачивать свои покупки посредством пластиковой карты. Даже в сравнительно небольшом магазине или ресторане установлено несколько касс, оснащённых терминалами для оплаты пластиковыми картами. А по стране таких терминалов сотни тысяч, и ещё есть банкоматы, терминалы самообслуживания и т.д. Каждая операция с картой сопровождается транзакцией в банковской и смежных системах. Например, только за 2015 год в Альфа-Банке было проведено свыше 1 млрд транзакций [3] по пластиковым картам.

Ещё одной областью, требующей средств Big Data, является медицина. Современное медицинское оборудование за один сеанс обследования снимает огромное количество параметров, и хранить результаты требуется пожизненно, а с развитием генетических анализов желательно будет хранить данные и предыдущих поколений [4]. Это упростит выявление генетической предрасположенности к тем или иным заболеваниям. Для медицинских систем крайне важно не только хранить, но и быстро извлекать данные, поскольку от оперативности их получения может зависеть человеческая жизнь.

Источниками, генерирующими большие и разнообразные данные, являются социальные сети, крупные научно-технические установки (например, большой адронный коллайдер), оцифрованные государственные архивы и библио-

теки, а также крупные промышленные предприятия (рис. 1). И это далеко не полный список областей, нуждающихся в технологиях Big Data.

МЕТОДЫ И СРЕДСТВА BIG DATA

Накопленная информация для многих организаций является важным активом, однако мало уметь накапливать, нужно ещё обрабатывать и извлекать из неё пользу. На данный момент существует три основных технологии работы с Big Data: NoSQL, MapReduce и Hadoop.

NoSQL

NoSQL – это ряд подходов для построения баз данных, основанных на отличных от реляционных моделей принципах [5]. Основной причиной появления подобных баз данных стала необходимость параллельной обработки неструктурированных данных с возможностью масштабирования (рис. 2). Базы NoSQL используют не SQL-запросы, а SQL-подобный синтаксис и не структурированы.

Существует много реализаций таких БД, решающих вопрос представления данных по-своему, зачастую весьма специфично. Эти решения допускают не-

ограниченное формирование записей и хранение данных в виде ключа-значения. Наиболее яркими примерами NoSQL являются открытая СУБД MongoDB, Amazon DynamoDB и открытая СУБД OrientDB.

MapReduce

MapReduce – это модель распределённой обработки данных, предложенная компанией Google для обработки больших объёмов данных на компьютерных кластерах (рис. 3). Работа MapReduce состоит из двух основных стадий: Map и Reduce [6, 7].

Стадия Map. На этой стадии происходит предварительная обработка данных. Для этого один из компьютеров (называемый главным узлом – master node) получает входные данные задачи, разделяет их на части и передаёт другим компьютерам (рабочим узлам – worker nodes) для предварительной обработки.

Стадия Reduce. На данной стадии происходит свёртка предварительно обработанных данных. Главный узел получает ответы от рабочих узлов и на их основе формирует результат – решение задачи, которая изначально формулировалась.

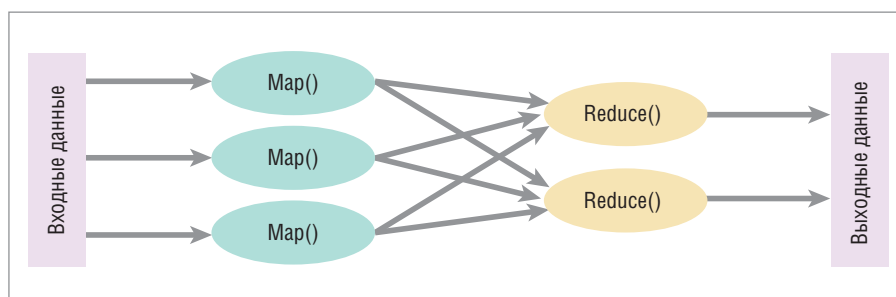


Рис. 3. Обобщённая схема технологии MapReduce

(источник: <https://www.ibm.com/developerworks/cloud/library/cl-openstack-deployhadoop/>)

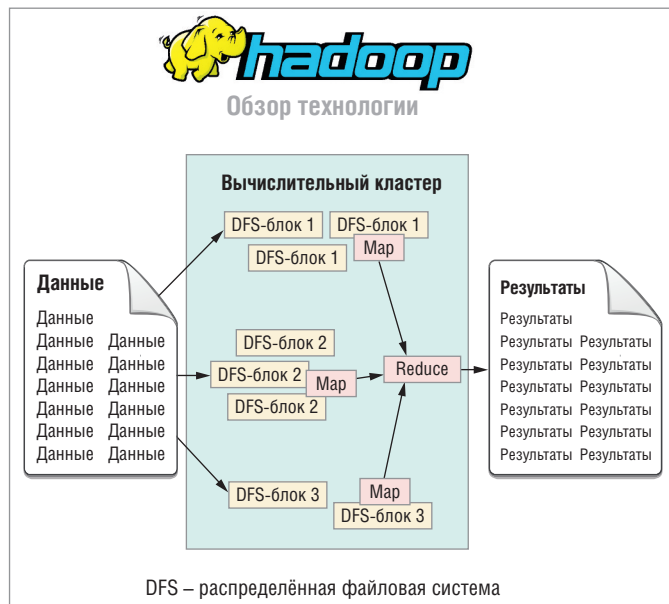


Рис. 4. Обобщённая схема технологии Hadoop

Преимущество модели MapReduce заключается в том, что она позволяет распределённо производить операции предварительной обработки и свёртки. Операции предварительной обработки работают независимо друг от друга и могут производиться параллельно. Аналогично множество рабочих узлов могут осуществлять свёртку – для этого необходимо только, чтобы все результаты предварительной обработки с одним конкретным значением ключа обрабатывались одним рабочим узлом в один момент времени.

Эффективность Map-Reduce по сравнению с последовательными алгоритмами может показаться неочевидной, но именно возможность обработки больших объёмов данных на множестве серверов делает её привлекательной для разработчиков.

Hadoop

Одной из лидирующих технологий, относящихся к классу Big Data, является открытая платформа Hadoop. Данная технология представляет собой открытую платформу программного обеспечения для распределённого хранения и распределённой обработки больших объёмов данных на вычислительных кластерах, построенных из стандартных аппаратных средств.

Hadoop позволяет не только сократить время на обработку и подготовку данных для аналитических систем, но и значительно расширяет возможности по анализу, позволяя оперировать новой информацией, слабо структурированными или совсем не структурированными данными.

Технология также используется для оптимизации существующих процессов обработки данных, что позволяет значительно сократить затраты на хранение и обработку и при этом обеспечить эффективность работы с данными.

Hadoop состоит из двух ключевых компонентов (рис. 4):

- распределённой файловой системы Hadoop (HDFS), которая отвечает за хранение данных на кластере Hadoop;
- системы MapReduce, предназначенной для вычислений и обработки больших объёмов данных на кластере.

Распределённая файловая система Hadoop (HDFS – Hadoop Distributed File System) создана для хранения очень большого объёма информации и обеспечения высокой скорости доступа к этой информации. Файлы хранятся в избыточной форме на нескольких машинах для обеспечения их устойчивости при возможных ошибках и высокой доступности параллельным приложениям.

Hadoop обеспечивает эффективную обработку большого объёма данных путём подключения нескольких компьютеров для параллельной работы. Теоретически суперкомпьютер с 1000 процессоров будет стоить куда больше, чем 1000 компьютеров с одним процессором или 250 четырёхъядерных компьютеров. Hadoop соединит эти небольшие по стоимости машины в один экономически выгодный вычислительный кластер.

Следует обратить внимание на то, что технология даёт возможность осуществлять сложную обработку любых файлов, в том числе неструктурированных, благодаря чему такие данные могут быть эффективно проанализированы и ис-



Рис. 5. Пирамида интеграции технологий автоматизации от уровня техпроцесса до уровня отрасли

пользованы. В частности, это свойство позволяет системам жизнеобеспечения города осуществлять on-line мониторинг информации в социальных сетях и быстро реагировать на сообщения об авариях в сетях водоснабжения, теплоснабжения, газоснабжения и т.п. (информация об авариях в уличных сетях в Интернете появляется раньше, чем звонки на «горячую» линию предприятия).

BIG DATA в АСУ ТП

Оснащение производства современными системами автоматизации приводит к оцифровке всех получаемых данных, и это создаёт у персонала предприятия иллюзию их доступности. Но оцифровано не значит доступно [8].

Данные о технологических процессах есть в АСУ ТП агрегатов, данные о производстве хранятся в MES-системах, данные о заказах содержатся в ERP-системе (рис. 5). Нет только средств, которые могут агрегировать все данные и увязать их между собой.

В случае поиска причин брака и закономерностей, приводящих к его появлению, а также при решении оптимизационных задач необходимо собрать и сопоставить между собой отчёты совершенно разнородных систем. Но горизонт времени зачастую составляет не более нескольких дней, при этом данные не прекращают поступать, постоянно меняя общую картину.

И здесь на помощь могут прийти технологии класса Big Data. Они обеспечат создание единого информационного пространства для всех уровней автоматизации. Особенно эффективны данные технологии при построении авто-

матризованных систем оперативного управления производственно-технологическими процессами больших географически распределённых предприятий городского хозяйства и промышленных предприятий, предъявляющих повышенные требования к масштабируемости, надёжности и безопасности, а также требующих реализации специальных функций пользователя. К таким предприятиям относятся нефте- и газотранспортные сети, химические предприятия, предприятия генерации и распределения электроэнергии.

Технологии Big Data обеспечат высокую производительность в сложных сетевых и резервированных системах, гарантируют получение высокопроизводительных коммуникационных соединений от полевого уровня до станций управления, от отдельного производственного объекта до центра управления. Гарантируется высокая степень доступности и достоверности информации, а также быстрое взаимодействие компонентов системы между собой. Изменение приложений можно будет выполнять без остановки работы системы.

Важно отметить, что технологии Big Data не только обеспечивают сбор дан-

ных с распределённых объектов АСУ ТП, но и позволяют работать с распределёнными локальными базами АСУ ТП. Таким образом, информация о техническом объекте предварительно обрабатывается и хранится непосредственно на нём, а централизация касается только аналитики и задач оптимизации.

Технологии Big Data уже активно внедряются в АСУ ТП. Это системы управления металлургическими предприятиями [8, 9], системы испытаний авиадвигателей [10], системы предсказания аварийных ситуаций для опасных производств [11].

Но ещё больше остаётся АСУ ТП, где есть потребность в средствах Big Data, но они пока не внедрены, далее перечислены некоторые из них.

- **Транспортировка нефти** – это сложная задача, связанная не столько с самой технологией перекачки, сколько с прогнозированием качества нефти (зависит от содержания серы), поступающей на нефтеперерабатывающие предприятия, которое определяется множеством факторов: качество нефти на месторождениях, скорость перекачки по трубопроводной сети, качество и объём нефти в буферных ём-

костях трубопроводов. Чем более сложная трубопроводная сеть, тем сложнее спрогнозировать, какое качество нефти и через какое время будет на том или ином предприятии. А это напрямую влияет на экономику производства.

- **Химическое производство** требует равномерного поступления в заданной пропорции множества компонентов, при этом необходимо на разных этапах обеспечивать технологию процесса, соблюдать требуемое давление, температуру, концентрацию. Количество параметров, определяющих стоимость итогового продукта, достигает нескольких десятков, а сами параметры собираются с десятков территориально распределённых датчиков и установок.
- **Метрополитен** – это самый сложный объект, и выход из строя любой из его подсистем приводит к транспортному коллапсу. Поэтому крайне необходимо не только постоянно мониторить состояние оборудования и обстановки на объектах метро, но и выполнять прогнозирование для своевременного выполнения ремонтных работ, перенаправления потоков пас-

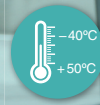


Обновите свои решения для видеонаблюдения с QTS Gateway

IS-453S

Промышленный сетевой накопитель/видеорегистратор с 4 отсеками для жёстких условий эксплуатации или мобильных применений

- Широкий диапазон рабочих температур
- Поддержка кросс-платформенного доступа и различных операционных систем



*Распространяется IEL



QTS-Gateway

Иновационная и интуитивно-понятная операционная система

IVS-300-UTL3

Система для видеонаблюдения на транспорте

- Легко соединяется с устройствами для видеонаблюдения
- Поддерживает облачные решения для видеонаблюдения



IEI Integration Corp.

No. 29, Zhongxing Rd., Xizhi Dist., New Taipei City 221, Taiwan

TEL : +886-2-86916798 / +886-2-26902098 FAX : +886-2-66160028 sales@ieiworld.com www.ieiworld.com



Реклама

сажиров, планирования расписания движения.

ЗАКЛЮЧЕНИЕ

Системы АСУ ТП уже достигли высокого уровня контроля и управления, инженеры могут контролировать производство из любой точки мира и получать детальную информацию о каждом устройстве и установке, но теперь настала пора совершить качественный рывок и включить их в единое информационное пространство предприятия для решения задач анализа и оптимизации производства.

Функциональность Big Data позволяет обеспечить своевременное поступление, индексацию и классификацию данных АСУ ТП, увязку с данными MES и ERP-систем. А это, в свою очередь, позволяет оперативно проанализировать и получить качественно новые знания о процессах на предприятии, которые могут помочь снизить издержки и повысить прибыль. ●

ЛИТЕРАТУРА

1. Панельная сессия «Умный город. Кто может это себе позволить?» [Электронный ресурс] // Сайт ТАСС. – Режим

доступа : <http://tass.ru/pmef-2016/article/3352672>.

2. Данные ударят по газам [Электронный ресурс] // Сайт «Российской газеты». – Режим доступа : <https://rg.ru/2016/11/23/sistema-umnyj-gorod-smozhet-uluchshit-zhizn-naseleniia.html>.
3. Количество транзакций по картам Альфа-Банка превысило миллиард за год [Электронный ресурс] // Сайт Альфа-Банка. – Режим доступа : <https://alfabank.ru/press/news/2016/1/26/35345.html>.
4. Big Data. Большие данные в медицине [Электронный ресурс] // Сайт Medspecial.ru. – Режим доступа : <http://medspecial.ru/news/1/28048/>.
5. SQL и NoSQL: разбираемся в основных моделях баз данных [Электронный ресурс] // Сайт «Типичный программист». – Режим доступа : <https://tproger.ru/translations/sql-nosql-database-models/>.
6. MapReduce [Электронный ресурс] // Сайт Wikipedia. – Режим доступа : <https://ru.wikipedia.org/wiki/MapReduce>.
7. Big Data от А до Я. Часть 1: Принципы работы с большими данными, парадигма MapReduce [Электронный ресурс] // Сайт Хабрахабр. – Режим доступа : <https://habrahabr.ru/company/dca/blog/267361/>.

8. Технология Big Data в металлургии [Электронный ресурс] // Сайт компании «ДАТА-ЦЕНТР Автоматика». – Режим доступа : http://www.dc.ru/activities/asu_tp/big_data.

9. Yandex Data Factory научит металлургов расходовать сырье с Big Data [Электронный ресурс] // Сайт ООО «Роем». – Режим доступа : <https://roem.ru/18-08-2015/203479/mmk-yndx/>.

10. Михаил Зырянов. BIG DATA 2016: Большие Данные для практических задач [Электронный ресурс] // Сайт издательства «Открытые системы». – Режим доступа : <http://www.osp.ru/cw/2016/08/13049514/>.

11. Компанией Data-analytic при поддержке технологий Big Data было разработано типовое решение уровня VI по предсказанию аварийных ситуаций в режиме реального времени АСУ ТП для опасных производств [Электронный ресурс] // Сайт агрегатора новостей по теме исследования искусственного интеллекта. – Режим доступа : http://ai-news.ru/2015/07/kompanij_data_analytic_pri_podderzhki_tehnologij_big_data_bylo_razrabotano_349356.html.

E-mail: ssa-company@rambler.ru

Elektro-Automatik

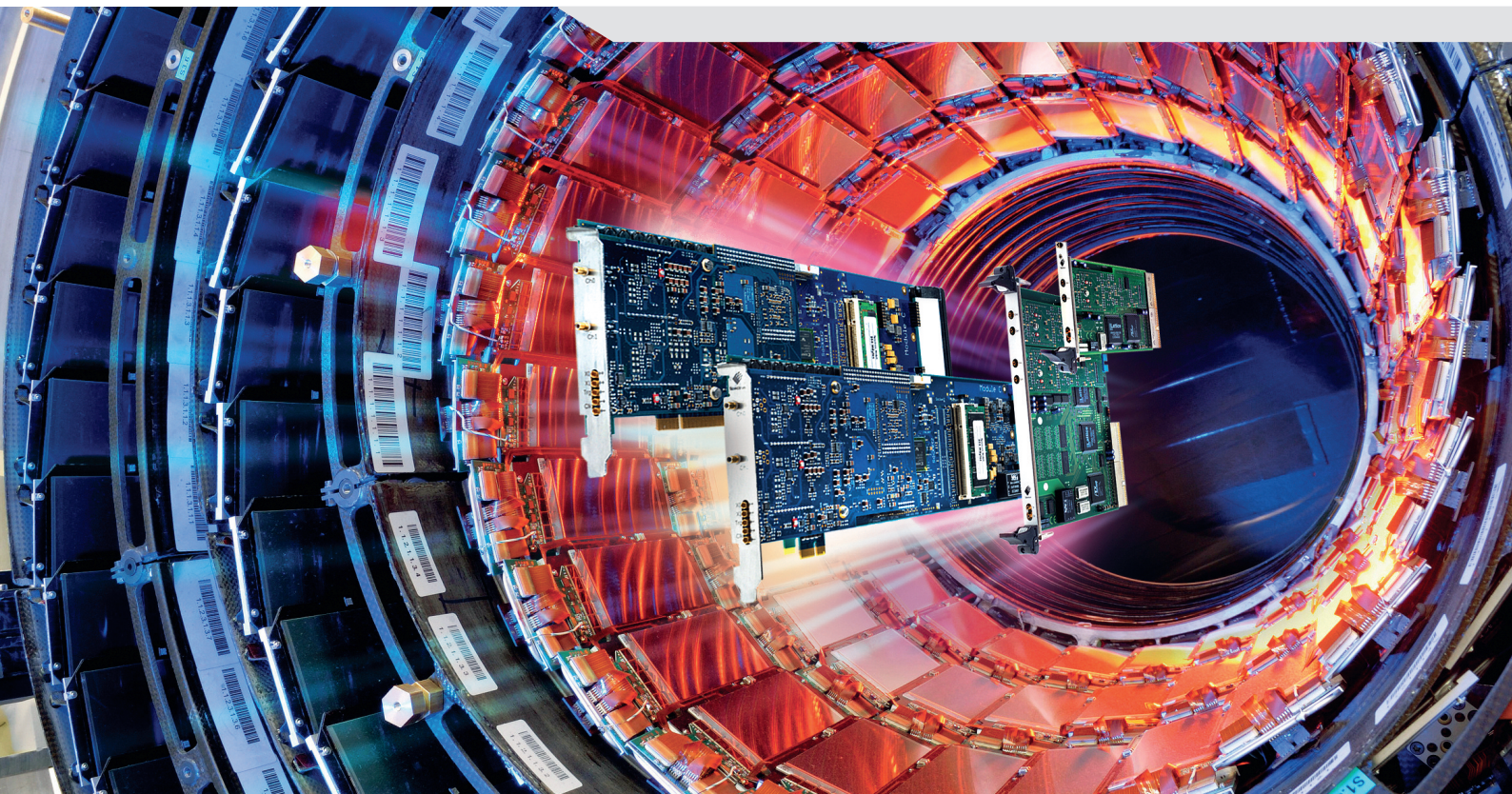
ЭЛЕКТРОННЫЕ ПРОГРАММИРУЕМЫЕ НАГРУЗКИ
Для лабораторий и промышленности

Функции:
Испытания постоянной мощностью, током, напряжением, сопротивлением
Запись результатов испытаний
Рекуперация энергии до 95%

ОФИЦИАЛЬНЫЙ ДИСТРИБЬЮТОР ПРОДУКЦИИ ЭЛЕКТРО-AUTOMATIK

ProSoft®

Тел.: (495) 234-0636 • Факс: (495) 234-0640 • info@prosoft.ru • www.prosoft.ru



Для широкого спектра решений по сбору данных и генерации сигналов

PCI/PCI-X и PCI Express

- Свыше 200 моделей плат
- До 16 синхронных каналов
- Разрешение от 8 до 16 бит
- Частота опроса до 1 ГГц
- Встроенная память до 4 Гбайт
- Тактирование и многомодульная синхронизация

6U CompactPCI

- Около 80 вариантов модулей
- До 16 каналов
- Разрешение до 16 бит
- Частота опроса до 500 МГц

3U PXI

- Более 45 моделей
- Соответствие стандарту PXI
- Межмодульная синхронизация
- Тактирование 10 МГц
- Память до 512 Мбайт

Программное обеспечение



- Собственное ПО SBench 6
- Поддержка ОС Windows, Linux
- Разработка систем сбора и записи данных по ТЗ заказчика
- Индивидуальное консультирование по выбору оборудования для конкретных применений

LXI-системы сбора сигналов



- Более 60 моделей
- Соответствие стандарту LXI
- Число каналов 2–48
- Частота опроса до 500 МГц
- Разрешение от 8 до 16 бит
- Полоса частот от 100 кГц до 250 МГц



ОФИЦИАЛЬНЫЙ ДИСТРИБЬЮТОР ПРОДУКЦИИ SPECTRUM

МОСКВА Тел.: (495) 234-0636 • Факс: (495) 234-0640 • info@prosoft.ru • www.prosoft.ru
С.-ПЕТЕРБУРГ Тел.: (812) 448-0444 • Факс: (812) 448-0339 • info@spb.prosoft.ru • www.prosoft.ru
АЛМА-АТА Тел.: (727) 220-7140/7141 • sales@kz.prosoft.ru • www.prosoft-kz.com
ВОЛГОГРАД Тел.: (8442) 260-048 • volgograd@prosoft.ru • www.prosoft.ru
ЕКАТЕРИНБУРГ Тел.: (343) 376-2820; 356-5111 • Факс: (343) 310-0106 • info@prosoftsystems.ru • www.prosoftsystems.ru
КАЗАНЬ Тел.: (843) 203-6020 • Факс: (843) 203-6020 • info@kzn.prosoft.ru • www.prosoft.ru
КРАСНОДАР Тел.: (861) 224-9513 • Факс: (861) 224-9513 • krasnodar@prosoft.ru • www.prosoft.ru
Н. НОВГОРОД Тел.: (831) 215-4084 • nnovgorod@prosoft.ru • www.prosoft.ru
НОВОСИБИРСК Тел.: (383) 202-0960; 335-7001/7002 • Факс: (383) 230-2729 • info@nsk.prosoft.ru • www.prosoft.ru
ОМСК Тел.: (3812) 286-521 • Факс: (3812) 315-294 • omsk@prosoft.ru • www.prosoft.ru
ПЕНЗА Тел.: (8412) 494-971 • Факс: (8412) 494-971 • penza@prosoft.ru • www.prosoft.ru
САМАРА Тел.: (846) 277-9166 • Факс: (846) 277-9165 • info@samara.prosoft.ru • www.prosoft.ru
УФА Тел.: (347) 292-5216/5217 • Факс: (347) 292-5218 • info@ufa.prosoft.ru • www.prosoft.ru
ЧЕЛЯБИНСК Тел.: (351) 239-9360 • chelyabinsk@prosoft.ru • www.prosoft.ru