



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**  
Федеральное государственное автономное образовательное учреждение высшего образования  
**«Дальневосточный федеральный университет»**

---

**ШКОЛА ЭКОНОМИКИ И МЕНЕДЖМЕНТА**

**Базовая кафедра современного банковского дела**

Тарабанов Максим Александрович

**МЕХАНИЗМ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИИ BIG DATA В  
УПРАВЛЕНИИ БИЗНЕС-ПРОЦЕССАМИ БАНКА: ВОЗМОЖНОСТИ И  
РИСКИ**

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**  
по образовательной программе подготовки  
бакалавров  
по направлению 38.03.01 Экономика  
«Банковское дело»

г. Владивосток

2018



**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное автономное образовательное учреждение  
высшего профессионального образования

**«Дальневосточный федеральный университет»**

---

**ШКОЛА ЭКОНОМИКИ И МЕНЕДЖМЕНТА**

**Базовая кафедра современного банковского дела**

**З А Д А Н И Е**

на выпускную квалификационную работу

*студенту* Тарабанову Максиму Александровичу *группы* Б1401Бдб  
(фамилия, имя, отчество)

на тему «Механизм использования технологии Big Data в управлении бизнес-процессами  
банка: возможности и риски»

Вопросы, подлежащие разработке (исследованию):

- Большие Данные: понятие, сущность и основные характеристики;
- Технологии и методики анализа Больших Данных;
- Большие Данные и управление бизнес-процессами банка;
- Возможности и риски Больших Данных в банковской сфере;
- Проблемы использования Больших Данных в банках;
- Перспективы использования Больших Данных в ПАО «Сбербанк России».

Основные источники информации и прочее, используемые для разработки темы:  
Труды ведущих специалистов, изучающие новые технологии, научные статьи,  
законодательные и нормативные документы, инструктивные материалы по теме  
исследования, внутренние документы кредитной организации, информационный  
портал ПАО «Сбербанк России», правления банка, электронные ресурсы.

Срок представления работы « \_\_\_\_ » \_\_\_\_\_ 2018 г.

Дата выдачи задания « \_\_\_\_ » \_\_\_\_\_ 2018 г.

Руководитель ВКР канд. экон. наук, доцент \_\_\_\_\_ Ю. Б. Шатов  
(должность, уч. звание) (подпись)

Задание получил \_\_\_\_\_ М. А. Тарабанов  
(подпись)

## Оглавление

Введение.....	4
1 Теоретические аспекты технологии Больших Данных.....	8
1.1 Большие Данные: понятие, сущность и основные характеристики.....	8
1.2 Технологии и методики анализа Больших Данных .....	19
1.3 Большие Данные и управление бизнес-процессами банка .....	30
2 Анализ использования Больших Данных в банковской деятельности .....	39
2.1 Возможности и риски Больших Данных в банковской сфере .....	39
2.2 Проблемы использования Больших Данных в банках.....	46
2.3 Перспективы использования Больших Данных в ПАО «Сбербанк России».....	54
Заключение .....	68
Список использованных источников .....	72

## Введение

С приходом новых технологий, инструментов и средств коммуникаций, таких, как социальные сети, количество данных, производимых людьми, растет с каждым годом в геометрической прогрессии. Соотношение коэффициента полезности при этом уменьшается. Следовательно, вся генерируемая информация может быть использована для определенных целей только после предварительной и тщательной обработки.

Термин «Big Data» означает большие работы (коллекции, потоки) данных, которые не могут быть обработаны традиционными компьютерными техниками. Этот термин означает не само понятие «большие данные», а предмет исследования, который включает в себя различные инструменты, техники и платформы.

Большие данные включают в себя информацию, генерируемую различными системами и приложениями. Некоторые из сфер, которые попадают под определение «Big Data»: черный ящик: информационная составляющая часть вертолета, самолета, морского/космического корабля. Данные подобного рода включают в себя запись голосов экипажа (микрофоны и наушники), информацию о характеристиках объекта управления; социальные медиа: включают данные, распространяемые через социальные сети; фондовые биржи: хранение информации о сделках купли-продажи между компаниями-партнерами; энергосистемы: подобного рода данные содержат информацию о узлах и нагрузках энергетической сети; транспортные системы: модели, характеристики, расстояния - все информация о транспорте и дорожных сетях; поисковые системы: инженерный поиск информации различных базах данных.

Как следствие, термин «Big Data» включает большой объем, высокую скорость обработки и широкое разнообразие данных и делится на три типа: структурные данные - реляционные БД; полуструктурированные данные - XML-файлы; неструктурированные данные - файлы формата Word, PDF, Text, медиа-журналы.

Большие данные действительно имеют решающее значение для нашей жизни и становятся одной из самых важных технологий в современном мире. Самыми распространенными и известными являются лишь несколько преимуществ. Например, использование информации, хранящейся в социальных сетях, маркетинговые агентства изучают обратную связь на свои кампании, акции, и другие рекламные носители.

В свою очередь, использование информации в социальных медиа-системах, таких как предпочтения и восприятие продукта потребителями, компании и розничные организации планируют свое производство. Касательно такой сферы, как медицина, применимость данных о предыдущей истории болезни пациентов способствует обеспечению лучшего и более быстрого обслуживания.

Большие технологии передачи данных играют важную роль в обеспечении детального анализа, который способствует принятию более точных решений, что в свою очередь приводит к повышению эффективности эксплуатации, снижению затрат и снижению рисков для бизнеса.

Для использования возможностей больших данных требуется инфраструктура, которая может управлять и обрабатывать огромные объемы структурированных и неструктурированных данных в реальном времени и может защитить конфиденциальность и безопасность данных. Существуют различные технологии на рынке от различных поставщиков, включая такие компании, как Google, IBM, Microsoft, SAP и др.

Большие Данные получили широкое распространение во многих отраслях бизнеса. Их используют в здравоохранении, телекоммуникациях, торговле, логистике, в финансовых компаниях, а также в государственном управлении.

Большие Данные дают возможность проанализировать кредитоспособность заемщика, также они полезны для кредитного скоринга и андеррайтинга. Внедрение технологий Больших Данных позволит сократить время рассмотрения кредитных заявок. С помощью Больших Данных можно

проанализировать операции конкретного клиента и предложить подходящие именно ему банковские услуги.

В настоящее время предприятиям приходится работать с большими объемами информации, которая часто обновляется и приходит из разных источников. С помощью технологий Big Data предприятия могут анализировать огромные массивы данных и выявлять полезные закономерности, дающие им конкурентные преимущества. Для более легкого восприятия и быстрого принятия управленческих решений необходимо представить результаты анализа данных визуально.

На данный момент есть несколько видов представления массивов данных. Но существующие методы визуализации еще недостаточно развиты и требуют усовершенствования. В связи с вышеизложенным, механизм использования технологии Big Data в управлении бизнес-процессами является актуальным в настоящее время.

Целью выпускной квалификационной работы является изучение механизма использования технологии Big Data в управлении бизнес-процессами банка: возможности и риски. Для достижения намеченной цели в работе поставлены следующие задачи:

- изложить теоретические аспекты технологии Больших Данных (большие данные: понятие, сущность и основные характеристики; технологии и методики анализа Больших Данных; большие данные и управление бизнес-процессами банка);

- провести анализ использования Больших Данных в банковской деятельности (возможности и риски Больших Данных в банках; проблемы использования Больших Данных в банковской деятельности; перспективы использования Больших Данных в банках).

Предмет исследования - механизм использования технологии Big Data в управлении бизнес-процессами.

Объект исследования технологии Big Data в управлении бизнес-процессами банков.

Исследования проведены с использованием таких общенаучных методов, как исторический, аналитический, монографический, системно–структурный анализ и синтез и конкретных прикладных методов, как экономико–статистический, графический и других.

Теоретической и методологической основой работы явилась учебная и периодическая литература, а также труды таких ученых как: Бачило И.Л., Биктимирова М.Р., Блинова Н., Бодрова А.А., Будзко В. И., Вахрамеева К., Веретенникова А. В., Волковой Ю. С. И т.д.

Исследованию Больших Данных посвящены труды как зарубежных, так и российских ученых: James Manyika, Michael Chui, Бодров А.А., Топорков В.В., Будзко В.И. и т.д.

Существенный вклад в изучение этой технологии вносят крупные мировые компании, такие как: McKinsey & Company, CNews Analytics, SAP, Oracle, IBM, Microsoft, Teradata и многие другие. Они занимаются обработкой и анализом данных и на основе Больших Данных создают программно-аппаратные комплексы.

Выпускная квалификационная работа включает введение, две главы, заключение, список использованных источников, а также приложения.

В первой главе выпускной квалификационной работе изложены теоретические аспекты технологии Больших Данных: дано понятие, сущность и основные характеристики Больших Данных; охарактеризованы технологии и методики анализа Больших Данных; большие данные и управление бизнес-процессами банка.

Во второй главе выпускной квалификационной работе проведен анализ использования Больших Данных в банковской деятельности: возможности и риски Больших Данных в банках; проблемы использования больших данных в банковской деятельности; перспективы использования Больших Данных в банках.

# **1 Теоретические аспекты технологии Больших Данных**

## **1.1 Большие Данные: понятие, сущность и основные характеристики**

Как известно «Большие Данные» (Big Data) сегодня являются одним из ключевых драйверов (стимулов) развития информационно-коммуникационных технологий (ИКТ). Это направление развития ИКТ, относительно новое для России, получило широкое распространение в западных странах. Связано это с тем, что в эпоху информационных технологий, особенно после бума социальных сетей, стало накапливаться значительное и все возрастающее количество информации, связанной с каждым пользователем интернета, что в конечном счете дало развитие направлению Больших Данных [3].

Большие Данные - это не отдельная технология, а скорее сочетание опробованных и вновь внедряемых технологий, которое позволяет компаниям извлекать из имеющихся данных информацию для использования в их деятельности.

Большие Данные можно определить, как способность управлять большими объемами разнородных данных со скоростью, достаточной для анализа таких данных в реальном времени и своевременного реагирования. Большие Данные обладают тремя основными признаками:

- объем: количество данных;
- скорость: быстрота обработки данных;
- вариативность: количество различных типов данных.

Термин «Большие Данные» вызывает множество споров. Многие полагают, что этот термин характеризует лишь объем накопленной информации, но не стоит забывать о технической стороне: названное направление развития ИКТ включает в себя технологии хранения и вычисления, а также сервисные услуги.

Следует отметить, что к названной сфере относится обработка именно большого объема информации, который затруднительно обрабатывать



традиционными способами. Например, международная исследовательская компания Forrester определяет это понятие как технологию в области аппаратного и программного обеспечения, которая объединяет, организует, управляет и анализирует данные, характеризующиеся «четырьмя V»: объемом (Volume), разнообразием (Variety), изменчивостью (Variability) и скоростью (Velocity):

- volume – это очень большой объем информации, накопленный в базах данных, его трудоемко обрабатывать и хранить традиционными средствами СУБД; поэтому востребованы новые подходы и усовершенствованные инструменты обработки этих данных;

- variety – это разнообразие (многообразие) форматов данных (главный критерий Больших Данных): большие массивы данных, поступающие из разных источников в различных форматах, разной степени структурированности – табличные данные в СУБД, иерархические данные, текстовые документы, видео, изображения, аудиофайлы и т. д.; поэтому востребована возможность одновременной обработки структурированной и неструктурированной разноформатной информации [33].

Главное отличие структурированной информации – в том, что она может быть классифицирована. Примером может служить информация о клиентских транзакциях. Неструктурированная информация включает в себя видео-, аудиофайлы, свободный текст, информацию, поступающую из социальных сетей. На сегодняшний день 80% появляющейся информации входит в группу неструктурированной. Такая информация нуждается в комплексном анализе, чтобы сделать ее полезной для дальнейшей обработки:

- variability – изменчивость информации: например, таковой является информация, непрерывно поступающая с датчиков некоторых устройств или из интернета и имеющая важное значение для анализа, прогнозирования и принятия решений;

- velocity – скорость накопления и обработки данных; данный признак указывает как на увеличивающуюся скорость накопления данных (90%

информации было собрано за последние 2 года), так и на скорость их обработки; в последнее время в ряде задач стали более востребованы технологии обработки данных в реальном времени. К перечисленным свойствам Больших Данных сегодня добавляют:

– veracity – достоверность данных: все большую значимость пользователи стали придавать достоверности имеющихся данных. Так, у интернет-компаний есть проблема по разделению действий, проводимых на сайте компании роботом и человеком, что приводит, в конечном счете, к затруднению анализа данных;

– value – ценность накопленной информации: Большие Данные должны быть полезны компаниям и приносить им определенную выгоду, например, должны помогать в усовершенствовании бизнес-процессов, составлении отчетности или оптимизации расходов. Таким образом, при соблюдении указанных выше условий накопленные объемы данных можно относить к числу больших [50].

Цикл управления Большими Данными представлен рисунком 1.



Источник: [13]

Рисунок 1 - Цикл управления Большими Данными

Из рисунка 1 видно, что первым этапом цикла является сбор данных, затем следует их систематизация и обобщение. После этого данные могут быть проанализированы с учетом конкретной задачи. Особенно важен вопрос

проверки данных. При объединении данных из разных источников нужно убедиться, что они друг с другом соотносятся. Кроме того, некоторые источники данных могут содержать конфиденциальную информацию, для которой необходимо обеспечить соответствующий уровень защиты и управления.

Большие данные - это огромные массивы данных, которые на порядок больше (объем); разнообразнее, включая в себя структурированные, полуструктурированные и неструктурированные данные, (диверсификация) и быстрее (скорость передачи и обработки), чем все данные, с которыми организации до сих пор приходилось иметь дело. Этот поток данных создается подключенными устройствами - от ПК и смартфонов до датчиков, например, устройств считывания RFID и уличных камер. Кроме того, эти данные гетерогенны и передаются в различных форматах: текст, документы, изображения, видео и многое другое [51].

Большие данные (англ. Big Data) - серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста информации [26].

Согласно отчету McKinsey Institute, «Большие данные: новый рубеж для инноваций, конкуренции и производительности» (Big data: The next frontier for innovation, competition and productivity), термин «Большие Данные» относится к наборам данных, размер которых превосходит возможности типичных баз данных (БД) по занесению, хранению, управлению и анализу информации [2].

«Большие Данные» предполагают нечто большее, чем просто анализ огромных объемов информации. Проблема не в том, что организации создают огромные объемы данных, а в том, что большая их часть представлена в формате, плохо соответствующем традиционному структурированному формату БД - это веб-журналы, видеозаписи, текстовые документы, машинный код или, например, геопространственные данные. Всё это хранится во

множестве разнообразных хранилищ, иногда даже за пределами организации. В результате, корпорации могут иметь доступ к огромному объему своих данных и не иметь необходимых инструментов, чтобы установить взаимосвязи между этими данными и сделать на их основе значимые выводы.

Кроме того, данные сейчас обновляются все чаще и чаще, поэтому традиционные методы анализа информации не дают необходимой скорости и качества обработки огромных объемов постоянно обновляемых данных, что в итоге и открывает дорогу технологиям больших данных.

Понятие больших данных подразумевает работу с информацией огромного объема и разнообразного состава, весьма часто обновляемой и находящейся в разных источниках в целях увеличения эффективности работы, создания новых продуктов и повышения конкурентоспособности.

Бизнес-анализ является описательным процессом анализа результатов, достигнутых бизнесом в определенный период времени, между тем как скорость обработки больших данных позволяет сделать анализ предсказательным, способным предлагать бизнесу рекомендации на будущее.

Технологии Больших Данных позволяют также анализировать больше типов данных в сравнении с инструментами бизнес-аналитики, что дает возможность фокусироваться не только на структурированных хранилищах.

Большие Данные и бизнес-аналитика имеют одинаковую цель (поиск ответов на вопрос), и отличаются друг от друга по трем аспектам.

1. Большие Данные предназначены для обработки более значительных объемов информации, чем бизнес-аналитика, и это, конечно, соответствует традиционному определению Больших Данных.

2. Большие Данные предназначены для обработки более быстро получаемых и меняющихся сведений, что означает глубокое исследование и интерактивность. В некоторых случаях результаты формируются быстрее, чем загружается веб-страница.

3. Большие Данные предназначены для обработки неструктурированных данных, способ использования которых мы только начинаем изучать после

того, как смогли наладить их сбор и хранение, требуются алгоритмы и возможность диалога для облегчения поиска тенденций, содержащихся внутри этих массивов [33].

Сегодня компании должны обрабатывать колоссальное количество данных в объемах, которые трудно представить, это приводит к тому, что традиционные базы данных не могут справиться с такой задачей, и это приводит к необходимости внедрять технологии Больших Данных. В таблице 1 представлена сравнительная характеристика Больших Данных и традиционных баз данных.

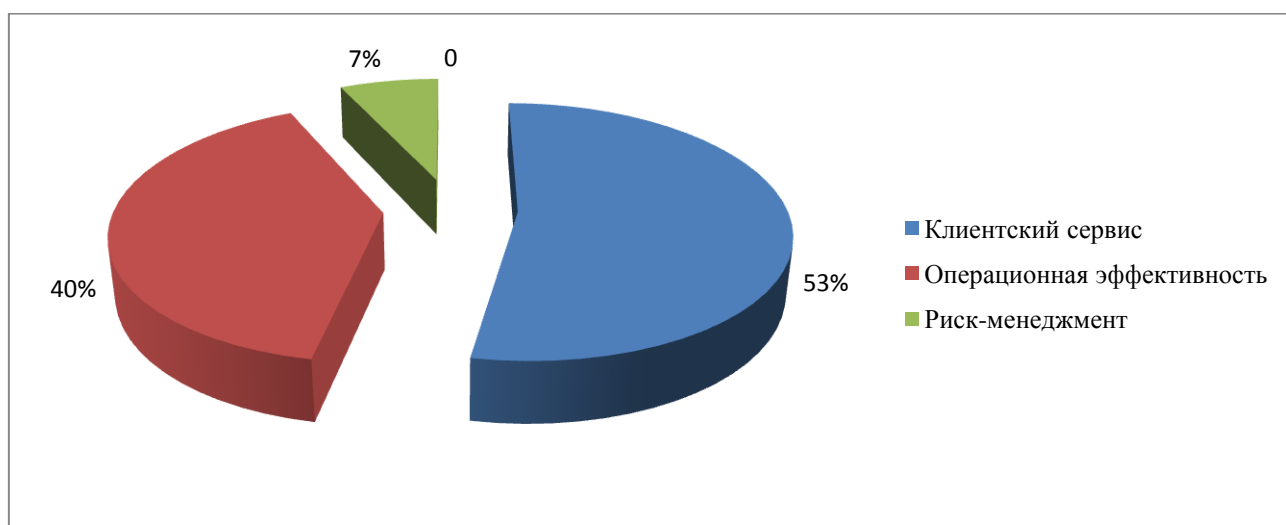
Таблица 1 - Сравнительная таблица традиционной и базы Больших Данных

Характеристика	Традиционная база данных	База Больших Данных
Объем информации	От гигабайт ( $10^9$ байт) до терабайт ( $10^{12}$ байт)	От петабайт ( $10^{15}$ байт) до эксабайт ( $10^{18}$ байт)
Способ хранения	Централизованный	Децентрализованный
Структурированность данных	Структурирована	Полуструктурирована и неструктурирована
Модель хранения и обработки данных	Вертикальная модель	Горизонтальная модель
Взаимосвязь данных	Сильная	Слабая

Источник: [26]

Сфера использования технологий Больших Данных весьма обширна.

Сферы применения Больших Данных представлены рисунком 2.



Источник: [26]

Рисунок 2 - Сферы применения Больших Данных

Большинство компаний использует Большие Данные в сфере клиентского сервиса, второе по популярности направление – операционная эффективность, в сфере управления рисками Большие Данные на текущий момент времени менее распространены.

Следует отметить, что Большие Данные являются одной из самых быстрорастущих сфер информационных технологий: согласно статистике общий объем получаемых и хранимых данных удваивается каждые 1,2 года. За период с 2015 по 2017 годы количество данных, ежемесячно передаваемых мобильными сетями, выросло на 81%.

По оценкам Cisco, в 2017 году объем мобильного трафика составил 2,5 эксабайта (единица измерения количества информации, равная 1018 стандартным байтам) в месяц, а уже в 2019 году он будет равен 24,3 эксабайтам. Таким образом, Большие Данные – это уже устоявшаяся сфера технологий, несмотря на относительно молодой возраст, получившая распространение во многих сферах бизнеса и играющая немаловажную роль в развитии компаний [30].

Большие Данные получили широкое распространение во многих отраслях бизнеса – их используют в здравоохранении, телекоммуникациях, торговле, логистике, финансовых компаниях, а также в государственном управлении.

Розничная торговля. В базах данных розничных магазинов может быть накоплено множество информации о клиентах, системе управления запасами и поставками товарной продукции. Эта информация может быть полезна во всех сферах деятельности магазинов. Так, например, с помощью накопленной информации можно управлять поставками товара, его хранением и продажей, а также прогнозировать спрос и поставки товара. Система обработки и анализа данных может решить и другие проблемы ритейлера, например, оптимизировать затраты или подготавливать отчетность.

Финансовые услуги. Большие Данные дают возможность проанализировать кредитоспособность заемщика, также они полезны для кредитного скоринга и андеррайтинга. Внедрение технологий Больших Данных

позволяет сократить время рассмотрения кредитных заявок, а также проанализировать операции конкретного клиента и предложить банковские услуги, подходящие именно ему.

В телекоммуникационной отрасли широкое распространение Большие Данные получили у сотовых операторов. Операторы сотовой связи наравне с финансовыми организациями имеют одни из самых объемных баз данных, что позволяет им проводить наиболее глубокий анализ накопленной информации. Главной целью анализа данных являются удержание существующих клиентов и привлечение новых. Для этого компании проводят сегментацию клиентов, анализируют их трафики, определяют социальную принадлежность абонентов. Помимо использования Больших Данных в маркетинговых целях, эти технологии применяются для предотвращения мошеннических финансовых операций.

Горнодобывающая и нефтяная промышленности. Большие Данные используются как при добыче полезных ископаемых, так и при их переработке и сбыте. На основании поступившей информации предприятия могут делать выводы об эффективности разработки месторождения, отслеживать график капитального ремонта и состояния оборудования, прогнозировать спрос на продукцию и цены.

Наибольшее распространение Большие Данные получили в телекоммуникационной отрасли, а также в инжиниринге, ИТ, в финансовых и государственных предприятиях. Менее популярны Большие Данные в образовании и здравоохранении.

На сегодняшний день Большие Данные активно внедряются в зарубежных компаниях. Такие компании, как Nasdaq, Facebook, Google, IBM, VISA, Master Card, Bank of America, HSBC, AT&T, Coca Cola, Starbucks и Netflix, уже активно используют ресурсы Больших Данных.

Сферы применения обработанной информации разнообразны и варьируются в зависимости от отрасли и задач, которые необходимо решить.

Компания HSBC использует технологии Больших Данных для

противодействия мошенническим операциям с пластиковыми картами. С помощью этих технологий компания увеличила эффективность службы безопасности в 3 раза, распознавание мошеннических инцидентов – в 10 раз. Экономический эффект от внедрения таких технологий превысил 10 млн. долл. США.

Система Антифрод VISA позволяет в автоматическом режиме вычислить операции мошеннического характера, в настоящее время она позволяет предотвратить мошеннические платежи на сумму до 2 млрд. долл. США ежегодно.

Суперкомпьютер Watson компании IBM анализирует в реальном времени поток данных по денежным транзакциям. По данным IBM, Watson на 15% увеличил количество выявленных мошеннических операций, на 50% сократил ложные срабатывания системы и на 60% увеличил сумму денежных средств, защищенных от транзакций такого характера [30].

Компания Procter & Gamble (P&G) с помощью Больших Данных проектирует новые продукты и составляет глобальные маркетинговые акции.

P&G создала специализированные офисы Business Spheres, где можно просматривать информацию в реальном времени. Таким образом, у менеджмента компании появилась возможность мгновенно проверять гипотезы и проводить эксперименты. В P&G считают, что Большие Данные существенно помогают в прогнозировании деятельности компании.

Ритейлер офисных принадлежностей OfficeMax с помощью технологий Больших Данных анализируют поведение клиентов.

Анализ Больших Данных позволил увеличить B2B выручку на 13%, уменьшить затраты на 400 тыс. долларов США в год [30].

По мнению компании Caterpillar, ее дистрибьюторы ежегодно упускают от 9 до 18 млрд. долл. США прибыли только из-за того, что не внедряют технологии обработки Больших Данных. Последние позволили бы клиентам более эффективно управлять парком автомобилей за счет анализа информации, поступающей с датчиков, установленных на них. На сегодняшний день уже



есть возможность анализировать состояние ключевых узлов, их степени износа, управлять затратами на топливо и техническое обслуживание.

Компания Luxottica group является производителем спортивных очков таких марок, как Ray-Ban, Persol и Oakley. Она применяет технологии Больших Данных для анализа поведения потенциальных клиентов и «умного» смс-маркетинга. В результате Luxottica group выделила более 100 миллионов наиболее ценных клиентов и повысила эффективность маркетинговой кампании на 10%.

С помощью Yandex Data Factory разработчики игры World of Tanks анализируют поведение игроков. Технологии Больших Данных позволили проанализировать поведение 100 тысяч игроков World of Tanks с использованием более 100 параметров (информация о покупках, играх, опыт и др.). В результате анализа был получен прогноз оттока пользователей. Данная информация позволяет уменьшить уход пользователей и работать с участниками игры адресно. Разработанная модель оказалась на 20–30% эффективнее стандартных инструментов анализа игровой индустрии.

Министерство труда Германии использует Большие Данные в работе, связанной с анализом поступающих заявок на выдачу пособий по безработице. Так, проведенный анализ информации показал, что 20% пособий выплачивалось незаслуженно. В результате министерство сократило расходы на 10 млрд. евро.

По мнению IDG Enterprise, в 2017 расходы компаний на сферу Больших Данных составили в среднем 7,4 млн. долл. США на компанию (у крупных компаний это примерно 13,8 млн. долл. США, у малых и средних – 1,6 млн. долл. США). Больше всего средств инвестировано в такие области, как анализ и визуализация данных и их сбор. Согласно имеющейся информации, инвестиции в 2017 году были использованы на улучшение качества данных, совершенствование планирования и прогнозирования, а также увеличение скорости обработки данных [30].

Компаниями финансового сектора, по данным Bain Company's Insights

Analysis, были произведены значительные инвестиции (в 2017 году потрачено около 6,4 млрд. долл. США на технологии Больших Данных, средний темп роста инвестиций составит 22% до 2020 года; интернет-компании потратили по-рядка 2,8 млрд. долл. США, средний темп роста увеличения затрат на Большие Данные составит 26%).

Согласно результатам исследования CNews Analytics и Oracle, уровень зрелости российского рынка Big Data за последний год повысился. Респонденты, представляющие 108 крупных предприятий из разных отраслей, продемонстрировали более высокую степень осведомленности об этих технологиях, а также сложившееся понимание потенциала подобных решений для своего бизнеса.

В России накоплено 155 эксабайт информации, что составляет всего лишь 1,8% мировых данных. Объем информации к 2020 году достигнет 980 эксабайт и займет 2,2%. Таким образом, средний темп роста объема информации составит 36% в год.

Компания IDC оценивает рынок России в 340 млн долл. США, из них 100 млн долл. США – решения SAP, примерно 240 млн долл. США – аналогичные решения Oracle, IBM, SAS, Microsoft и др. Темп роста российского рынка Больших Данных составляет не менее, чем 50% в год [30].

Прогнозируется сохранение позитивной динамики в этом секторе российского рынка ИТ, даже в условиях общей стагнации экономики. Это связано с тем, что бизнес по-прежнему предъявляет спрос на решения, позволяющие повысить эффективность работы, а также оптимизацию расходов, улучшение точности прогнозирования и минимизировать возможные риски компании.

Таким образом, Большие Данные (англ. Big Data) - серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста информации.

Это огромные массивы данных, которые на порядок больше (объем); разнообразнее, включая в себя структурированные, полуструктурированные и неструктурированные данные, (диверсификация) и быстрее (скорость передачи и обработки). Этот поток данных создается подключенными устройствами - от ПК и смартфонов до датчиков, например, устройств считывания RFID и уличных камер. Кроме того, эти данные гетерогенны и передаются в различных форматах: текст, документы, изображения, видео и многое другое.

Управление данными и их анализ дают любой организации, независимо от ее размера и области деятельности, значительные преимущества и открывают большие перспективы. Однако с развитием экономики в целом и каждой отдельной компании сбор информации о потребителях, продуктах и услугах становится сложнее. Когда речь идет о небольшом количестве клиентов, которые приобретают один и тот же продукт одним и тем же путем, вести учет таких сведений не трудно. Но со временем растут и компании, и их рынки сбыта; конкуренция заставляет создавать новые линейки продукции, способы продвижения и продажи товаров также диверсифицируются - и объем информации возрастает многократно. Трудности с использованием данных возникают не только в бизнесе: например, научно-исследовательским организациям не хватает вычислительных мощностей для работы со сложными моделями, обработки изображений и других источников научных данных.

## **1.2 Технологии и методики анализа Больших Данных**

Технологии Big Data - серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия.

MapReduce - модель вычислений для пакетной обработки больших объемов данных, разработанная и используемая в компании Google для широкого круга приложений. Модель MapReduce отличается простотой и удобством использования, скрывая от пользователя детали организации вычислений на кластерной системе. Пользователю достаточно описать

процедуру обработки данных в виде двух функций – map и reduce, после чего система автоматически распределяет вычисления по кластеру, обрабатывает отказы машин, балансирует нагрузку и координирует взаимодействия между машинами.

Модель MapReduce накладывает ряд ограничений на программу для того, чтобы автоматизировать системное распараллеливание, запуск и управление вычислениями на кластере. С одной стороны, это значительно упрощает задачу программиста и практически не требует от него специальной квалификации. С другой стороны, накладываемые системой ограничения, не позволяют реализовать в ней решение произвольных задач. Однако, как показывает опыт последних лет, модель MapReduce может успешно применяться при решении широкого круга задач. Примерами задач, хорошо укладывающихся в данную модель, являются поиск в тексте, сортировка данных, индексирование документов, вычисление индексов цитируемости, статистический анализ, машинное обучение, обработка изображений и т.д. [50].

Используемая в Google реализация модели вычислений MapReduce является закрытой разработкой. Тем не менее, основные принципы данной реализации хорошо известны по публикациям в научных изданиях и докладам разработчиков на конференциях.

Реализация MapReduce в Google ориентирована на вычислительную инфраструктуру, состоящую из большого числа недорогих серверов из массовых комплектующих. Постоянный рост объемов, обрабатываемых данных требует соответствующего наращивания вычислительных ресурсов, в связи с чем используемая вычислительная среда должна обладать высокой масштабируемостью.

Этим обуславливается отказ от специализированных суперкомпьютерных архитектур в пользу более экономичных и масштабируемых, но менее надежных кластерных систем из серверов массового производства. Вычислительная инфраструктура Google насчитывает сотни тысяч серверов, размещенных в нескольких центрах обработки данных по всему миру.

Эффективная реализация MapReduce невозможна без эффективной организации хранения данных на кластерной системе. Для этой цели в Google применяется распределенная файловая система Google File System (GFS).

Как и любая распределенная файловая система, GFS ориентирована на обеспечение высокой производительности, масштабируемости, надежности и доступности. Отличия архитектуры GFS от других распределенных файловых систем обусловлены спецификой приложений и вычислительной инфраструктуры Google. Отметим главные особенности распределенной файловой системы GFS: высокая отказоустойчивость, ориентация на хранение файлов большого размера, оптимизация под операции записи в конец файла, эффективное использование сетевых ресурсов и оптимизация под высокую агрегированную пропускную способность, нестандартный интерфейс файловой системы, ослабленная модель целостности данных [50].

В реализации MapReduce применяется ряд оптимизаций, позволяющих повысить эффективность вычислений. Например, при распределении map-задач по машинам в кластере учитывается то, каким образом входные данные размещены в GFS. Управляющий сервер пытается отправить map-задачу на машину, хранящую соответствующий фрагмент входных данных, или же на машину, наиболее близкую к данным в смысле сетевой топологии. Подобная стратегия позволяет существенно снизить объем данных, передаваемых по сети во время запуска задания, и, тем самым, уменьшить время выполнения задания.

Платформа распределенных вычислений Hadoop разрабатывается на принципах open source в рамках организации Apache Software Foundation. Платформа ориентирована на поддержку обработки больших объемов данных на кластерных системах и заимствует многие идеи у закрытых технологий Google, таких как MapReduce, GFS и BigTable, фактически предоставляя их открытые реализации [50].

Распределенная файловая система Hadoop File System (HDFS), по сути, является общедоступным аналогом закрытой технологии GFS. HDFS обладает высокой отказоустойчивостью и нацелена на поддержку приложений,

связанных с обработкой больших объемов данных. Поэтому акцент делается на обеспечении высокой пропускной способности при доступе к данным в потоковом режиме и оптимизации хранения файлов большого размера. HDFS жестко следует модели однократной записи файла с последующим многократным чтением.

В настоящее время платформа Hadoop достигла достаточного уровня зрелости и масштабируемости для использования ее в реальных приложениях на больших кластерных системах.

Данные технологии применяются для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения информации по многочисленным узлам вычислительной сети.

Существует множество разнообразных методик анализа массивов данных, в основе которых лежит инструментарий, заимствованный из статистики и информатики.

*A/B testing.* Методика, в которой контрольная выборка поочередно сравнивается с другими. Тем самым удается выявить оптимальную комбинацию показателей для достижения, например, наилучшей ответной реакции потребителей на маркетинговое предложение. Большие данные позволяют провести огромное количество итераций и таким образом получить статистически достоверный результат.

*Association rule learning.* Набор методик для выявления взаимосвязей, т.е. ассоциативных правил, между переменными величинами в больших массивах данных. Используется в *data mining* [50].

*Classification.* Набор методик, которые позволяют предсказать поведение потребителей в определенном сегменте рынка (принятие решений о покупке, отток, объем потребления и проч.). Используется в *data mining*.

*Cluster analysis.* Статистический метод классификации объектов по группам за счет выявления наперед не известных общих признаков. Используется в *data mining*.

Crowdsourcing. Методика сбора данных из большого количества источников.

Data fusion and data integration. Набор методик, который позволяет анализировать комментарии пользователей социальных сетей и сопоставлять с результатами продаж в режиме реального времени.

Data mining. Набор методик, который позволяет определить наиболее восприимчивые для продвигаемого продукта или услуги категории потребителей, выявить особенности наиболее успешных работников, предсказать поведенческую модель потребителей [50].

Ensemble learning. В этом методе задействуется множество предикативных моделей за счет чего повышается качество сделанных прогнозов.

Genetic algorithms. В этой методике возможные решения представляют в виде «хромосом», которые могут комбинироваться и мутировать. Как и в процессе естественной эволюции, выживает наиболее приспособленная особь.

Machine learning. Направление в информатике (исторически за ним закрепилось название «искусственный интеллект»), которое преследует цель создания алгоритмов самообучения на основе анализа эмпирических данных.

Natural language processing (NLP). Набор заимствованных из информатики и лингвистики методик распознавания естественного языка человека.

Network analysis. Набор методик анализа связей между узлами в сетях. Применительно к социальным сетям позволяет анализировать взаимосвязи между отдельными пользователями, компаниями, сообществами и т.п.

Optimization. Набор численных методов для редизайна сложных систем и процессов для улучшения одного или нескольких показателей. Помогает в принятии стратегических решений, например, состава выводимой на рынок продуктовой линейки, проведении инвестиционного анализа и проч.

Pattern recognition. Набор методик с элементами самообучения для предсказания поведенческой модели потребителей [50].

Regression. Набор статистических методов для выявления

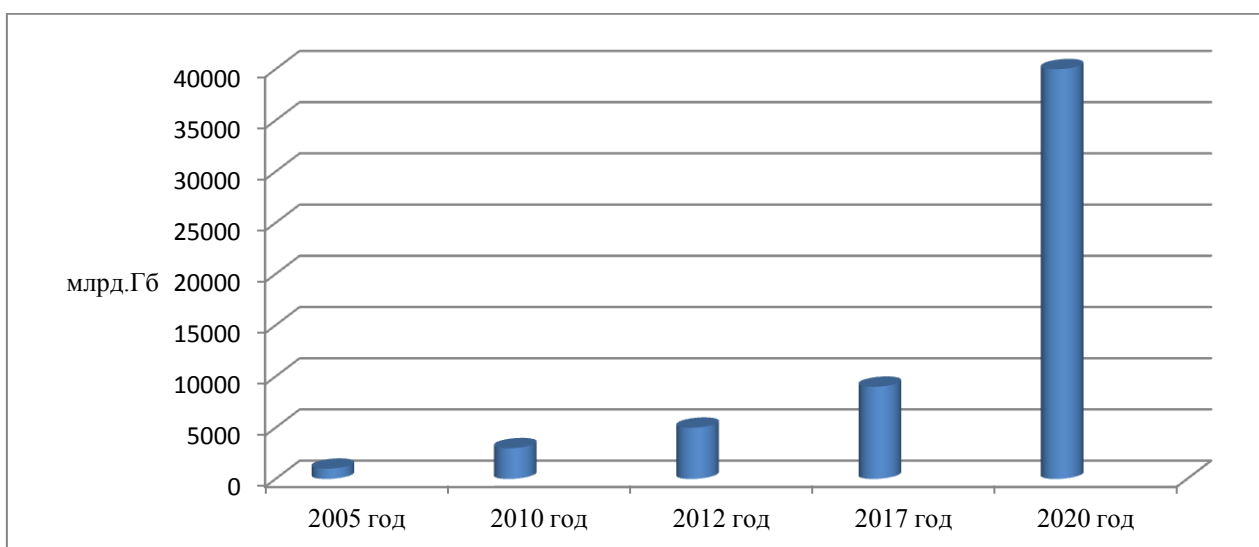
закономерности между изменением зависимой переменной и одной или несколькими независимыми. Часто применяется для прогнозирования и предсказаний. Используется в data mining.

Simulation. Моделирование поведения сложных систем часто используется для прогнозирования, предсказания и проработки различных сценариев при планировании.

Time series analysis. Набор заимствованных из статистики и цифровой обработки сигналов методов анализа, повторяющихся с течением времени последовательностей данных. Одни из очевидных применений – отслеживание рынка ценных бумаг.

В настоящее время множество компаний следят за развитием технологий Big Data. Аналитическая компания IDC представила в 2017 г. отчет «Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East», в котором предсказывалось, что объемы информации будут удваиваться каждые 2 года в течение следующих 8 лет [21].

За ближайшие 7 лет количество данных в мире достигнет 40 ЗБ (1 ЗБ = 1021 байт), а это значит, что на каждого жителя Земли будет приходиться по 5200 ГБ данных (рисунок 3).



Источник: [21]

Рисунок 3 - Общий объем цифровых данных в мире



В современных условиях организации создают большое количество неструктурированных данных, таких как текстовые документы, изображения, видеозаписи, машинные коды, таблицы и т. д. Вся эта информация хранится во множестве репозиториях, порой даже за пределами организации.

Компании могут иметь доступ к огромному массиву собственных данных и не иметь необходимых инструментов, которые могли бы установить взаимосвязи между этими данными и сделать на их основе значимые выводы.

Традиционные методы анализа информации не могут угнаться за огромными объемами постоянно растущих и обновляемых данных, что в итоге и открывает дорогу технологиям Big Data.

Можно выделить следующие особенности технологий Big Data:

- работа с информацией огромного объема и разнообразного состава;
- информация весьма часто обновляется и находится в разных источниках;
- качественно отличающийся метод открывающей аналитики для выявления практических знаний, которые непосредственно монетизируются в прибыль;
- наглядное отображение отчетов и возможности сценарного анализа («что, если...»);
- цель применения технологий Big Data - увеличение эффективности работы, создание новых продуктов и повышение конкурентоспособности.

Согласно отчету компании McKinsey «Global Institute, Big data: The next frontier for innovation, competition, and productivity», данные стали важным фактором производства наряду с трудовыми и капитальными ресурсами. Использование больших данных станет основой конкурентного преимущества и роста компаний [23].

Объем информации на предприятии неуклонно растет за счет данных, полученных с датчиков, измерительных и «умных» устройств. Самыми перспективными устройствами считаются датчики, которые могут передавать данные в режиме реального времени.

Все устройства на предприятии с помощью таких датчиков могут быть объединены в сеть, а технологии Big Data позволят обрабатывать информацию, поступающую с них, и проводить необходимые мероприятия в автоматическом режиме.

Например, предприятия могут с помощью датчиков получать ежеминутные данные о состоянии своего оборудования и на основе этих данных предсказывать оптимальное время для замены и обслуживания. Слишком ранняя замена приведет к дополнительным расходам, а поздняя - к потере прибыли вследствие простоя оборудования.

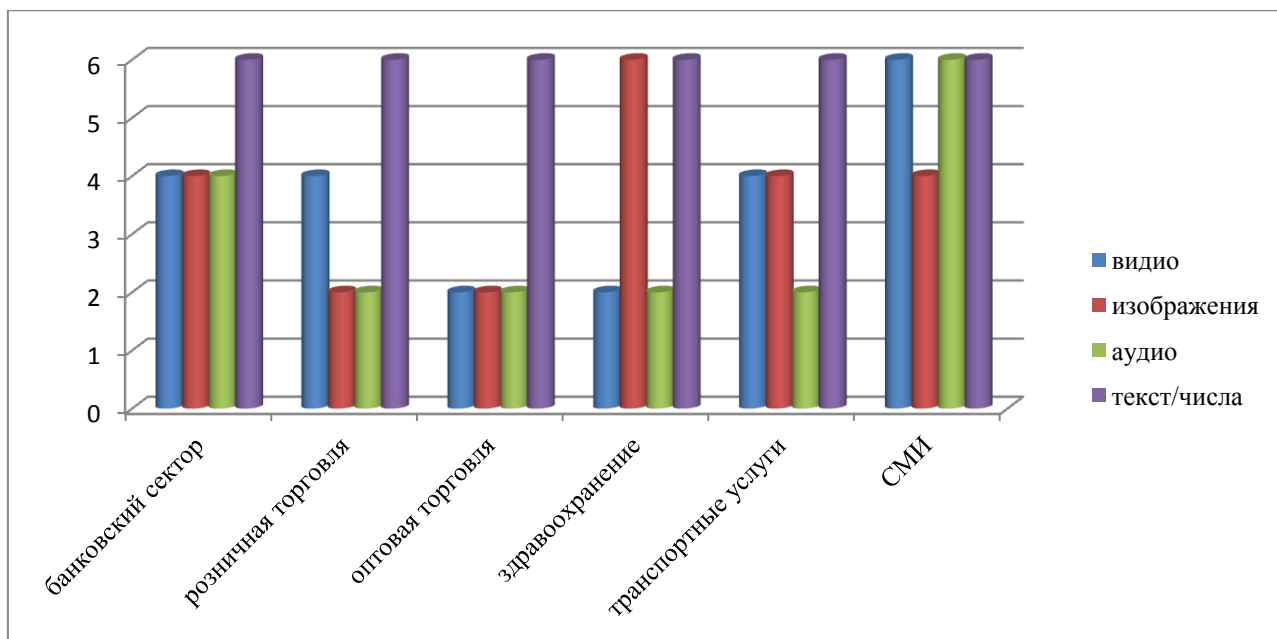
По оценке компании Cisco, к 2019 г. будет существовать более 1,7 млрд межмашинных соединений [23].

Технологии Big Data могут быть полезны для решения следующих задач:

- прогнозирование рыночной ситуации;
- маркетинг и оптимизация продаж;
- эффективное сегментирование клиентов;
- совершенствование товаров и услуг;
- принятие более обоснованных управленческих решений на основе анализа Big Data;
- оптимизация портфеля инвестиций;
- повышение производительности труда;
- эффективная логистика;
- мониторинг состояния основных фондов.

Методика и инструменты работы со структурированными данными уже давно созданы.

Это реляционная модель данных и системы управления базами данных. Но в современных условиях предприятиям нужно обрабатывать большие объемы неструктурированных данных различных типов (рисунок 4), а для этой работы прежние методы не совсем подходят.



Источник: [23]

Рисунок 4 - Превалирующие типы информации для разных сфер деятельности

В настоящее время все более популярной становится модель работы с Big Data, реализованная в проекте Apache Hadoop.

Большинство продуктов для работы с Big Data обладают высокоэффективной системой обработки огромных объемов информации и ее аналитики в реальном времени. Ожидаемый эффект от внедрения Big Data может варьироваться в зависимости от типа деятельности и реализуемой политики конкретного предприятия. При работе с Большими Данными применяют методы манипуляции знаниями: различные методы теории распознавания и классификации, методы разведывательного анализа и обобщения данных, интеллектуальные подходы в виде генетических алгоритмов, нейросетей и других ответвлений искусственного интеллекта.

Главные задачи платформы Hadoop - хранение, обработка и управление данными. Основными составляющими платформы Hadoop являются:

- отказоустойчивая распределенная файловая система Hadoop Distributed File System (HDFS), при помощи которой осуществляется хранение;
- программный интерфейс Map Reduce, который является основой для написания приложений, обрабатывающих большие объемы структурированных

и неструктурированных данных параллельно на кластере, состоящем из тысяч машин.

Платформа Hadoop позволяет сократить время на обработку и подготовку данных, расширяет возможности по анализу, позволяет оперировать новой информацией и неструктурированными данными. Результаты проекта по внедрению технологии Hadoop подтверждают целесообразность ее использования (таблица 2).

Таблица 2 - Результаты проекта

Платформа	Описание оборудования	Примерная стоимость оборудования, руб	Среднее время работы одного отчета, мин
БД Oracle	Сервер класса	12 млн.	59
Кластер Hadoop	10 рабочих станций	250 тыс.	66
Hadoop с учетом оптимального кластера	10 рабочих станций	300 тыс.	40

Источник: [21]

Решения, построенные на базе технологии Hadoop, обладают рядом существенных преимуществ. Основные из них приведены в таблице 3.

Таблица 3 - Преимущества решения на базе Hadoop

Преимущество	Краткое описание
Снижение времени на обработку данных	При обработке данных на кластере можно существенно сократить время на обработку данных
Снижение стоимости оборудования	Применение технологии Hadoop позволяет сократить затраты на оборудование, требуемое для хранения и обработки данных, в десятки раз
Повышение отказоустойчивости. Технология позволяет построить отказоустойчивое решение	Выход из строя одного или нескольких узлов кластера влияет только на производительность системы, при этом система продолжает корректно работать и предоставлять сервис конечным пользователям.
Линейная масштабируемость	Решение позволяет наращивать производительность просто за счет добавления новых узлов кластера. При этом производительность кластера возрастает линейно.
Работа с неструктурированными данными	Технология позволяет осуществлять сложную обработку любых файлов, в том числе неструктурированных, благодаря чему такие данные могут быть эффективно обработаны и использованы

Источник: [21]

В России доступны решения, использующие технологию Big Data от ведущих производителей (Cisco, HP, IBM, Microsoft, Oracle, Apache), но проектов по реализации очень мало.

Отечественный рынок находится в зачаточной стадии развития, но все без исключения аналитики прогнозируют взрывной рост технологий Big Data.

В октябре 2017 г. корпорация EMC провела исследование среди российских компаний, в ходе которого было выявлено, что использование Big Data ведет к существенному улучшению процессов принятия решений, повышает конкурентоспособность компаний и упрощает управление рисками:

- 70 % респондентов в России считают, что анализ данных их компании поможет принимать более обоснованные решения, а 35 % подтверждают, что высшее руководство их компаний полагается на результаты анализа Big Data при принятии стратегических бизнес-решений;

- 31 % респондентов сообщили, что их компании получили конкурентное преимущество в результате внедрения технологий Big Data, а 51 % считают, что отрасли, в которых используются такие технологии, покажут наиболее быстрый рост;

- 51 % респондентов согласны, что технологии анализа Big Data помогут в выявлении и предотвращении кибератак. Это может оказаться ключевым фактором, так как только 67 % респондентов в России уверены, что они смогут полностью восстановить все свои данные [33].

В современном мире, где информация часто обновляется и поступает из разных источников, предприятиям приходится работать с огромными массивами данных.

Технологии Big Data позволяют предприятиям хранить, структурировать и анализировать большие объемы информации. Это помогает руководству предприятия находить связь между различными факторами и использовать эту привилегию для получения благоприятного эффекта.

Одним из наиболее перспективных программных обеспечений для работы с Big Data, оптимизированных для промышленных предприятий, является

платформа «Hadoop», разработанная компанией «Apache Software Foundation». Большие данные - это не очередной ажиотаж на ИТ-рынке, это системный, качественный переход к составлению цепочек ценностей, основанных на знаниях.

По эффекту его можно сравнить с появлением доступной компьютерной техники в конце прошлого века. Сейчас эта технология находится в фазе ожидания инвесторов: они следят, схлынут ли спекуляции вокруг новой технологии, или же это значимая инновация в стадии проникновения на рынок. В ближайшие 5 лет произойдет исправление недостатков технологии, и к 2018 г. начнется ее широкое распространение.

Таким образом, технологии Big Data - серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия.

Реализация MapReduce в Google ориентирована на вычислительную инфраструктуру, состоящую из большого числа недорогих серверов из массовых комплектующих. Постоянный рост объемов обрабатываемых данных требует соответствующего наращивания вычислительных ресурсов, в связи с чем используемая вычислительная среда должна обладать высокой масштабируемостью.

### **1.3 Большие Данные и управление бизнес-процессами банка**

Текущие сложные экономические условия сказываются на всех сферах, в том числе и на банковской сфере. Просроченная задолженность по кредитам растет и встает вопрос, как управлять ей, как минимизировать ее. Очень остро стоит вопрос получения актуальной контактной информации и данных о клиентах, для установления контакта и урегулирования просроченной задолженности. Такими данными могут быть как место жительства и регистрации клиента, и его контактная информация (телефоны, электронные адреса), так и информацию по его счетам и продуктам, которыми он пользуется.

Стоит отметить, что у банков уже давно реализован программный механизм заведения клиента в свою базу, отслеживания использования того или иного продукта. Данные программы предоставляют большой набор функций по обработке и использованию данной информации. В частности, данные программы используются для отслеживания факта выхода клиента на просроченную задолженность по кредиту или кредитной карте. С помощью данных программ собирается вся информация о клиенте, которая есть в распоряжении банка, для установления контакта с клиентом как личного, так и удаленного по средствам смс, звонков, электронной почте.

Контакты с клиентом поддерживаются для разных целей: для предоставления информации о новых услугах банка, для информирования клиента об услуге, которой он уже пользуется, для информирования клиента о выходе на просроченную задолженность, для урегулирования вопроса по погашению просроченной задолженности [23].

Big Data - это методы работы с большим объемом структурированных и неструктурированных данных. Большие данные характеризуются тремя характеристиками:

1. Volume - большой объем данных, который постоянно увеличивается.
2. Velocity - скорость работы с подобными данными.
3. Variety - разнообразность хранимых и обрабатываемых данных.

Реляционные базы данных уже не позволяют хранить данные и предоставлять решения для быстрой обработки именно больших объемов данных. Здесь речь идет о терабайтах информации. Прежние программы и аппаратные средства не позволяют в полной мере проанализировать и обработать такие большие объемы данных.

Более того, Big Data позволяет работать как со структурированной информацией, так и с неструктурированными данными, такими как картинки, текст, видео.

Уже существует достаточно большой набор инструментов, технологий, позволяющих работать с Big Data:

1. NoSQL;
2. MapReduce;
3. Hadoop.

Эти и другие продукты предназначены для работы с Big Data работают по принципу «шеринга». Данный принцип может быть объяснен следующим образом. Существует одна база данных, которая принимает запрос на получение определенных данных [23].

Эта база данных пересылает данный запрос на множество других баз данных, и собирает с них уже сводную информацию по запросу. Тем самым база данных первополучатель запроса имеет дело уже со значительно меньшей по объему информацией и более структурированной.

Данная технология позволяет ускорить время обработки запросов, которые требуют анализа больших объемов данных имеющих разные характеристики.

Направление по применению Big Data в работе банка, в сфере сбора просроченной задолженности считается перспективным так, как для сбора просроченной задолженности необходим контакт с должником. Для установления связи с клиентом, нужна актуальная контактная информация.

Благодаря технологии Big Data становится возможным обрабатывать не только те данные, которые есть у банка для поиска подобной информации, но искать и контактную информацию и в других местах. Например, в социальных сетях, форумах и других ресурсах, где клиент мог бы оставить свои актуальные данные. Для обработки таких больших объемов информации и требуется технологии Big Data.

Более того информация в социальных сетях может быть представлена и в формате текста или даже фото или видео. Следовательно, переход на работу с большими объемами данных как никогда актуален в данной сфере. Для поиска



и обработки данных клиентов в социальных сетях, базах данных мобильных операторов как раз и необходима Big Data.

Однако в этом случае речь идет не о традиционном представлении о Big Data, как о статистическом инструменте, а о распределенной Big Data. То есть, поиск и анализ данных ведется не в одной централизованной базе данных, а сразу в нескольких. Это обусловлено тем, что поиск ведется в нескольких независимых базах данных. Так, уже существует множество примеров применения Big Data для сборов информации в социальных сетях.

76% банков заявляют, что Big Data позволяют привлекать новых клиентов, лучше взаимодействовать с ними и поддерживать их лояльность. Можно с уверенностью утверждать, что более половины аналитических работ выполняется по заказу маркетологов. Глубокая сегментация клиентской базы, поиск неявных закономерностей, персонализация продуктовых предложений - вот основные задачи по этому направлению, которые в крупных розничных банках решаются с помощью Big Data. Банкам нужны новые источники информации, позволяющие посмотреть на клиента шире, формировать новые уникальные предложения [48].

Примером использования Big Data в российских банках может служить проект банка «Тинькофф». В банке уже восемь лет используются BI-системы, создано традиционное хранилище данных Data Warehouse (DWH), в котором накоплено 80 Тб данных (в течение двух лет планируется расширить объем данных до 700 Тб), а система Big Data стала дополнением к этой системе.

«Big Data - это прежде всего внешние данные (управление хранилищем данных и отчетности банка «Тинькофф»).

Перед управлением DWH была поставлена большая задача - анализировать интересы и поведение посетителей сайта банка. Внешние источники порождали колоссальный объем текстовых полуструктурированных данных, что, конечно, для традиционного DWH, построенного в банке на массивно параллельной СУБД Greenplum, совсем не подходило.

В итоге в банке был развернут кластер Hadoop на основе дистрибутива Cloudera, который лег в основу целевого хранилища данных, а точнее озера данных, для внешних данных. Были выделены основные концептуальные слои данных:

RAW - слой сырых данных, куда загружаются файлы, логи, архивы самых разнообразных форматов данных,

ODD - Operational Data Definition, куда загружаются данные в формате, приближенном к реляционному, в том числе данные, полученные в результате предобработки данных из RAW перед загрузкой в DDS;

DDS - Detail Data Store, где собирается консолидированная модель детальных данных. Для хранения данных в этом слое была выбрана концепция Data Vault;

MART - прикладные витрины данных.

С Hadoop работают Data Scientist и системы Machine Learning. Этим в банке занимается специальная лаборатория Big Data. Она же отвечает за подготовку и качество данных, подготовку данных для маркетинга. Данные из Hadoop и СУБД Greenplum используются для систем бизнес-аналитики, пользователями которой могут быть уже обычные бизнес-службы банка. Data Scientist генерируют много гипотез и предлагают их бизнесу, который дает их оценку - обратную связь

Hadoop - фреймворк, предназначенный для построения распределённых приложений для работы с данными очень большого объёма, при котором приложение разбивается на множество независимых частей, каждая из которых может исполняться на отдельном узле, многими считается чуть ли не синонимом термина Big Data.

Big Data может быть хорошим инструментом подготовки данных для анализа состояния не только клиентов, но и для подготовки отчетов о состоянии самого банка, повышая эффективность существующих BI-систем.

К ним относятся:

- ad hoc-отчетность - отчетность по желанию и в любой момент времени;

- регулярная отчетность — принятые в банке формы отчетности, на базе которых производятся какие-то действия, например, строится KPI. К ней же относится отчетность для регуляторов;

- прогнозная отчетность, например, сколько будет у банка клиентов и куда лучше вкладывать деньги;

- операционная отчетность - понимание того, что на данный момент происходит в банке — объем транзакций, число выданных кредитов [23].

Примером такого использования технологий Big Data может служить все тот же банк «Тинькофф». Добавление внешних источников информации и обработка их с помощью технологий Big Data позволяют значительно увеличить количество данных, доступных для анализа, дают возможность проанализировать рынок и получить понимание тенденций его развития, оценить реальное положение банка на рынке и сделать более точные прогнозы.

Другой пример - проекты банков по анализу Больших Данных для оптимизации размещения и загрузки банкоматов. Лишнюю наличность в банкоматах держать невыгодно, но и раздражать клиентов, не имеющих возможности снять деньги, банкам не хочется. Решения, основанные на анализе данных, поступающих с банкоматов, и информации из всевозможных внешних источников, позволяют построить модель, прогнозирующую потребность в наличности. На основе оценки оптимального объема денег в банкоматах можно составить график инкассации банкоматов, снизить число инкассаций и жалоб от недовольных клиентов.

Еще один аспект работы с клиентами - идентификация подозрительных транзакций, и тут роль глубокой аналитики переоценить невозможно. Проблема организации эффективной борьбы с мошенничеством присуща не только финансовому сектору. Профильные решения класса Data Mining не новы для

рынка, но тут задача должна решаться оперативно, и потому используются решения категории in-memory - вычисления в оперативной памяти.

Например, банк HSBC внедрил технологии больших данных в состав решения для противодействия мошенничеству с кредитными картами. В результате эффективность службы по выявлению случаев мошенничества повысилась в три раза, а точность его выявления - в десять раз. За первые же две недели эксплуатации семь специалистов службы безопасности HSBC выявили новые криминальные группы и схемы с общим потенциальным ущербом более \$10 млн.

Колоссальные перспективы раскрывает перед банком сочетание Big Data и геоаналитики. Например, клиент заявил о потере карты и одновременно по этой же карте в магазине совершена покупка. А банк оперативно идентифицировал местонахождение телефона клиента и может доказать, что клиент пытается смошенничать (если телефон и карта находятся рядом). Или заблокировать транзакцию, не дожидаясь заявления клиента о потере карты (если телефон и карта находятся на приличном расстоянии друг от друга).

В годовом отчете Сбербанка за 2017 год, говорится, что анализ Больших Данных по активности клиентов позволил банку снизить уровень неработающих кредитов и сократить риски. Это, в свою очередь, «привело к снижению процентных ставок по кредитам, формированию специальных предложений с более интересными условиями для разных сегментов заемщиков» [55].

В Сбербанке использование технологий Big Data помогает точнее определить текущий риск-профиль клиента, его интересы и потребности, что в итоге позволяет банку делать «своевременные и персонализированные предложения» в части предоставляемых услуг. Например, использование информации по движению средств по счетам клиентов, анализируя их структуру трат, банк может оценить умение клиента распоряжаться его денежными средствами. Это, в свою очередь, напрямую влияет на вероятность невозврата денежных средств в срок.

Спектр задач, связанных с анализом данных в банке, очень широкий: от анализа клиентского опыта для предоставления оптимального кредитного предложения, до риск-менеджмента, до управления информационной безопасностью и оптимизации ИТ-процессов, добавили в Сбербанке.

На сегодняшний день в распоряжении Альфа-Банка находится большое количество данных, консолидированных в одном месте - Едином Корпоративном хранилище данных, которое Центр ВІ совместно с бизнес-подразделениями банка развивает и наполняет различной аналитикой уже более 10 лет.

Для развития Хранилища данных в Альфа-Банке успешно используются оптимизированные программно-аппаратные комплексы Oracle Exadata и Oracle Big Data Appliance. Первые применяются для работы с реляционными базами данных, вторые - с неструктурированными данными. Oracle Big Data Appliance позволяет хранить информацию о клиентах в виде файлов, гибко настраивать структуру файлов по полям, если она меняется.

Альфа-Банк начал использовать технологии Big Data с хранения глубокой кредитной истории по кредитным заявкам в соответствии с требованиями Basel II, поэтому Oracle Big Data Appliance изначально приобреталась банком как решение для удовлетворения требований регуляторов.

Банк планирует использовать сбор информации о клиенте из разных источников, анализ неструктурированных данных для формирования целевого предложения для клиента.

Технологии больших данных можно применять не только для анализа клиентской среды, но для управления рисками, взаимоотношениями с клиентами, использовать в маркетинге, борьбе с мошенничеством. Применение этих технологий в финансовом секторе имеет большой потенциал, поэтому направление Big Data будет развиваться и далее.

Возможно дальнейшее развитие применения Big Data в банковской сфере. Одним из перспективных направлений в этой сфере является определение и анализ факторов, которые необходимо учитывать при выдаче кредитов. Речь

идет о создании модели с множеством переменных, которая позволила бы анализировать различные параметры, влияющие на потенциальный риск невозврата кредита или выход клиента на просроченную задолженность. Будет возможно непосредственно влиять на качество кредитного портфеля банка и сделать его более здоровым.

Таким образом, направление по применению Big Data в работе банка, в сфере сбора просроченной задолженности считается перспективным так, как для сбора просроченной задолженности необходим контакт с должником. Для установления связи с клиентом, нужна актуальная контактная информация.

Благодаря технологии Big Data становится возможным обрабатывать не только те данные, которые есть у банка для поиска подобной информации, но искать и контактную информацию и в других местах. Например, в социальных сетях, форумах и других ресурсах, где клиент мог бы оставить свои актуальные данные.

Применение технологий Big Data для поиска контактной информации о клиентах в целях реализации функции банка по сбору просроченной задолженности является актуальным направлением.

Уже существующая практика применения технологий больших объемов данных для схожих целей создает практическую базу для эффективного применения Big Data указанной функции банка. Однако применение подобной технологии только этим не ограничивается и может быть направлено на создание многофакторной модели анализа клиента на стадии оценки заявки на выдачу кредита. Подобная модель позволила бы минимизировать риски невозврата кредита или выхода клиента на просроченную задолженность. В купе, данные решения позволили бы банку оздоровить кредитный портфель и минимизировать экономические риски.

## **2 Анализ использования Больших Данных в банковской сфере**

### **2.1 Возможности и риски Больших Данных в банках**

Риск – это ситуативная характеристика деятельности любого производителя, в том числе банка, отображающая неопределенность ее исхода и возможные негативные последствия в случае неуспеха. Под риском следует понимать не только возможность банка остаться в убытке. Стоит отметить, что на протяжении всего процесса функционирования банка существует малая, либо большая вероятная работы исключительно на прибыль, т.к. основная масса банков стремится к получению максимально высокой прибыли, что повысит и рыночную конкурентоспособность, а также уровень привлекательности для клиентов [29].

Банковская деятельность, как и других финансовых и нефинансовых институтов, связана с рисками. Само понятие «риск» многогранно, но для банка оно означает вероятность понесения потерь или ухудшение ликвидности вследствие наступления неблагоприятных событий.

Риск чаще всего реализуется в результате несоответствия прогнозов реально экономическим событиям. Выделяют следующие виды банковских рисков:

- кредитный риск. Подразумевает под собой возможность падения прибыли и даже потерю части акционерного капитала в результате того, что заемщик оказывается не способным погашать и обслуживать долг;
- риск ликвидности. Возможная угроза прибыли и акционерному капиталу банка в результате затруднения в получении средств путем реализации части активов или приобретения нового займа по приемлемой цене;
- процентный риск. Отражает вероятность банка потерять доход в результате непогашения процентных платежей заемщиком;
- риск текущих расходов. Возможное снижение прибыли банка из-за непредвиденных расходов на содержание аппарата сотрудников и прочих расходов, обеспечивающих нормальный ритм работы учреждения;

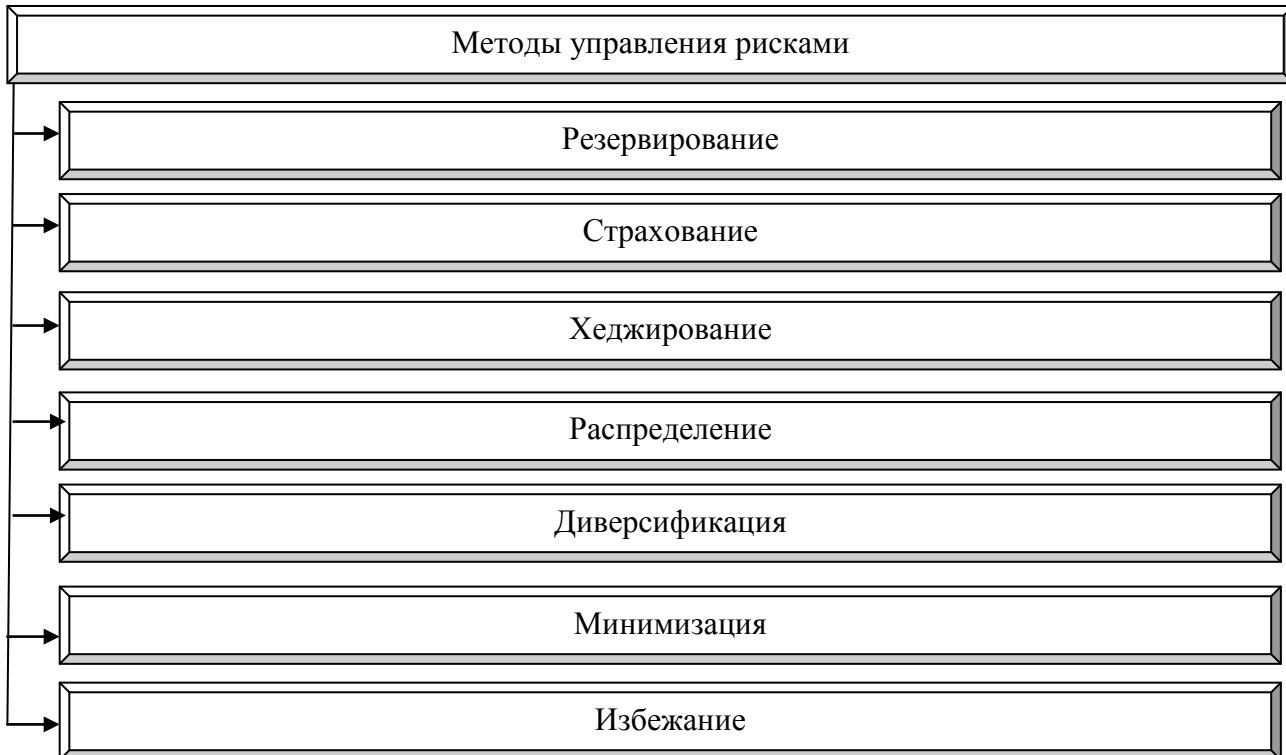
– валютный риск. Опасность валютных потерь, связанных с изменением курса иностранной валюты по отношению к национальной валюте при проведении международных кредитных, валютных и расчетных операций;

– риск неплатежеспособности банка. Использование банком акционерного капитала для погашения своих обязательств при отсутствии каких-либо других источников.

Понимание сущности риска приносит мало пользы без возможности воздействия на них. Процесс управления рисками представляет собой совокупность действий, которые направлены на выявление проблем риска и разработку способов и методов их решения [29].

Целью управления рисками является эффективность управления банком в условиях неопределенности. Основной задачей риск-менеджмента является минимизация негативных влияний рисков на финансовые результаты банков.

Методы регулирования рисками, с учетом формы риска, продемонстрированы на рисунке 5.



Источник: [18]

Рисунок 5 - Методы управления рисками



Управление рисками очень важно для банков. Эта культура должна быть частью организационной культуры и операционной модели банка для того, чтобы анализ рисков использовался практически при принятии решений.

Для того, чтобы практически определять эффективность того или иного решения по привлечению капитала или вложению капитала, необходим количественный анализ рисков и регулярный мониторинг рисков, в том числе мониторинг внешних источников информации. Следование той или иной формальной модели количественной оценки рисков также уменьшает необходимые затраты на соответствие нормам регулирования банков.

Для внедрения подобного подхода требуется изменить сознание людей, которые сегодня работают с рисками. Необходимо предоставить людям действительно удобные инструменты. И, наконец, необходимо преобразовать регламенты работы так, чтобы они соответствовали гибкости предлагаемого подхода.

Система, основанная на Big data, позволяет объединить следующие направления анализа рисков на основе общей архитектуры рисков, информационной модели, системы отчетности и аналитики, включая ноу-хау, эвристики и аналитики в реальном времени:

- финансовые риски;
- риски несоответствия регулированию;
- операционные риски;
- риски финансовых преступлений (AML, мошенничество, Case Management);
- it-риски (безопасность, непрерывность процессов, достоверность данных) [18].

Банки стремятся выявлять проблемы на ранних стадиях и прилагают все усилия для взаимовыгодного решения проблем с задолженностью у клиентов. В текущей сложной экономической ситуации разработан ряд антикризисных мероприятий в части работы с проблемными активами: совершенствование кадров, работающих в подразделениях по работе с проблемными кредитами;

создание антикризисных штабов, работающих с крупнейшими клиентами; комплекс мер по упрощению реструктуризации кредитов физических лиц.

Мошенники стараются всяческими путями добраться до финансовых средств онлайн пользователей, использующих для бытовых нужд платежные системы, с их электронными кошельками.

Интересует кибернетических жуликов, прежде всего, реализуемые большинством банковских организаций системы «Интернет-банкинга», то есть услуги интерактивного доступа к персональным банковским счетам.

Время от времени онлайн-финансы пользователей подвергаются атакам вредоносных программ, компьютерных вирусов, позволяющих мошенникам похитить их конфиденциальные данные, пароли доступа, финансовые реквизиты, логины и тому подобную информацию.

Получив такую информацию, онлайн-мошенники могут получить доступ к банковским счетам, привязанным к системе Сбербанк ОнЛайн, участником которой является пользователь. Мошенники могут вывести средства с его счета на собственный счет, при этом аферисты часто предварительно проводят денежные средства через банковские счета таких же жертв, дабы сложнее было проследить их финансовые реквизиты, где осядут означенные денежные средства. Так они запутывают свои следы в системе электронных платежей.

Интернет-банкинг – это сервис, активно пропагандируемый отечественными банками, так же подвержен атакам кибернетических мошенников, как и системы электронных платежей, интерактивные магазины. Причина тому, в первую очередь, – наивность клиентов означенного банковского сервиса, их безграничная уверенность в безопасности электронных платежей посредством данного инновационного сервиса.

Основные риски, конечно, связаны с несанкционированным доступом к банковским счетам пользователя услуги Сбербанк ОнЛайн. Клиент банка рискует потерять безвозвратно все хранящиеся на его банковских счетах денежные средства. Разрыв доступа к электронной сети, нестабильный доступ к интернету, прочие технические проблемы существенно повышают его риски,

играют на руку кибернетическим жуликам.

Описанные неполадки могут привести к некорректной обработке финансовой операции, когда средства с баланса система списала, но адресат платежа средства так и не получил, то есть платеж «затерялся».

ПАО «Сбербанк России», дабы обезопасить онлайн взаиморасчеты, использует различные меры информационной защиты. Им активно применяется технология SSL-шифрования данных. Данный протокол позволяет организовать безопасный шлюз связи, где перехват конфиденциальных данных практически невозможен. Сервер банковской организации посредством системы шифрования SSL верифицирует права доступа клиента сервиса Сбербанк ОнЛайн.

Система одноразовых паролей, которые владельцы банковских карт получают в процессе использования банкомата, на сегодня редко практикуется отечественными банками. Эксперты предупреждают тех граждан, чьи банки все же используют данную технологию защиты, заботиться, чтобы означенные пароли не стали известны посторонним лицам.

ПАО «Сбербанк России» столкнулся с проблемой фейковых приложений «Сбербанк Онлайн» на операционной системе Android.

Фейковое приложение маскируется под оригинал, заражает трояном мобильное устройство. С помощью вируса злоумышленники получают доступ к данным клиента для доступа к реальному аккаунту в системе Сбербанка.

Уязвимость связана с платформой Android, которая позволяет пользователям загружать нелицензионные программы.

Для борьбы с трояном ПАО «Сбербанк России» в собственное приложение внедрил антивирус, который выявляет вредоносные приложения и блокирует их установку. Достаточно скачать официальное приложение «СбербанкОнлайн» из GooglePlay». Банк также рекомендует скачивать приложения только с официальных ресурсов и внимательно следить, на какие данные приложение запрашивает доступ.

О массовом заражении смартфонов системы Android сообщили эксперты

Group-IB. Они выявили троян, который маскируется под приложение одного из трех крупнейших российских банков. Приложение попадает в телефон не через GooglePlay, а через поисковой запрос - когда пользователь ищет приложение банка, оно выдается ему в результате поиска. В таком случае пользователь не подозревает, что самостоятельно установил вирус на свой смартфон.

Многие банки используют систему многоуровневой защиты, то есть помимо стандартной идентификации клиента по логину и паролю используют другие методы защиты, в особенности, когда речь идет об осуществление расходных операций (онлайн платежи, денежные переводы). Используется система СМС верификации посредством мобильной связи, технология цифровой подписи, специальные электронные устройства и тому подобное.

Система СМС верификации наиболее популярна, проста, эффективна, не требует специального оборудования, установки программного обеспечения. Все операции участник сервиса Сбербанк ОнЛайн подтверждает, вводя в своем электронном бумажнике одноразовый пароль, полученный на привязанный к его аккаунту (счету) номер сотового телефона.

Технология цифровой подписи или ЭЦП – это система генерации индивидуальных кодов, используемых для подтверждения финансовых операций пользователя. Данная технология позволяет эффективно идентифицировать пользователя в платежной системе, правда имеет один недостаток. Мошенники могут получить означенный индивидуальный код, если зарадят заблаговременно компьютерное оборудование пользователя вредоносным софтом.

Специальные электронные устройства менее популярны, когда речь идет о сервисе интернет-банкинга, ведь пользователь должен приобрести внешнее устройство, способное генерировать одноразовые пароли. Подключается данное специальное оборудование к ПК через порт USB, обычно работает без специальных программных приложений. Многие эксперты считают данную систему защиты безнадежно устаревшей, несовершенной версией технологии ЭЦП. Сбербанк настоятельно рекомендует клиентам сервиса Сбербанк ОнЛайн

придерживаться простых, но весьма эффективных мер предосторожности. Онлайн мошенники постоянно ищут новые лазейки, слабые места в системах электронной защиты банков.

Клиенты не должны полностью полагаться на систему защиты банковской сети, надеяться, что в случае подобного банковское предприятие вернет похищенные деньги. В действительности менее половины, пострадавших получают означенные компенсации от банка в 100% объеме, а 10-20% банк выплачивает компенсацию частично. Около 40% пострадавших, не получают от банка ничего. Именно поэтому, используя услугу Сбербанк Онлайн, необходимо соблюдать ряд простых правил.

Рекомендуется клиентам всегда вводить электронный адрес банка, сервис которого они используют, только вручную, без гиперссылок, которые размещены на посторонних веб-ресурсах. Существует риск угодить на поддельный сайт, созданный мошенниками. Мошенники могут украсть логин и пароль доступа, тем самым проникнув в чужой аккаунт и получить конфиденциальные данные.

Всегда нужно использовать только сложные, трудные для подбора пароли доступа. Опасно, когда пароль имеет связь с личностью пользователя, его датой рождения, кличкой его собаки или именем его ребенка. Пароль доступа должен содержать сложную композицию из букв и цифр. Запомнить подобные пароли сложно, но все же нельзя, чтобы пароли от разных онлайн сервисов (Сбербанк Онлайн, платежные системы, аккаунты на различных социальных проектах, форумах, почтовой службе) пользователя полностью или частично совпадали.

Необходимо быть крайне осторожным, пользуясь бесплатным беспроводным интернетом (Wi-Fi), если тот не защищен системой паролей. Работая с платежными системами посредством подобного интернет соединения можно и не заметить, как мошенники перехватят конфиденциальную информацию (пароли доступа, коды активации платежных операций и тому подобное), ведь классически такое открытое соединение имеет слабую защиту.

Рекомендуется использовать соединение «https» зашифрованного типа,

ведь работа с системой Сбербанк ОнЛайн сопряжена с передачей конфиденциальных данных. Определить это визуально можно, если посмотреть в адресную строку браузера, где должен отображаться соответствующий символ (например, замок).

Сегодня наработаны методы и инструментарий борьбы с техническим взломом, эксплойтами, DoS- и DDoS-атаками. Почти все кредитные учреждения имеют соответствующие средства противодействия.

Поэтому злоумышленники все чаще снова стали использовать методы социальной инженерии, и сейчас на новом витке развивается борьба с этими методами, которые действуют на невнимании человека и его откровенном обмане. Мошенники изучают процессы банка, поведение его сотрудников, «поведение» систем и подделываются под них, имитируя реальные ситуации.

## **2.2 Проблемы использования Больших Данных в банковской деятельности**

Массовое внедрение технологий анализа Больших Данных осложнено тем, что банки зачастую используют разрозненные или просто устаревшие платформы. Тем не менее, уже есть примеры того, как сотрудники, отвечающие за информационную безопасность, предотвращали мошеннические операции. Помимо технологии Big Data эксперты также считают, что бороться с мошенниками позволяет внедрение современных систем идентификации пользователей. Одним из примеров является так называемая непрерывная поведенческая идентификация, анализирующая поведение клиентов на протяжении длительного времени. Это делается при помощи привязки счета к мобильному телефону.

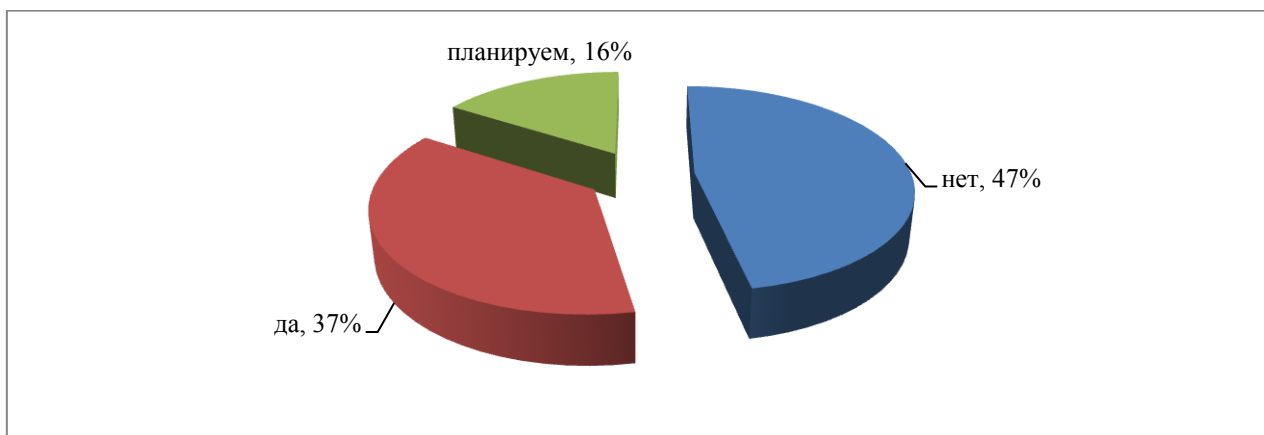
Большие Данные способны решать практически все ключевые задачи банков: привлечение клиентов, повышение качества услуг, оценка заемщиков, противодействие мошенничеству и др. Повышая скорость и качество формирования отчетности, увеличивая глубину анализа данных, участвуя в противодействии отмыванию незаконных средств, эти технологии помогают

банкам соответствовать требованиям регуляторов.

Основные задачи, для которых банки используют технологии анализа Больших Данных, – это оперативное получение отчетности, скоринг, недопущение проведения сомнительных операций, мошенничества и отмывания денег, а также персонализация предлагаемых клиентам банковских продуктов. Исследование CNews Analytics показало, что практика применения Больших Данных в передовых российских банках углубляется, хотя круг банков-инноваторов не становится шире. Основными причинами, тормозящими адаптацию технологий на российском рынке, эксперты называют сокращение ИТ-бюджетов на развитие и неочевидную эффективность больших данных.

В 2017 г. CNews Analytics провел опрос среди тридцати крупнейших российских банков по совокупным активам, чтобы узнать, какие технологии Больших Данных они применяют и с какими целями. По сравнению с опросом 2016 г., число банков топ-30, сообщивших о применении технологий Больших Данных, увеличилось, но это изменение связано скорее с изменением состава топ-30, чем с запуском новых проектов по анализу больших данных. Например, к тридцатке крупнейших присоединился банк «ФК Открытие», применявший эти технологии и ранее [32].

Характеристика использования технологии Больших Данных топ-30 российских банков представлено рисунком 6.



Источник:[32]

Рисунок 6 - Характеристика использования технологии Больших Данных ТОП-30 Российских банков

В 2017 г. представители 11 финансовых учреждений (37%) ответили, что их банк применяет современные технологии для анализа Больших Данных. В 5 банках (16%) планируют внедрять эти технологии, и число таких банков сократилось по сравнению с опросом 2016 г. Скорее всего это связано с общей тенденцией к сокращению ИТ-бюджетов на развитие в свете сложных экономических условий. В 14 крупнейших банках страны (47%) технологии не используются и не планируются к внедрению [32].

Крупные банки в ближайшие 3-5 лет будут наращивать экспертизу и проекты в области Больших Данных. Остальные игроки будут довольствоваться доступными SaaS-решениями, либо пытаться сделать что-то самостоятельно, но это не будет эффективно, так как профит маловат, а требуемые вложения в инфраструктуру и команду все еще очень высоки.

Использование технологий Больших Данных в ТОП-30 Российских банках представлено таблицей 4.

Таблица 4 - Использование технологий Больших Данных в ТОП-30 Российских банках

Применяют	Планируют к внедрению	Не применяют и не планируют
Сбербанк	ВТБ	Россельхозбанк
Газпромбанк	Юникредит Банк	Банк Москвы
ВТБ24	Нордеа Банк	НКЦ АКБ
Альфа-Банк	МДМ Банк	Росбанк
ФК Открытие		Московский Кредитный Банк
Промсвязьбанк		Россия
Райффайзенбанк		Санкт-Петербург
Русский Стандарт		Ак Барс
ХМБ Открытие		Уралсиб
Ситибанк		Связь-Банк
ХКФ Банк		СМП Банк
Тинькофф Банк		Банк Зенит
		Бинбанк
		Глобэксбанк

Источник: [32]

Текущая практика применения Больших Данных в российских банках находится на этапе становления.



Причины медленной адаптации Больших Данных в России видится в настороженном отношении ИТ-специалистов заказчиков к новым технологиям. Они не испытывают уверенности в том, что технологии Больших Данных помогут решать задачи в полном объеме. К тому же специалистов по этим технологиям на рынке явно недостаточно.

Оценка кредитоспособности клиента представляет собой неотъемлемый этап кредитного процесса. Важность этого процесса подтверждается высокой ценой ошибки выдачи кредита клиентам, неспособным вовремя и полностью выполнить условия кредитного договора. Для минимизации риска неплатежей кредитные организации используют и совершенствуют систему оценки кредитоспособности потенциальных заемщиков.

В российской банковской практике для оценки кредитоспособности юридических лиц используется метод коэффициентов, который дает разностороннюю оценку клиента: ликвидности, финансового левериджа, оборачиваемости капитала, прибыльности, обеспечения долга. Одним из методов оценки кредитоспособности физических лиц является скоринговые модели, которые представляют собой различные математические методы оценки благонадежности клиента. В основном такие модели применяются в «экспресс» - кредитовании и выдаче кредитных карт и оценивают клиента по ключевым характеристикам: его характер, финансовые возможности, обеспеченность кредита и условия предоставления кредита.

Одним из недостатков такой системы является тот факт, что оценка заемщику дается на основе сведений, содержащихся в предоставленной им анкете. При таком подходе до 25% заемщиков становятся некредитоспособными, например, студенты или амбициозные предприниматели без значительных средств и заработка. С развитием «финтеха» и внедрением систем по работе с «Большими Данными» данная проблема начинает находить решения.

Работа с Большими Данными позволяет провести более детальный анализ кредитоспособности клиента, увидеть не только его финансовый, но и

психологический портрет. Оценка направленности платежей и денежных переводов, личных качеств, профилей и активности в социальных сетях – все это позволяет уже сегодня финтех-компаниям предвидеть риски возможных неплатежей клиентов.

В целом, можно выделить несколько основных моментов, которые значительно отличают финтех - оценку кредитоспособности от традиционной банковской оценки:

1. Использование данных из социальных сетей. Сегодня социальные сети представляют собой ценный источник информации о заемщике: контакты, записи, фотографии, сообщества – по ним можно судить не только о интересах клиента, но и о его личных качествах. Некоторые финтех-стартапы прибегают к подтверждению кредитоспособности у друзей потенциального заемщика.

2. Использование мобильных данных. Периодичность и время звонков, история запросов в сети интернет, анализ мобильных оповещений – это лишь небольшая часть информации, которая используется финтех-стартапам для оценки кредитоспособности.

3. Агрегирование различных источников данных. Агрегирование позволяет проводить двойную проверку сведений, предоставленных заемщиком, например, использование данных из социальных сетей для проверки достоверности информации о карьере заемщика.

Очевидно, что оценка клиента по таким критериям традиционными методами может занять много времени. Для минимизации временных и финансовых затрат финтех-компаниями были изобретены роботы-оценщики. Они используют технологии искусственного интеллекта и адаптивного обучения, которые позволяют дать комплексную оценку заемщику, используя более тысячи фрагментов неструктурированных данных социальных сетей, мобильных данных и т.д.

Разработчики финтех-продуктов полагают, что при оценке традиционным способом банки упускают порядка 80% информации о клиенте, которая может быть использована для определения уровня его кредитного риска. Что не менее

важно, такого рода комплексный анализ может и должен использоваться кредитными организациями для непрерывного мониторинга клиента, даже после выдачи кредита. Например, он поможет понять для чего клиент хочет взять новый кредит: для выплаты процентов по-старому или для расширения бизнеса.

Одним из ярких и наиболее успешных финтех-стартапов, является проект Airе, миссией которого является помощь в присвоении кредитного рейтинга тем, кого традиционные скоринговые методы считают некредитоспособными. Особенностью проекта является возможность заемщиков следить за своим текущим уровнем кредитоспособности.

Использование технологий работы с «Большими Данными» при оценке кредитоспособности имеет место и в России. Однако в отличие от стран запада, где финтех развивается за счет вновь образующихся стартапов, в России финтех растет за счет ведущих банков: «Сбербанк России», «Тинькофф банк» и «Альфа-банк». Кроме того, банки ВТБ24, «Хоум Кредит» вместе с другими участниками финансового рынка и международной платежной системой MasterCard создали первый в России отраслевой финтех-акселератор «Финтех Лаб».

На предварительных обсуждениях наибольшей поддержкой пользовались темы, связанные с идентификацией и биометрией, роботизацией, машинным обучением и маркетингом, основанным на данных. «Тинькофф Банк» запустил собственную школу разработки и аналитики в сфере финтеха «Tinkoff Fintech School».

Тем не менее, анализ позволяет сделать вывод: финтех в России в основном развивается в сфере платежей и онлайн-банкинга. Разработка финтех-проектов оценки кредитоспособности представлена только небольшим числом проектов. Среди них совместный проект Mail.ru Group и бюро кредитных историй «Эквифакс». Данный проект использует систему искусственного интеллекта для оценки уровня риска заемщиков, признанных нежелательными традиционными методами. Mail.ru Group уже имеет клиентов, пользующихся

новым сервисом, но их имена и стоимость услуги не раскрываются.

Другой российский финтех-стартап FscoreLab запустил свой пилотный проект скоринга, основанный на анализе фотографии заемщика. Сервис использует нейросеть, обученную на 600 тысячах реальных случаев по выплате кредитов, взятых из баз Бюро кредитных историй (БКИ).

Программа изучила фотографии людей с просроченными кредитами и самостоятельно выявила черты, свойственные лицам должников. Тем не менее, многие эксперты настроены скептически к таким проектам, подчеркивая, что определение кредитоспособности по одному фото может рассматриваться лишь как дополнительный инструмент при формировании кредитного рейтинга.

При всей привлекательности финтех-проектов, интерес к ним со стороны инвесторов начинает затухать.

По данным CB Insights, в 2017 г. венчурные инвестиции в финтех-стартапы во всем мире сократились по сравнению с зафиксированным годом ранее максимумом на 13% и составили \$12,7 млрд. Количество сделок также уменьшилось на 1%. В России же мнения инвесторов разделились: одни считают, что финтех представляет собой очередной финансовый пузырь, другие – что развитию финтеха мешает жесткая политика мегарегулятора, третьи полагают, что в России нет преград для развития новых методов оценки кредитоспособности уже сегодня [34].

В заключение, хотелось бы отметить, что развитие альтернативных систем оценки кредитоспособности клиентов позволит, с одной стороны, банкам расширить активные операции, а с другой стороны, откроет доступ к финансовым ресурсам тысячам нуждающимся в них клиентам. Особенностью развития российских проектов является их создание не как новых стартапов, а как проектов внутри существующих кредитных организаций.

Несмотря на наличие проблем, связанных с недостатком инвестиций в подобные проекты, развитие финтех-проектов в России будет набирать скорость, поскольку связано с объективным желанием кредитных организаций получить всестороннюю оценку кредитоспособности клиента и

минимизировать потери по ссудам.

Монетизация технологий больших данных – до сих пор открытый вопрос.

Перспективно выглядят технологии по прескорингу клиентов на основании данных из социальных сетей и иной информации, например, истории поисков и посещений сайтов клиентами.

В Росбанке сейчас сделан фокус на развитии корпоративного хранилища данных, повышении качества информации о клиентах и продуктах – своего рода фундаменте с точки зрения аналитики.

На следующем этапе банк планирует перейти к эффективному использованию имеющейся клиентской информации для получения объемного (360°) видения клиента для персонализации услуг.

Ежедневно в мире производится и накапливается все больше данных. По оценке McKinsey Global Institute, их прирост составляет 40% в год, то есть с 2010 года по 2020 год общемировой объем данных увеличится в 40 раз.

Аналитики Gartner считают, что уже в 2017 году более 90% бизнес-руководителей будут рассматривать информацию как стратегический актив. К сожалению, из-за недостатка нужных знаний или инструментов оценить реальный экономический эффект от ее использования смогут не более 10% из них [34].

В этой ситуации банки не являются исключением: объем данных, которые они накапливают, уже сейчас довольно велик. В скором времени поток информации станет таким обильным, что стандартные технологии уже не смогут с ним справиться, и финансовым институтам надо быть к этому готовыми.

Классические реляционные СУБД, используемые для аналитических задач, имеют известные ограничения. При изучении клиентской базы риск-аналитики вынуждены использовать «репрезентативные» выборки, как правило, не превышающие 1% от общего числа клиентов. Кроме того, оценка динамики за длительные периоды времени требует значительных вычислительных ресурсов.

В итоге возникают две серьезные проблемы: огромный объем разнородных данных и низкая производительность аналитических систем. Простое наращивание мощностей уже не способно их решить. Поэтому для многих задач целесообразно использовать технологии массово-параллельных вычислений, которые обозначают термином Big Data - «большие данные».

Суть технологии - в распределении обработки данных по множеству независимых взаимозаменяемых узлов (серверов). Это позволяет выполнять задачу по частям и дает необходимый уровень производительности, устойчивости и масштабируемости. Кроме того, технологии «Больших Данных» оптимизируют загрузку данных, а также обеспечивают оперативный доступ к ним.

### 2.3 Перспективы использования Больших Данных в ПАО «Сбербанк России»

ПАО «Сбербанк России» провел большую работу по обеспечению высокой надежности своих систем. В 2017 году решаются задачи с применением технологий Больших Данных (таблица 5).

Таблица 5 - Решение задач ПАО «Сбербанк России» с применением технологий Больших Данных

Название банка	Технологии, используемые для работы с большими данными	Задачи, решаемые с применением технологий для обработки больших данных
Сбербанк	Teradata, Cloudera Hadoop, Impala, Zettaset, стек продуктов Apache (Hadoop, HBase, Hive, Mahout, Oozie, Zookeeper, Flume, Solr, Spark и пр.), специализированные базы данных (Neo4j, MongoDB и т.д.), различные аналитические инструменты, собственные решения в области data minig, predictive/prescriptive-аналититки, обработки текстов на естественном языке.	<ul style="list-style-type: none"> <li>- Управление рисками,</li> <li>- противодействие мошенничеству,</li> <li>- сегментация клиентов,</li> <li>- оптимизация эквайринговой сети,</li> <li>- управление оттоком,</li> <li>- расчет бонусов для сотрудников массового сегмента,</li> <li>- технологии по работе с большими данными рассматриваются к применению в транзакционных и расчетных системах банка.</li> </ul>

Источник: [57]

Преимущественно, задачами с применением технологий Больших Данных является: быстрое формирование финансовой отчетности, управление взаимоотношениями с клиентами, персонализация услуг, борьба с мошенничеством, кредитный скоринг и управление оттоком клиентов.

Среди важных вех этой работы – организация георезервирования сервисов контактного центра ПАО «Сбербанк России»; создание ядра новой высоконадежной локальной вычислительной сети; работа клиентских сервисов при совершении операций в интернет-магазинах, переводах, выдаче кредитов, обслуживании через удаленные каналы в режиме Stand-In 24 × 7 в период инцидентов и технологических работ.

Простои критичных автоматизированных систем ЦОД «Южный порт» не превышают 1,6 часа в год. Данный ЦОД сертифицирован по программе Tier Certification Operational Sustainability, Uptime Institute, уровень GOLD.

Высококритичные сервисы транспортирования данных между автоматизированными системами ПАО «Сбербанк России» переведены в режим функционирования 99,999 %, то есть простой системы составляет не более 5 минут в год. Это обеспечивает непрерывность предоставления основных услуг частным и корпоративным клиентам [57].

В системе «Сбербанк Онлайн» выделен пилотный блок для сотрудников, в котором происходит тестирование новых версий Сбербанк Онлайн до масштабного тиражирования, что минимизирует риски и сокращает сроки внедрения.

В ПАО «Сбербанк России» внедрены сквозной производственный процесс и ресурсное планирование, благодаря чему усилился контроль за запуском и реализацией проектов, сократилась средняя длительность проектов с 30 до 18 месяцев. Новый процесс реализации непроежных задач позволил сократить срок их внедрения в 1,9 раза. Выросла удовлетворенность внутренних клиентов, которая в области реализации ИТ-составляющей проектов выросла в 3,8 раза, в области реализации непроежных задач – в 3 раза. ПАО «Сбербанк России» завершил трансформацию ИТ-организации.

Создана платформа для технологической трансформации.

В ПАО «Сбербанк России» началась Agile-трансформация, которая заключается в переходе на метод гибкой разработки, получившей название Sbergile. Sbergile-команды обеспечены базовой автоматизацией, разработан процесс итеративной разработки сервисов [57].

В ПАО «Сбербанк России» создан единый процесс управления операционным и ИТ-производством, инцидентами и технологическими стандартами. На 13 % сокращена численность функции сопровождения клиентских операций. Трансформированы региональные центры сопровождения клиентских операций в г. Хабаровске и Воронеже. Сопровождение ИТ-операций обеспечено во всех часовых поясах.

Внедрена программа «Платформа поддержки развития бизнеса (18+)» Платформа призвана стать универсальным конструктором для создания бизнес-приложений.

Практически подтверждена производительность и масштабируемость In-Memory Data Grid архитектуры, в частности достигнута высокая производительность в 35 тыс. транзакций в секунду.

Создано единое информационное пространство, куда успешно загружены данные по 100 млн клиентов. Разработаны механизмы аудита, авторизации, доступа к данным и их пакетной обработки.

Внедрены важнейшие сервисы для бизнеса: единый профиль клиента Розничного блока, единый каталог продуктов и тарифов в части вкладов и банковских карт, динамическое ценообразование. Запущены первые продуктовые фабрики: переводов P2P, торгового эквайринга, вкладов. Команда Программы получила статус разработчиков open-source-сообщества Apache Software Foundation. Проекты Программы получили возможность развивать open-source-компоненты технологического стека платформ.

Внедрена программа «Единая фронтальная система» Цель Программы – создать единый стандарт во всех каналах обслуживания клиентов. Основной акцент Программы в 2017 году был сделан на росте активных продаж частным



клиентам через контактный центр, повышении лояльности корпоративных клиентов за счет сервиса удаленного резервирования счета без визита в офис ПАО «Сбербанк России», снижении стоимости услуг внешних контактных центров корпоративных клиентов.

С технической стороны для этого была создана единая библиотека интерфейсных компонентов базовых системных сервисов, которые используются для создания пользовательского интерфейса. Использование библиотеки позволяет повысить на 30–35 % скорость разработки экранных форм и снизить стоимость их разработки на 15–20 % [57].

Разработан ряд open-source-компонент, которые представлены для переиспользования в свободный доступ интернет-сообществу. Внедрен конвейер автоматической сборки приложений, и пилотируется технология автоматического развертывания системы на все среды.

Использование технологии DevOps приведет к существенному снижению time-to-market и позволит в разы быстрее выводить продукты на рынок. Функционал дистанционного открытия счетов, зарплатных проектов, корпоративных карт перенесен на новую цифровую корпоративную платформу. Это является первым шагом на пути к переходу на Единую фронтальную систему.

Создано мобильное рабочее место агента прямых продаж, что позволит планировать встречи и оптимизировать маршруты передвижения с учетом географического расположения клиентов. Программа полностью реализуется по методу Agile. От идеи до открытия проходит восемь недель. По Программе работают более 90 Agile-команд.

В 2017 году удалось сформировать лучшую команду ИТ-специалистов и бизнес-экспертов. Команда насчитывает более 1 тыс. сотрудников из бизнес-блоков Сбербанка и 17 центров компетенции «Сбербанк-Технологии». Для привлечения лучших специалистов Сбербанк провел день открытых дверей и Международный дизайн-хакатон [57].

Внедрена программа «Фабрика данных». Цель Программы – обеспечить

Группе условия для достижения конкурентной скорости вывода на рынок новых продуктов, монетизации данных, повышения скорости принятия управленческих решений, снижения стоимости владения данными. Программа объединила активности по созданию data-сервисов и развитию инфраструктуры с учетом актуальных тенденций в построении корпоративных хранилищ данных и аналитических платформ.

Ключевые проекты Программы:

- профиль клиента «4D» – повышает полноту информации и глубину истории о корпоративном клиенте;
- «Массовая персонализация» – повышает эффективность одноименных процессов розничного бизнеса за счет быстрого получения достоверных сведений о клиентах на основе данных;
- «Бутиковый конвейер» – увеличивает доход от клиентов СІВ за счет сокращения сроков и повышения эффективности принятия решений в части информации о клиентах;
- проект «Геомаркетинг 2.0» – предоставляет внешним клиентам ПАО «Сбербанк России» информацию об экономическом потенциале отдельных географических локаций.

В рамках Программы увеличена производительность аналитического хранилища данных. Создан новый важнейший элемент архитектуры – облако данных – это распределенное хранилище данных для последующей обработки, куда загружены первые данные крупнейших систем ПАО «Сбербанк России» – Единая корпоративная система и Единый кредитный портфель. Запущена область экспериментов с данными и проверки гипотез моделей для бизнес-пользователей. ПАО «Сбербанк России» удалось сократить до 10 дней время разовой поставки данных по запросам подразделений Сбербанка (ранее срок составлял более четырех месяцев).

Внедрена программа «Централизация 3.0». Цель Программы – завершить централизацию ландшафта, существенно повысив экономическую эффективность ИТ-активов.

В 2017 году в рамках Программы выведены из эксплуатации 682 нецелевые автоматизированные системы (при плане 410) и два ЦОД. В 2018 году планируется вывести еще 270 нецелевых систем и семь центров обработки данных и заменить ИТ-оборудование.

Основой изменяющегося ИТ-ландшафта ПАО «Сбербанк России» должна стать технологическая платформа, которая будет выступать средой, обеспечивающей функционирование бизнеса и позволяющей участникам экосистемы взаимодействовать и создавать ценность.

Технологическая платформа будет включать в себя инфраструктуру, данные и средства их обработки и анализа, приложения, средства разработки, API.

ПАО «Сбербанк России» активно применяет новейшие технологии для создания инновационных сервисов и услуг. В частности, методы работы с Big Data использованы в разработке универсальной чат-платформы для мессенджеров Telegram и Facebook.

Проведен пилот системы биометрической идентификации на основе сетчатки глаза для устройств самообслуживания. При запуске проекта «Оплата по биометрии» в торговой сети «Азбука вкуса» было проведено закрытое тестирование для сотрудников Сбербанка по отпечатку пальца.

Реализован пилот по разработке математической модели управления инкассацией и наличностью в устройствах самообслуживания ПАО «Сбербанк России», который призван сократить простой устройств и сэкономить на оптимизации хранимых объемов наличных средств.

Успешно завершена начатая 4,5 года назад Программа «Автоматизация систем управления рисками на финансовых рынках».

Созданы системы для контроля лимитов рыночного и кредитного рисков, установления лимитов на финансовые институты. Внедрены системы контроля рыночности и независимой верификации цен. Уникальность этой ИТ-программы для России в том, что до сих пор не было положительного опыта решения аналогичных задач.

В итоге при создании риск-инфраструктуры было сэкономлено около 360 млн рублей. ПАО «Сбербанк России» внедрил автоматизированную систему мониторинга кредитных заявок для выявления мошеннических действий со стороны корпоративных клиентов. Система обрабатывает большие массивы данных In-Memory, что позволяет производить проверки в режиме реального времени. ПАО «Сбербанк России» автоматизировал систему контроля качества андеррайтинга.

Запланированное до конца 2018 года создание инструментария анализа рисков позволит улучшить качество кредитного портфеля и обеспечить поддержание высокого уровня экспертизы.

Успешно завершён пилот по развёртыванию международного карточного процессингового центра. Первым клиентом стал «БПС-Сбербанк» Республики Беларусь.

Сбербанк ввёл систему, автоматизирующую сбор просроченной задолженности на этапе позднего сбора, в том числе с использованием мобильного приложения коллектора.

ПАО «Сбербанк России» предложил клиентам «Личный кабинет инвестора», который дал им доступ к просмотру остатков по брокерским счетам, маржинальным показателям, реестру сделок, информации по вводу/выводу денежных средств, движению ценных бумаг между счетами клиента.

База данных СМС-сервиса «Мобильный банк» переведена на новый высокопроизводительный ресурс для выдерживания нагрузки в 18 млн входящих СМС от клиентов в сутки.

Создан прототип портала для внешних партнеров ПАО «Сбербанк России», опубликован сервис с открытым интерфейсом (API) и развернут инструмент управления жизненным циклом API для подключения партнеров «Сбербанк-Мессенджер», а также публикации API корпоративных сервисов и привлечения новых партнеров.

Создан новый канал обработки мультимедиа-сообщений в контактном центре – «Текстовый чат», что сократит расходы на оплату телефонного

трафика и снизит темпы роста нагрузки на операторов.

ПАО «Сбербанк России» внедрил у себя технологию «прямых расчетов», что позволило проводить платежи в расчетной системе Сбербанка в режиме реального времени. Таким образом, время прохождения платежа между клиентами ПАО «Сбербанк России» было сокращено с 45 до 6 минут.

ПАО «Сбербанк России» ввел упрощенную схему рассмотрения обращений корпоративных клиентов, в результате чего до 50 % финансовых претензий решаются за один день, обращения по самоинкассации решаются в течение часа. Максимальное время рассмотрения обращений корпоративных клиентов не превышает двух дней. Создана Политика по управлению качеством данных, разработана уникальная для российского рынка Программа обучения по работе с данными, которую первыми прошли 80 топ-менеджеров ПАО «Сбербанк России».

В ПАО «Сбербанк России» появился новый бизнес – монетизация данных. Сформирован базовый портфель продуктов Сбербанка на основе агрегированных данных.

Заключены первые сделки с внешними клиентами на продукты по построению модели склонности к покупке и сегментированию клиентской базы для таргетированных кампаний.

Запущен портал «Открытые данные» – уникальный информационный продукт на основе технологий Big Data, который представляет агрегированные данные экономической активности населения и бизнеса.

В ПАО «Сбербанк России» созданы лаборатории инноваций, которые на текущий момент имеют средний срок разработки прототипа инновационного продукта не более пяти месяцев.

Среди инновационных инициатив, которые ПАО «Сбербанк России» тестирует или внедряет, можно назвать следующие:

– учет и управление доверенностями, электронными закладными на недвижимость, денежными переводами, учет факторинговых сделок на базе технологии блокчейн;

- построение совместно с Федеральной антимонопольной службой системы электронного документооборота на базе технологии блокчейн;
- автоматизированное построение графиков работы сотрудников в офисах обслуживания клиентов;
- использование технологии Platform as a Service (PaaS) и внедрение нового продукта с минимальным ценным для клиента функционалом (MVP), что будет сокращать время вывода на рынок новых продуктов Сбербанка;
- использование технологии автоматического общения с клиентами в текстовых каналах обслуживания (мессенджеры и СМС), создание интеллектуальных цифровых помощников для решения проблем клиентов и универсальной платформы для чат-ботов;
- создание мобильного банковского приложения нового поколения, основанного на технологии мессенджера; создание экосистемы общения клиентов, простых каналов коммуникации, управления счетами и денежными средствами, переводами; исследование маркетплейсов для партнеров, оказание дополнительных услуг клиентам;
- исследование востребованности сервисов, основанных на предоставлении реальных API ПАО «Сбербанк России» для разработки приложений внешними разработчиками; цель – создать сообщество, использующее платформу ПАО «Сбербанк России» для развития внешних продуктов;
- разработка корпоративного файлового хранилища, при котором Сбербанк сможет хранить все рабочие документы в своей облачной среде, предоставлять доступ к ним с рабочих станций и мобильных устройств всем своим работникам, выполняя требования безопасности.

В качестве ключевых планов развития технологического ландшафта на 2018 год ПАО «Сбербанк России» наметил реализацию всех базовых технологических сервисов ИТ-платформы и бизнес-сервисов для частных клиентов, развитие экосистемы технологического партнерства Open API,

внедрение алгоритмов искусственного интеллекта и машинного обучения для задач Сбербанка, подключение не менее двух бизнесов экосистемы.

ПАО «Сбербанк России» запустил информационный продукт на основе технологии работы с большим объемом данных (Big Data), который позволит желающим следить за агрегированными данными финансовой активности клиентов банка на сайте финансовой организации.

ПАО «Сбербанк России» будет на регулярной основе публиковать такие данные, как выручка юридических лиц в разрезе отраслей и регионов, уровень доходов населения (зарплаты, пенсии, стипендии, пособия), склонность к сбережениям и потреблению.

В ПАО «Сбербанк России» считают, что такие данные будут интересны бизнес-аналитикам, маркетологам, кредитным аналитикам, риск-менеджерам, студентам и преподавателям (экономистам, политологам и социологам), журналистам, пишущим на экономические и общественно-политические темы, составителям обзоров, отчетов и аналитических записок.

Проект «Открытые данные» использует существующую ИТ-инфраструктуру банка, в том числе продукты открытого программного обеспечения. Решение на основе технологии Big Data, которое позволяет в открытом доступе представить агрегированные данные по финансовой активности клиентов, было создано внутренними ресурсами Сбербанка.

Банковский сектор остается на передовой цифровой трансформации бизнеса, и текущая рыночная ситуация, с одной стороны, и растущая конкуренция, с другой, заставляют его искать новые подходы к управлению активами, повышению эффективности своей работы и поиску новых точек роста. Безусловно, технологии Big Data - это крайне перспективное направление комплексного анализа самой разнообразной информации, инструмент для проведения многофакторного исследования банковских данных.

Применение такого подхода финансовыми организациями призвано решать сразу несколько важных задач: во-первых, оперативно и точно

оценивать кредитные риски, не допускать мошеннических действий, ну и, конечно, увеличивать объемы продаж, попадая четко в цель с персонализированными предложениями банковских услуг клиентам.

Эффект от применения Больших Данных проявляется в разработке новых бизнес-моделей, повышении эффективности работы сотрудников и продуктивности взаимоотношений с клиентами, оптимизации операционной деятельности, новых возможностях управления рисками и финансами.

Именно сейчас технология достигла высокого уровня развития, опираясь на алгоритмы обработки Больших Данных. За микросекунды принимаются решения, что продать и что купить, исходя из анализа сделок на рынках, которых происходит за день миллионы. Это настоящие Большие Данные.

Что касается отношений с клиентами, то в банках Большие Данные используются для целевого маркетинга и удержания клиентов. Целевой маркетинг важен скорее для потребительского банка, чем для инвестиционного.

Потребительский банк с помощью алгоритмов обработки Больших Данных повышает точность определения целевых групп клиентов для тех или иных продуктов, например, образовательных кредитов. Но удержать клиентов важно и для инвестиционного банка – Большие Данные дают возможность лучше понимать специфику работы корпоративных клиентов, их отношение к различным банковским сервисам и т. д.

В области управления рисками и финансами для инвестиционного банка важны алгоритмы вычисления риска дефолта контрагента. Когда банк делает хеджирование или принимает решение о предоставлении кредита корпоративному клиенту, необходимо четко понимать риск дефолта и осмысленно определять размер кредита и процент по кредиту. И еще одна область применения Больших Данных в банковском бизнесе – прогнозы движения рынка. Предсказать будущее на 100% вряд ли можно, но Большие Данные помогают лучше понимать тенденции и принимать наиболее подходящие меры.



Тема Big Data, которая недавно была модным трендом для технологических лидеров вроде Amazon и Google, сегодня находит широкое применение в России, и не в последнюю очередь в банках. Особенно это заметно по тем банкам, для которых главный приоритет — клиентоориентированность и которые хотят привлечь интернет-активных (и прибыльных) потребителей. Особенно примечателен интерес к этой технологии в условиях кризиса, когда качество кредитных портфелей падает, а клиенты настроены сменить банк, если получают более привлекательное приложение.

В условиях Big Data дает инструменты, чтобы собирать информацию о клиенте, автоматически формировать для него предложения исходя из их индивидуальных потребностей и ограничений, а также использовать эту информацию для точной оценки платежеспособности человека [38].

По результатам опроса того же McKinsey, 76% банков заявляют, что Big Data позволяют привлекать новых клиентов, лучше взаимодействовать с ними и поддерживать их лояльность. Можно с уверенностью утверждать, что более половины аналитических работ выполняется по заказу маркетологов. Глубокая сегментация клиентской базы, поиск неявных закономерностей, персонализация продуктовых предложений — вот основные задачи по этому направлению, которые в крупных розничных банках решаются с помощью Big Data [32].

Банкам нужны новые источники информации, позволяющие посмотреть на клиента шире, формировать новые уникальные предложения.

Стратегия развития каналов также зачастую базируется на результатах аналитики Big Data. Вдобавок к стандартным возможностям анализирует отзывы в соцсетях (Facebook, Twitter, VKontakte), в «Народном рейтинге» на портале «Банки.ру» и на площадках, на которых размещены приложения банка. На базе анализа внутренней и внешней информации планируется развитие новых и модернизация уже имеющихся клиентских сервисов.

Еще один аспект работы с клиентами — идентификация подозрительных транзакций, и тут роль глубокой аналитики переоценить невозможно. Об этом

много говорилось, проблема организации эффективной борьбы с мошенничеством присуща не только финансовому сектору. Профильные решения класса Data Mining не новы для рынка, но тут задача должна решаться оперативно, и потому используются решения категории in-memory — вычисления в оперативной памяти.

Например, банк HSBC внедрил технологии больших данных в состав решения для противодействия мошенничеству с кредитными картами. В результате эффективность службы по выявлению случаев мошенничества повысилась в три раза, а точность его выявления — в десять раз. За первые же две недели эксплуатации семь специалистов службы безопасности HSBC выявили новые криминальные группы и схемы с общим потенциальным ущербом более \$10 млн.

Колоссальные перспективы раскрывает перед банком сочетание Big Data и геоаналитики. Например, клиент заявил о потере карты и одновременно по этой же карте в магазине совершена покупка. А банк оперативно идентифицировал местонахождение телефона клиента и может доказать, что клиент пытается смошенничать (если телефон и карта находятся рядом). Или заблокировать транзакцию, не дожидаясь заявления клиента о потере карты (если телефон и карта находятся на приличном расстоянии друг от друга).

Вендоры, ИТ-консультанты, говоря о Big Data, зачастую имеют в виду решение новых задач, чуть ли не открытие новых горизонтов. Конечно, большим объемам информации присущи свои законы. Однако, если можно так выразиться, на бытовом уровне банкам в первую очередь важны оперативность и соблюдение заданных сроков.

Из всего колоссального накопленного количества информации, по различным оценкам участников рынка банковской автоматизации, в жизни организации используется в лучшем случае половина информации. И это самый оптимистичный вариант. Порой КПД составляет единицы процентов. Поэтому вопрос монетизации хранения данных так же вечен для банков, как проблема загрязнения окружающей среды - для промышленных мегаполисов.

И только технологии Big Data могут обеспечить экономическое обоснование этого хранения, превращая «чулан» в эффективно организованную «кладовую». Ведь без Big Data извлечь ценность из накопленного богатства информации невозможно.

Преимущественно, задачами с применением технологий Больших Данных является: быстрое формирование финансовой отчетности, управление взаимоотношениями с клиентами, персонализация услуг, борьба с мошенничеством, кредитный скоринг и управление оттоком клиентов.

Управление данными и их анализ дают любой организации, независимо от ее размера и области деятельности, значительные преимущества и открывают большие перспективы.

Направление по применению Big Data в работе банка, в сфере сбора просроченной задолженности считается перспективным так, как для сбора просроченной задолженности необходим контакт с должником. Для установления связи с клиентом, нужна актуальная контактная информация.

Благодаря технологии Big Data становится возможным обрабатывать не только те данные, которые есть у банка для поиска подобной информации, но искать и контактную информацию и в других местах. Например, в социальных сетях, форумах и других ресурсах, где клиент мог бы оставить свои актуальные данные. Для обработки таких больших объемов информации и требуется технологии Big Data.

## Заключение

Большие Данные (англ. Big Data) - серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста информации.

Это огромные массивы данных, которые на порядок больше (объем); разнообразнее, включая в себя структурированные, полуструктурированные и неструктурированные данные, (диверсификация) и быстрее (скорость передачи и обработки). Этот поток данных создается подключенными устройствами - от ПК и смартфонов до датчиков, например, устройств считывания RFID и уличных камер. Кроме того, эти данные гетерогенны и передаются в различных форматах: текст, документы, изображения, видео и многое другое.

Управление данными и их анализ дают любой организации, независимо от ее размера и области деятельности, значительные преимущества и открывают большие перспективы.

Направление по применению Big Data в работе банка, в сфере сбора просроченной задолженности считается перспективным так, как для сбора просроченной задолженности необходим контакт с должником. Для установления связи с клиентом, нужна актуальная контактная информация.

Благодаря технологии Big Data становится возможным обрабатывать не только те данные, которые есть у банка для поиска подобной информации, но искать и контактную информацию и в других местах. Например, в социальных сетях, форумах и других ресурсах, где клиент мог бы оставить свои актуальные данные. Для обработки таких больших объемов информации и требуется технологии Big Data.

Более того информация в социальных сетях может быть представлена и в формате текста или даже фото или видео. Следовательно, переход на работу с большими объемами данных как никогда актуален в данной сфере. Для поиска

и обработки данных клиентов в социальных сетях, базах данных мобильных операторов как раз и необходима Big Data.

Работа с Большими Данными позволяет провести более детальный анализ кредитоспособности клиента, увидеть не только его финансовый, но и психологический портрет. Оценка направленности платежей и денежных переводов, личных качеств, профилей и активности в социальных сетях – все это позволяет уже сегодня финтех-компаниям предвидеть риски возможных неплатежей клиентов.

ПАО «Сбербанк России» провел большую работу по обеспечению высокой надежности своих систем. В 2017 году решаются задачи с применением технологий Больших Данных.

Внедрены важнейшие сервисы для бизнеса: единый профиль клиента Розничного блока, единый каталог продуктов и тарифов в части вкладов и банковских карт, динамическое ценообразование. Запущены первые продуктовые фабрики: переводов P2P, торгового эквайринга, вкладов. Команда Программы получила статус разработчиков open-source-сообщества Apache Software Foundation. Проекты Программы получили возможность развивать open-source-компоненты технологического стека платформ.

Рынок Больших данных (Big Data) вызывает все больший интерес коммерческих и государственных структур по всему миру. Огромные объемы накопленной информации позволяют проводить качественные оценочные и прогнозные работы.

Глобальный рынок Big Data вышел на уровень активного роста. Широкий спектр предложений подталкивает компании к поиску новых решений и оптимизации существующих.

По информации International Data Corporation (IDC), объем хранимых данных в мире с 2010 года увеличился в 50 раз. К 2020 году он вырастет еще в 15 раз и достигнет отметки в 40 ZB (зеттабайт). Рост в основном будет происходить за счет развивающихся рынков.

Эта тенденция неизбежно подтолкнет руководство банков к инвестициям

в IT-инфраструктуру: системы хранения, управления и защиты информации, а также оборудование, коммуникации, персонал.

Без выработки собственной IT-стратегии и стратегии управления данными, без выхода на рынок онлайн-услуг, без создания доверительных и долгосрочных отношений с клиентами, банки будут терять свой бизнес, постепенно превращаясь в бек-офисы социальных сетей и сервисных IT-компаний.

Первые будут эффективнее сегментировать клиентов, определять их предпочтения, предоставлять более удобный сервис, отбирая часть маржинального дохода у банков. Вторые будут эффективнее автоматизировать фронт - и бек-офисные процессы, предоставлять более удобные электронные каналы обслуживания и обработки транзакций клиентов. Это неизбежно снизит доходность банковского бизнеса. Большую часть своей маржи банки будут отдавать своим IT-подрядчикам. При этом зависимость от IT-партнеров будет только возрастать.

Выбор надежного IT-партнера и обеспечение эффективного корпоративного контроля над ним станет ключевым элементом IT-стратегии для большинства банков в ближайшие годы. Руководству банков неизбежно придется решать вопросы IT-стратегии: от модели бизнеса до модели данных банка, от модели данных до архитектуры приложений, от архитектуры приложений до управления и защиты информации, а также до оборудования, коммуникаций и удержания квалифицированного IT-персонала.

Появление новых технологий сблизило возможности использования «профилей клиентов» в банковском, телекоммуникационном и розничном секторе, где применяются схожие подходы: комбинируя и обогащая информацию, можно построить богатый поведенческий профиль клиента.

Если раньше профиль клиента включал в себя два-три десятка характеристик, то теперь их может быть значительно больше. Чтобы извлечь из этого выгоду, нужна накопленная статистика. Подобные методы уже используются при скоринге и оценке клиентов.

Новые подходы могут помочь в повышении уровня лояльности клиентов, сделать сервисы более удобными для них. По спектру услуг универсальные банки практически сравнялись. Клиенты смотрят на то, какой банк более удобен, технологичен, им важен персональный подход, когда, например, при звонке в колл-центр по номеру телефона, который привязан к карте, идентифицируется входящий звонок, и доступно больше информации по клиенту. Это способствует персонификации, лояльности, увеличению «жизненного цикла» пребывания клиентов в банке.

Банк планирует использовать сбор информации о клиенте из разных источников, анализ неструктурированных данных для формирования целевого предложения для клиента.

Технологии больших данных необходимо применять не только для анализа клиентской среды, но для управления рисками, взаимоотношениями с клиентами, использовать в маркетинге, борьбе с мошенничеством.

Применение этих технологий в финансовом секторе имеет большой потенциал, поэтому направление Big Data должно развиваться и далее.

Комплексы Oracle Big Data Appliance позволяют по-новому использовать данные, чтобы предоставлять клиентам услуги, не реализуемые ранее в силу технологических или экономических ограничений.

Общий подход Oracle при создании оптимизированных программно-аппаратных комплексов заключается в том, что за счет совместной разработки всех компонентов корпорация старается достичь максимальной производительности, а использование стандартных, хорошо отлаженных конфигураций повышает надежность комплекса.

Сочетая новые источники данных с аналитикой реального времени и поведенческой информацией, компании смогут разработать новое поколение приложений, способных оперативно адаптироваться и обучаться. Готовые комплексные решения, интегрирующие такие сервисы, как аналитика, исследование, подготовка и интеграция данных, упрощают разработку новых продуктов.

## Список использованных источников

1. Аналитический обзор рынка Big Data [Электронный ресурс]. - Электрон. дан. - Режим доступа: <https://habrahabr.ru/company/moex/blog/256747/>
2. Бачило, И. Л. Персональные данные в структуре информационных ресурсов. Основы правового регулирования./ И. Л. Бачило. - Минск: Беллитфонд, 2017. - 474 с.
3. Биктимиров, М. Р. Тенденции развития технологий обработки Больших данных и инструментария хранения разноформатных данных и аналитики. / М. Р. Биктимиров. // Future Banking. - 20.06.2017.
4. Блинов, Н. Банки и большие данные - знать о клиенте больше, чем он знает о себе сам. / Н. Блинов. // Future Banking. - 21.05.2017.
5. Бодров, А. А. Современные технологии анализа больших данных: новая философия знаний. / А. А. / Бодров. // Фундаментальные исследования. – 2015. – № 2-23. – С. 5295-5299.
6. Большие данные. Большие возможности [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://www.emc.com/ruru/big-data/index/>
7. Будзко, В. И. Системы высокой доступности и Большие Данные. / В.И. Будзко. // Большие данные в национальной экономике. - 2017. - №8. - С. 16. – С.45.
8. Будзко, В. И. BigData. Новый вызов [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://www.myshared.ru/>
9. Васильков, А. Data Collective: «большие данные» — приоритетное направление инвестиций./ А. Васильков [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://www.computerra.ru/97457/data-collective-relieson-big-data/>
10. Вахрамеев, К. СУБД для анализа Больших Данных. / К. Вахрамеев. // Открытые системы. СУБД. – 2017. – № 10.
11. Веретенников, А. В. BigData: анализ больших данных сегодня. / А.В. Веретенников. // Молодой ученый. - 2017. - №32.- С. 9-12.



12. Войниканис, Е. Право интеллектуальной собственности в цифровую эпоху. Парадигма баланса и гибкости. / Е. Войниканис. - М.: Юриспруденция, 2016. - 552 с.

13. Волкова, Ю. С. Большие Данные в современном мире. / Ю. С. Волкова. // Научно-методический электронный журнал «Концепт». – 2018. – №1. – С. 1171–1175.

14. Вуколов, Э. А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие для студентов вузов. / Э.С. Вуколов. – М.: ФОРУМ: ИНФРА-М, 2017. – 464 с

15. Галлини, Н. И. Информационная справочная система Yaltavernadainfocenter. / Н. И. Галлини. // Перспективы науки. Материалы I Международного заочного конкурса научно-исследовательских работ. Научно-образовательный центр «Знание». – Казань: ООО «Рукета Союз», 2017. – 212 с.

16. Галлини, Н. И. Проектирование информационной системы анализа и мониторинга показателей контингента обучающихся в организации высшего образования. / Н. И. Галлини. // Перспективы науки. - Материалы I Международного заочного конкурса научно-исследовательских работ. Научно-образовательный центр «Знание». – Казань: ООО «Рукета Союз», 2016. – 212 с.

17. Гурвиц, Джудит. Просто о больших данных / Гурвиц Джудит, Ньюджент Алан, Халпер Ферн, Кауфман Марсия [перевод с английского]. — М.: Эксмо, 2015. - 400 с.

18. Демина, М. И. Основные методы управления банковскими рисками в условиях нестабильной ситуации в стране. / М.И. Демина. // Научный альманах. - 2017. - № 1-1(27). – С.45.

19. Жаринов, Р. Возможности обезличивания персональных данных в системах, использующих реляционные базы данных./ Р. Жаринов. // Управление, вычислительная техника и информатика. – 2018. - № 2 (32) – С. 189.

20. Зарядов, И. С. Введение в статистический пакет R: типы переменных, структуры данных, чтение и запись информации, графика. / И. С. Зарядов. - М.: Издательство Российского университета дружбы народов, 2015. - 207 с.
21. Зудилова, Т. В. Обработка данных большого объема. / Т. В. Зудилова. // Естественные и технические науки. – 2017. - № 4.
22. Иванов, П. Д. Технологии Big Data и их применение на современном предприятии./ Наука и инновации, 2016 [Электронный ресурс]. -Электрон. дан. - Режим доступа: <http://engjournal.ru/catalog/it/asu/>
23. Измалкова, С. А. Использование глобальных технологий «Big data» в управлении экономическими системами. / С. А. Измалкова. // Экономические и юридические науки. - 2018. - № 1. - С. 151–158.
24. Казаков, Р. Технологии BIG DATA в управлении крупными банками. / Р. Казаков. // Бизнес-образование в экономике знаний. - 2015. - №2. – С.45.
25. Короткова, Т. «EMC Data Lake 2.0- средство перехода к аналитике больших данных и цифровой экономике» [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://bigdata.cnews.ru/news/2018>
26. Майер-Шенбергер, В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. / В. Майер-Шенбергер. - М.: Манн, Иванов и Фербер, - 2017. - 240 с.
27. Маковейчук, К. А. Применение технологий бизнес-анализа к большим данным в системах нового типа. / К. А. Маковейчук. // Технические науки. – 2017. - №58.
28. Мамедова, Г. А. Технологии больших данных в электронном образовании. / Г.А. Мамедова. // Open education. – 2016. - №6.
29. Мацкевич, Е. Д. Риски коммерческого банка. / Е. Д. Мацкевич. // Международный журнал прикладных и фундаментальных исследований. – 2017. – № 6-2. – С. 313-317.
30. Мырзагалын, Н. С. Анализ существующих подходов к автоматизации банковской деятельности. / Н.С. Мырзагалын. // Молодой ученый. - 2017. - № 9. - С. 235–237.

31. Немалевич, С. Анализ больших данных [Электронный ресурс]. - Электрон. дан. - Режим доступа: <https://republic.ru/biz/1152340/>
32. Носов, Н. Big Data в российских банках. Начало большого пути. / Н. Носов. // ITWEEK.18.03.2018.
33. Обзор рынка BigData [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://habrahabr.ru/company/moex/blog/256747/>
34. Ордынский, А. А. Оценка кредитоспособности с помощью BIG DATA: проблемы и перспективы внедрения в России. / А. А. Ордынский. // Экономические науки. – 2017. - №7. – С.51.
35. Паклин, Н. Б. Бизнес-аналитика: от данных к знаниям: учебное пособие / Н. Б. Паклин. – СПб: Питер, 2017. – 704 с.
36. Плохих, Ю. В. Проблемы автоматизации в банковской деятельности. / Ю.В. Плохих. // Молодой ученый. - 2017. - №20. - С. 410-412.
37. Савельев, А. И. Проблемы применения законодательства о персональных данных в эпоху «Больших данных» (Big Data). / А. И. Савельев. // Право. Журнал Высшей школы экономики. - 2017.- №1. - С. 43–66.
38. Самойлова, И. А. Технологии обработки больших данных./ И. А. Самойлова. // Молодой ученый. - 2017. - №49. - С. 26-28.
39. Семенов, Ю. А. Обзор по материалам ведущих фирм, работающих в сфере сетевой безопасности. [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://book.itep.ru/10/2018/>
40. Сухобоков, А. А. Влияние инструментария Big Data на развитие научных дисциплин, связанных с моделированием / А. А. Сухобоков. // Наука и образование: научное издание МГТУ им. Н. Э. Баумана. – 2016. – № 3. – С. 207–240.
41. Сухорослов, О. В. Новые технологии распределенного хранения и обработки больших массивов данных. / О. В. Сухорослов. // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы», 2018. - 40 с.

42. Тренды мирового e-commerce рынка в 2016–2017 годах [Электронный ресурс]. - Электрон. дан. - Режим доступа: [https:// habrahabr.ru/company/](https://habrahabr.ru/company/)
43. Фрэнк, Б. Революция в аналитике. Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики: Альпина Паблишер. Учебно-методическое пособие. / Б. Фрэнк. - Нижневартовск: Изд-во Нижневарт. гос. ун-та, 2016. - 227 с.
44. Фрэнкс, Б. Укрощение больших данных: как извлекать смысл из гигантских потоков данных с помощью продвинутой аналитики. / Б. Фрэнкс. - М.: Манн, Иванов и Фербер, 2016. - 352 с.
45. Цветков, В. Я. Естественное и искусственное информационное поле. / В. Я. Цветков. // Международный журнал прикладных и фундаментальных исследований. - 2017. - №5. - С. 178-180.
46. Черняк, Л. Большие данные-новая теория и практика. / Л. Черняк. //Открытые системы. СУБД - 2017. - №10. - С.18-25.
47. Чехарин, Е. Е. Большие данные: большие проблемы. / Е. Е. Чехарин. // Перспективы науки и образования. – 2017. - №4. - С.45.
48. Что такое Big Data (BigData) в маркетинге: проблемы, алгоритмы, методы анализа // Landing Page Generator [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://lpgenerator.ru/blog/2015/11/17/>
49. Big Data от А до Я. Часть 1: Принципы работы с большими данными, парадигма MapReduce // Хабрахабр [Электронный ресурс]. - Электрон. дан. - Режим доступа: <https://habrahabr.ru/company>
50. Best 10 Products Big Data [Электронный ресурс]. - Электрон. дан. - Режим доступа: <https://www.crn.ru/news>
51. Cisco: в период с 2012 по 2017 гг. Объем мобильного трафика [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://www.cisco.com>
52. IBM: Бизнес выбирает когнитивные решения [Электронный ресурс]. - Электрон. дан. - Режим доступа: <https://www.crn.ru/news>
53. Russian Digital Libraries Journal. - 2017. № 19.

54. RUN - Articles in the academy [Электронный ресурс]. - Электрон. дан. - Режим доступа: <http://naukarus.com/>

55. Russian habr news [Электронный ресурс]. - Электрон. дан. - Режим доступа: <https://habr.com/article/>

56. Started with big data [Электронный ресурс]. - Электрон. дан. - Режим доступа: <https://intel.com/>

57. Technologies big data in banks [Электронный ресурс]. - <https://ibs.ru/>



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Дальневосточный федеральный университет»

ШКОЛА ЭКОНОМИКИ И МЕНЕДЖМЕНТА

Базовая кафедра современного банковского дела

ОТЗЫВ РУКОВОДИТЕЛЯ

на выпускную квалификационную работу студента \_\_\_\_\_  
Тарабанова Максима Александровича  
(фамилия, имя, отчество)

специальность (направление) 38.03.01 Экономика  
профиль «Банковское дело» группа Б1401Бдб

Руководитель ВКР канд. экон. наук, доцент кафедры современного банковского дела  
Ю. Б. Шатов  
(ученая степень, ученое звание, и.о. фамилия)

на тему «Механизм использования технологии Big Data в управлении бизнес-процессами  
банка: возможности и риски»

Дата защиты ВКР « 25 » июня 2018 г.

Актуальность и практическая значимость выполненной ВКР на заданную тему обоснованы и представлены в формулировках цели и задач (Введение), исследование технологии Big Data (Глава 1), обобщение результатов анализа (Глава 2).

При подготовке ВКР студент проявил стремление и умение применить полученные знания общетеоретических и профильных дисциплин к раскрытию компетенций специалиста банковского дела, обладающего достаточным уровнем способностей самостоятельного решения прикладных задач в области анализа, обобщения факторов, выработки управленческих и организационных предложений. Работа написана корректно.

Заключение: заслуживает оценки «хорошо» и присвоения квалификации «бакалавр»

Руководитель ВКР канд. эк. наук, доцент  
(уч. степень, уч. звание)

Ю. Б. Шатов  
(и.о. фамилия)

« \_\_\_\_ » \_\_\_\_\_ 2018 г.

