

## Анализ данных на базе технологии частного облака

*Т.Н.Кондратьева, К.А. Эксузян*

*Донской государственной технической университет, Ростов-на-Дону*

**Аннотация:** Статья посвящена вопросу использования частного облака для реализации анализа данных. Современные cloud-технологии способны существенно сократить расходы, и сегодня многие компании все чаще переносят свои корпоративные системы и бизнес-приложения в облако, в том числе и процесс анализа данных. Целью является анализ котировок драгоценных металлов за период с 01.01.2018 по 08.01.2018, взятых с информационного портала Finam.ru для розничных трейдеров. Для достижения поставленной цели ставились следующие задачи: виртуализация сети, для работы частного облака, проектирование частного облака на базе Microsoft System Center Virtual Machine Manager 2012, разработка сервиса для прогноза данных по котировкам драгоценных металлов (золото, серебро, платина, палладий) с использованием алгоритма временных рядов и дерева принятия решений. Предметом исследования является технология частного облака, его использование для развертывания сервиса для анализа данных. В заключении приведены результаты работы сервиса, показано практическое использование облачных вычислений для анализа данных. Оценка качества построенных моделей позволяет считать ее удовлетворительной со статистической точки зрения.

**Ключевые слова:** облачные вычисления, частное облако, data mining, облачный сервис, виртуализация, анализ данных, прогнозирование.

На данный момент облачные технологии являются самой быстро развивающейся инновацией со времени появления сети Интернет. Глобальная сеть кардинально повлияла на развитие самых разных компаний по всему миру, и в облачных вычислениях заложен потенциал такого же уровня. На сегодняшний день в IT-сообществе распространено мнение о том, что современные cloud-технологии способны существенно сократить расходы, и многие компании все чаще переносят свои корпоративные системы и бизнес-приложения в облако, в том числе и процесс анализа данных [1].

На сегодняшний день самыми успешными компаниями в сегменте облачных технологий считаются компании Google, Amazon и Microsoft.

Предметом исследования данной работы является технология частного облака. Задачами являются: разработка общей концепции облачной платформы; проектирование и создание частного облака.

В качестве модели обслуживания выбраны модели «Программное обеспечение как услуга (SaaS)» и «Инфраструктура как услуга (IaaS)» [2].

Проанализированы модели развертывания облака, частное, публичное, общественное и гибридное облако. Была выбрана наиболее подходящая для поставленной задачи модель – частное облако. Являясь инфраструктурой, предназначенной для пользования одной организацией, которая может быть в собственности и эксплуатироваться самой организацией, частное облако стало самым удобным и оптимальным вариантом для работы сервиса.

Облако развернуто на платформе Windows Server 2012 Datacenter с поддержкой системы Windows Server Active Directory, виртуализации Hyper-V и анализа приложений с помощью System Center. Доступ к данным, хранящимся на сервере, осуществляется при помощи технологии доступа к данным ADO.Net.

Частное облако построено на базе технологии Microsoft System Center Virtual Machine Manager 2012. Преимущества создания частного облака есть в комплексном подходе Microsoft при внедрении облачных вычислений и изменении способа предоставления ИТ-услуг, кроме того, данный сервис значительно дешевле аналогичного решения от VMware (4-10 раз).

Исследование проведено на основе данных котировок драгоценных металлов (золото, серебро, платина, палладий), по ценам при закрытии рынка в течение недели с 01.01.2018 по 08.01.2018, скачанных с сайта - [www.finam.ru](http://www.finam.ru). Хранение данных осуществляется в базе СУБД MS SQL Server с помощью компонента Database Engine.

Для прогнозирования данных используется компонент Analysis Services, включающий в себя набор средств для работы с OLAP и интеллектуальным анализом данных. Данные хранятся в режиме отдельного хранения Multidimensional OLAP (MOLAP). Далее, данные обрабатываются, сохраняются и индексируются при помощи специального формата,

---

оптимизированного под многомерные данные. На рисунке 1, представлено окно просмотра котировок драгоценных металлов за указанный период (Рис.1 - График котировок драгоценных металлов (золото, серебро, платина, палладий)).

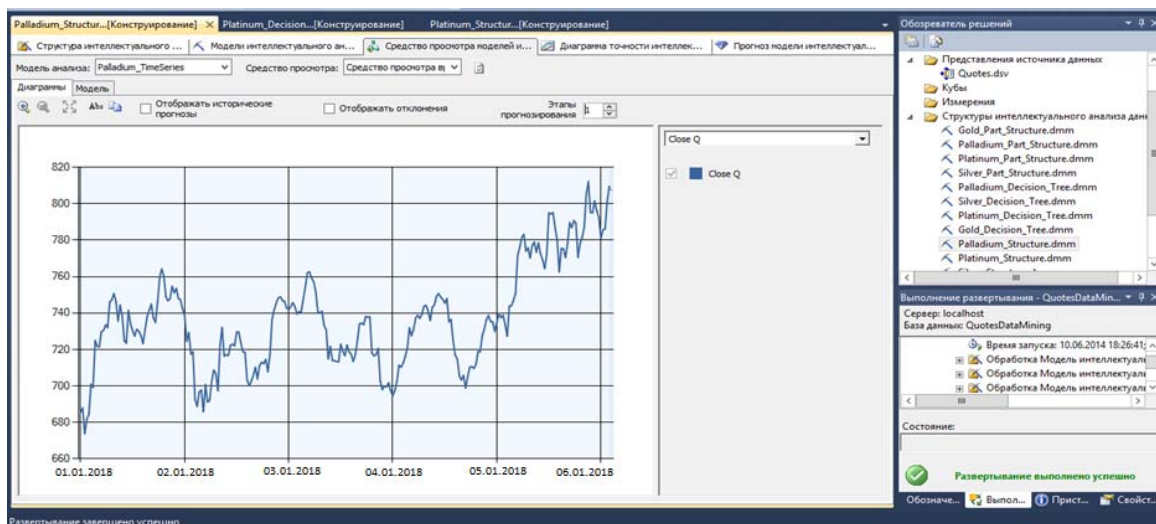


Рис. 1 - График котировок драгоценных металлов (золото, серебро, платина, палладий) (разработано автором)

Для интеллектуального анализа данных MS SQL Server предоставляет интегрированную среду для создания моделей интеллектуального анализа данных и работы с ними. Эта среда включает программу SQL Server Development Studio, которая включает в себя алгоритмы анализа данных и инструменты для создания запросов, облегчающие создание полноценного сервиса. Для прогнозирования данных используется компонент Analysis Services., который включают в себя набор средств для работы с OLAP и интеллектуальным анализом данных.

Анализ данных реализован с помощью двух алгоритмов прогнозирования: алгоритма временных рядов и дерева принятия решений. Для сравнения точности прогноза предоставлен выбор между ними: алгоритм временных рядов - Алгоритм 1, и дерева принятия решений - Алгоритм 2. Для алгоритма 1, так же предоставлен выбор метода прогноза: ARIMA,

ARTXP и MIXED. Критериями качества построенных моделей выбраны дисперсия прогноза и ошибка аппроксимации.

На рисунке 2 представлено главное окно сервиса с выводом графика с прогнозными значениями по двум алгоритмам, а так же основные критерии оценки их качества (Рис.2. Результат работы сервиса).

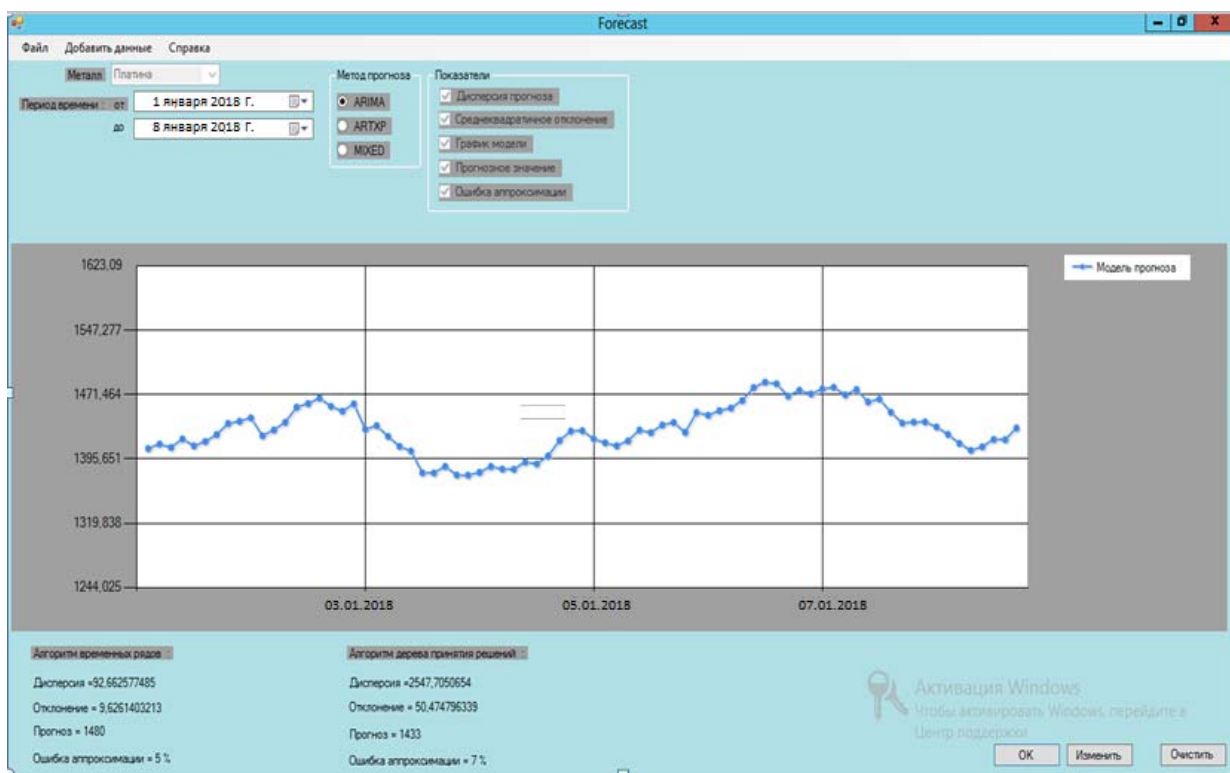


Рис. 2 - Результат работы сервиса (разработано автором)

Запрос прогнозного значения выполнен на языке DMX, с помощью алгоритма временных рядов (MIXED):

- подготовка данных;
- просмотр данных;
- построение модели, используя службу Microsoft Time Series;
- развертывание и обновление моделей.

Предлагаемый способ оценки можно считать приемлемым. Так как алгоритм ARTXP оптимизирован для прогнозирования следующего вероятного значения в ряду, а алгоритм ARIMA для повышения точности долгосрочного прогнозирования, то и результаты прогнозирования на

разработанном сервисе оказались ожидаемыми. Более точный прогноз показала модель ARTXP.

Аналогично строится модель и производится прогноз с алгоритмом дерева принятия решений.

В качестве критерия оценки прогноза используется, в том числе вычисление ошибки аппроксимации. Ошибка аппроксимации вычисляется в клиентской части приложения. Формула для расчета ошибки аппроксимации

$$A = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_p}{y_i} \right| \cdot 100\% \quad (1), \text{ где } y_i - \text{наблюдаемые уровни временного ряда, } y_p - \text{прогнозные значения временного ряда.}$$

Укрупненный алгоритм расчета ошибки аппроксимации имеет вид:

выбранные данные делим в пропорции 70:30. Получаем массивы a1 (объем массива 0,7n) и y (объем массива 0,3n), n – число уровней временного ряда;

инициализация. Выполняем прогноз по выборке a1 на 1 шаг согласно выбранному методу. Прогноз записываем в массив  $y_p$ ;

итерация. По данным массива a1 и полученным на предыдущем шаге прогнозам выполняем новый прогноз. Прогноз записываем в массив  $y_p$ ;

вычисление по формуле (1);

остановка.

Система была протестирована 100 раз в течение указанного периода по исходным данным. Прогноз строился по двум алгоритмам. Наиболее лучшие результаты занесены в таблицу 1.

Таблица 1 - Ошибки аппроксимации по результатам прогнозирования

n/n	Алгоритм 1		Алгоритм 2	
	Прогноз	Ошибка аппроксимации (%)	Прогноз	Ошибка аппроксимации (%)
1	1430	5,000	1480	7,000
2	1425	7,355	1476	8,932

---

---

3	1440	8,983	1465	11,763
4	1427	5,563	1491	12,672
5	1433	8,000	1477	10,934
...	...	...	...	...
97	1431	5,087	1490	8,457
98	1426	6,782	1493	9,294
99	1432	6,548	1470	11,354
100	1436	8,025	1472	12,001

Качество построенных моделей можно считать удовлетворительным, так как при использовании алгоритма временных рядов величина средней ошибки аппроксимации варьируется в допустимых пределах от 5% до 12%, наилучшее прогнозное значение было получено в результате работы Алгоритма 1, при выборе модели ARIMA (Рис.2.), с ошибкой в  $\bar{A} = 5\%$  (Табл.1.).

В случае вычисления прогнозного значения при использовании Алгоритма 2 (Рис.2.) наименьшая ошибка аппроксимации составила  $\bar{A} = 7\%$ . Что говорит о качестве обоих алгоритмов прогнозирования, с преимуществом у алгоритма временных рядов.

В работе построена общая концепция облачной платформы, спроектировано и развернуто частное облако, разработано приложение анализа данных с использованием инструментария Data Mining на базе облака.

В ходе выполнения работы была исследована технология частного облака. Разработанный сервис показал один из наилучших способов использования данной технологии. Преимущества данного сервиса:

Использование онлайн-сервиса без приобретения серверного и сетевого оборудования;

Уменьшение расходов на работу с информацией;



Возможность подключиться к облачному сервису с любого гаджета в любом месте, при условии подключения к Интернету;

Масштабируемость гибкость облачной модели на имеющихся ресурсах;

Возможность работы на разных платформах и с разными низкоуровневыми оболочками.

По итогам работы сервиса были реализованы поставленные задачи и получены удовлетворительные результаты анализа и прогнозирования данных.

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 17-01-00888 а*

### Литература

1. Алексеева Т.В., Использование облачных технологий для организации обучающего информационного пространства // Сборники конференций НИЦ Социосфера, 2012. № 17. С. 77-80.
2. Берёза Н.В., Современные тенденции развития мирового и российского рынка информационных услуг// Инженерный вестник Дона, 2012 №2. - URL: ivdon.ru/magazine/archive/n2y2012/758
3. Вебер, М. Бизнес-вычисления: практика расчета основных показателей / М.: Омега-Л, 2014. – С.143.
4. Кондратьева Т.Н. Прогнозирование тенденции финансовых временных рядов с помощью нейронной сети LSTM // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 9, №4. - С.5.



5. Кондратьева Т.Н., Комплексный подход в режиме торговой системе консервативного скальпинга // Инженерный вестник Дона 2013 №4 .- URL: [ivdon.ru/uploads/article/pdf/R\\_65\\_Kondratjeva.pdf\\_2090](http://ivdon.ru/uploads/article/pdf/R_65_Kondratjeva.pdf_2090)
6. Мисюра В.В., Кондратьева Т.Н., Бенгус Б.В. Сравнительный анализ методов прогнозирования тенденции развития // РГУПС, 2014. Ростов-на-Дону ISSN: 0201-727X. - 2014. - №2. - С. 124-128.
7. Облачные вычисления: обзор и рекомендации. Общая среда облачных вычислений - Рекомендации Национального Института Стандартов и Технологий (США), NIST, USA, 2007- С. 15.
8. Риз Д. Облачные вычисления // БХВ-Петербург, 2011- С. 27-29.
9. Фингар П., Dot.Cloud: облачные вычисления - бизнес-платформа XXI века // Аквамариновая книга, 2011, С. 256.
10. Эталонная архитектура облачных вычислений - Рекомендации Национального Института Стандартов и Технологий (США), NIST, USA, 2007- С. 3-7.
11. Безопасность для облачных вычислений // URL: [bytemag.ru/articles/detail.php?ID=17758](http://bytemag.ru/articles/detail.php?ID=17758), 03.03.2012.
12. Data mining (SSAS) // URL: [docs.microsoft.com/ru-ru/sql/analysis-services/data-mining/data-mining-ssas?view=sql-server-2017](http://docs.microsoft.com/ru-ru/sql/analysis-services/data-mining/data-mining-ssas?view=sql-server-2017)
13. Cloud computing // URL: [studylib.net/doc/7787068/cloud-computing](http://studylib.net/doc/7787068/cloud-computing).

### References

1. Alekseeva T.V., Ispol`zovanie oblachny`x texnologij dlya organizacii obuchayushhego informacionnogo prostranstva [Using cloud technologies to organize a learning information space]. Sborniki konferencij NICz Sociosfera, 2012. № 17. pp. 77-80.
  2. Beryoza N.V., Inženernyj vestnik Dona (Rus), 2012, №2. URL: [ivdon.ru/magazine/archive/n2y2012/758](http://ivdon.ru/magazine/archive/n2y2012/758)
-



3. Veber, M. Biznes-vy`chisleniya: praktika rascheta osnovny`x pokazatelej [Business computing: the practice of calculating key indicators]. M.: Omega-L, 2014. P.143.
4. Kondrat`eva T.N. Internet-zhurnal «NAUKOVEDENIE» Tom 9, №4. P.5.
5. Kondrat`eva T.N., Inženernyj vestnik Dona (Rus), 2013 №4. URL: [ivdon.ru/uploads/article/pdf/R\\_65\\_Kondratjeva.pdf\\_2090](http://ivdon.ru/uploads/article/pdf/R_65_Kondratjeva.pdf_2090)
6. Misyura V.V., Kondrat`eva T.N., Bengus B.V. RGUPS, 2014. Rostov-na-Donu ISSN: 0201-727X. 2014. №2. Pp. 124-128.
7. Oblachny`e vy`chisleniya: obzor i rekomendacii. Obshhaya sreda oblachny`x vy`chislenij [Cloud computing: overview and recommendations. Common Cloud Computing Environment]. Rekomendacii Nacional`nogo Instituta Standartov i Texnologij (SShA), NIST, USA, 2007. P. 15.
8. Riz D. Oblachny`e vy`chisleniya [Cloud computing]. BXV-Peterburg, 2011-S. 27-29.
9. Fingar P., Dot.Cloud: oblachny`e vy`chisleniya - biznes-platforma XXI veka [Dot.Cloud: cloud computing - a business platform of the XXI century]. Akvamarinovaya kniga, 2011, P. 256.
10. E`talonnaya arxitektura oblachny`x vy`chislenij [The reference architecture of cloud computing]. Rekomendacii Nacional`nogo Instituta Standartov i Texnologij (SShA), NIST, USA, 2007. Pp. 3-7.
11. Bezopasnost` dlya oblachny`x vy`chislenij [Security for cloud computing]. URL: [bytemag.ru/articles/detail.php?ID=17758](http://bytemag.ru/articles/detail.php?ID=17758), 03.03.2012.
12. Cloud computing. URL: [studylib.net/doc/7787068/cloud-computing](http://studylib.net/doc/7787068/cloud-computing).
13. Data mining (SSAS). URL: [docs.microsoft.com/ru-ru/sql/analysis-services/data-mining/data-mining-ssas?view=sql-server-2017](http://docs.microsoft.com/ru-ru/sql/analysis-services/data-mining/data-mining-ssas?view=sql-server-2017)