

УДК 004.6

DOI: [10.26102/2310-6018/2021.33.2.004](https://doi.org/10.26102/2310-6018/2021.33.2.004)

## Систематизация уровней зрелости корпоративной архитектуры в области анализа данных

Ф.Г. Логинов, С.А. Коваленко, Н.И. Гребенникова, В.А. Малиновкин

*Воронежский государственный технический университет,  
Воронеж, Российская Федерация*

**Резюме.** В настоящее время использование современных технологий и методов анализа данных позволяет создавать более совершенные информационные системы для организации качественного ведения бизнеса. Поэтому при проектировании корпоративной архитектуры данных при наличии существенных объемов информации необходимо учитывать множество факторов, систематизировать и корректно производить выбор определенного уровня зрелости. В данной статье представлен обзор технологий и методов в области анализа данных. Также были представлены и систематизированы уровни зрелости: «лоскутное одеяло», корпоративное хранилище данных, озера данных, машинное обучение, искусственный интеллект. Рассматриваемые средства имеют спрос при построении аналитической архитектуры компании. С помощью них можно осуществлять поиск данных, доступ к ним и обработку. Были рассмотрены и проанализированы преимущества и недостатки методов. Сравнение технологий производится по ряду характеристик, а именно: организация доступа к данным, способ хранения данных, процесс извлечения, преобразования данных, процесс построения бизнес-отчетов, использование машинного обучения, помощь в принятии решений. Эти аспекты являются основными при выборе средств построения корпоративной архитектуры в области анализа данных, так как они являются ключевыми при аналитической обработке данных.

**Ключевые слова:** аналитическая обработка данных, уровни зрелости корпоративной архитектуры в области анализа данных, хранилища данных, систематизация данных, проектирование корпоративной архитектуры данных

**Для цитирования:** Логинов Ф.Г., Коваленко С.А., Гребенникова Н.И., Малиновкин В.А. Систематизация уровней зрелости корпоративной архитектуры в области анализа данных. *Моделирование, оптимизация и информационные технологии*. 2021;9(2). Доступно по: <https://moitvivt.ru/ru/journal/pdf?id=988> DOI: 10.26102/2310-6018/2021.33.2.004

## Systematization of the maturity levels of the corporate architecture in the field of data analysis

F.G. Loginov, S. A. Kovalenko, N.I. Grebennikova, V.A. Malinovkin

*Voronezh State Technical University, Voronezh, Russian Federation*

**Abstract:** The use of modern technologies and data analysis methods allows you to create more advanced information systems for the organization of a high-quality business. Therefore, when designing a corporate data architecture, it is necessary to recognize many factors and correctly select a certain level of maturity. This article provides an overview of technologies and methods in the field of data analysis. The tools in question are in demand when building the company's analytical architecture. They can be used to search for, access, and process data. The advantages and disadvantages of the methods were considered. The comparison of technologies is made by several characteristics, namely: data access organization, data warehouse building method, the process of extracting, converting data, and the process of building a business report. These aspects are the main ones when choosing the tools for building a corporate architecture in data analysis since they are crucial in analytical data processing.

**Keywords:** analytical data processing, maturity levels of the corporate architecture in the field of data analysis, data warehouses, data systematization, designing a corporate data architecture

**For citation:** Loginov F. G., Kovalenko S. A., Grebennikova N. I., Malinovkin V. A. Systematization of levels of maturity of corporate architecture in the field of data analysis *Optimization and Information Technology*. 2021;9(2). Available from: <https://moitvivt.ru/ru/journal/pdf?id=988> DOI: 10.26102/2310-6018/2021.33.2.004 (In Russ).

## Введение

В любой сложной технико-экономической системе, от предприятия до отрасли, одно из центральных мест в современной теории и практике управления занимает проблема принятия эффективных решений.

Со временем сложилась картина неравномерного применения технологий и методов анализа данных в корпоративном секторе. Другими словами, одновременно существуют различные уровни зрелости корпоративной архитектуры в области анализа данных в зависимости от использования тех или иных технологий и методов. В работе предлагается классификация на основе следующих определений:

- «лоскутное одеяло» – данные не систематизируются, хранятся в разных ИС, используются время от времени;
- корпоративные хранилища данных – данные централизованно собираются и хранятся, выполняется первичная обработка данных и построение агрегированных отчетов и бизнес-аналитики;
- озера данных – помимо отчетов хранятся детальные данные и неструктурированные данные, такие решения могут быть горизонтально масштабированы;
- машинное обучение – использование методов и алгоритмов машинного обучения на прецедентах для анализа детальных данных и построения прогнозных моделей;
- искусственный интеллект – использование методов оптимизации и искусственного интеллекта для построения систем помощи принятия решений.

Целью работы является разработка дорожной карты развития аналитической архитектуры предприятия, которая будет основываться на уровне технологической зрелости и стоящих перед предприятием задачах. Для этого необходимо: изучить и систематизировать возможные состояния аналитической архитектуры, выделить ключевые критерии, определяющие отношение аналитической архитектуры к тому или иному уровню зрелости, определить последовательность перехода от одного уровня зрелости к другому.

## Технологии и методы в области анализа данных

### Лоскутное одеяло

Под «лоскутным одеялом» понимается такое построение аналитической архитектуры, когда в компании не существует единого хранилища данных и инструментов для их анализа. В результате общекорпоративный анализ данных затруднен или вообще невозможен. Для решения поставленной задачи аналитик предварительно должен вручную выполнить ряд задач:

- выяснить, существуют ли в компании необходимые данные;
- получить доступ к данным, загрузить себе на рабочую станцию;
- привести данные к единой форме.

Только после этого появится возможность приступить к работе над задачей. Так как современных инструментов для анализа нет, аналитик использует программы для работы с электронными таблицами, функциональность которых сильно ограничена. Зачастую эти обстоятельства не просто усложняют и замедляют работу, а вовсе делают ее невозможной.

### Корпоративные хранилища данных

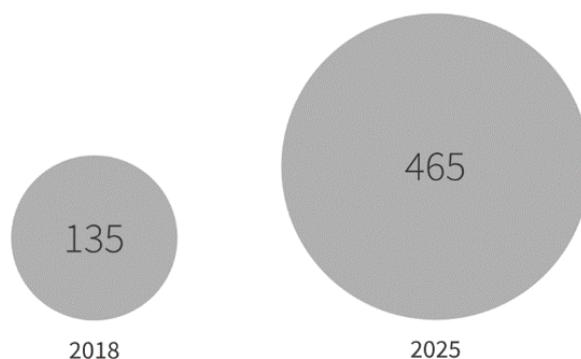
Для устранения препятствий, описанных в предыдущем разделе, на рубеже 1970-х и 1980-х годов возникла концепция корпоративных хранилищ данных (Data Warehouse), которая активно используется с 1990 года, когда в корпоративном секторе стали применяться различные информационные системы для учета показателей бизнеса. Такие приложения используются для автоматизации определенного производственного процесса, например, построение бухгалтерских отчетов, проведения транзакций и т.д. [1-3]

Таким образом, главной целью создания корпоративного хранилища данных (КХД) является измерение и описание корпоративных процессов.

### Озера данных

Причиной возникновения озер данных стало то, что количество генерируемых данных стремительно растет, меняются форматы хранения и скорость накопления [4-5]. На Рисунке 1 представлен прогноз роста объемов генерации информации.

Прогноз роста объемов генерации информации, млрд. гигабайт в день



Источник: IDC, МНИАП

Рисунок 1 – Прогноз роста объемов генерации информации (млрд. Гб в день)  
Figure1 – Forecast of growth of information generation volumes (billion Gb per day)

Ответом на эти три вызова: большой объем, необходимость работы в реальном времени и разнородность данных – стали решения класса больших данных и корпоративная архитектура озер данных, построенная на их основе. В основу технологий больших данных заложены следующие принципы:

– горизонтальная масштабируемость – возможность наращивать вычислительную мощность посредством добавления новых машин в кластер;

- распределенные файловые системы позволяют хранить и обрабатывать файлы любого формата: текстовые и бинарные – даже если их размер превышает емкость памяти одной машины в составе кластера;
- параллельные вычисления – способ обработки данных, при котором исходный массив информации разбивается на части, вычисления над которыми выполняются независимо.

Из-за своих свойств озера данных в меньшей степени подходят для построения интерактивной отчетности, но в гораздо большей степени помогают в работе аналитикам благодаря тому, что история хранения исходных данных практически не ограничена.

### Машинное обучение

Следующим шагом после развития компетенций в области работы с большими данными для большинства компаний становится применение алгоритмов и методов машинного обучения для анализа бизнес-процессов [6,7,8]. Наиболее частой задачей являются регрессия и прогнозирование. Применение машинного обучения без инфраструктуры хранения и обработки больших данных зачастую невозможно, т. к. для обучения моделей требуется обработать массивы информации, которые часто невозможно или слишком дорого хранить в КХД.

### Искусственный интеллект

Благодаря использованию технологий хранилищ и озер данных у аналитиков появилась возможность измерять и количественно оценивать бизнес-процессы компании, с распространением машинного обучения стало возможным строить прогнозные модели с целью точного планирования [9-11]. Переходом к следующему уровню зрелости корпоративной аналитической архитектуры стало применение технологий искусственного интеллекта для оптимизации бизнес-процессов и операционного управления. Под искусственным интеллектом в данной работе понимается способность системы правильно интерпретировать внешние данные, извлекать уроки из данных и использовать полученные знания для достижения конкретных целей и задач при помощи гибкой адаптации.

Принцип применения ИИ в задачах корпоративного управления можно сформулировать довольно просто. В сущности, это задача оптимизации, которая может быть представлена в виде функции:

$$y = f(x^*, r) \quad (1)$$

где  $f(x^*, r)$  – предиктивная модель машинного обучения, описывающая состояние объекта;  $x^*$  – массив характеристик объекта;  $y$  – критерий оптимизации, выбранный пользователем, например, максимизация или минимизация состояния объекта;  $r$  – ограничения, например, затраты.

В качестве примера можно привести задачу планирования геолого-технических мероприятий в нефтегазовой отрасли. В этом случае  $f(x^*, r)$  – совокупность моделей, прогнозирующая доход организации от продажи добытой нефти;  $x^*$  – массив параметров, которыми компания может управлять, например, обороты двигателей насосов, количество нагнетательных и добывающих скважин и т. д., а так же тех, которыми управлять не может, например, величина НДС или курсы валют;  $r$  – ограничения, например, затраты на проведение ГТМ или полка ОПЕС;  $y$  – в данном случае, очевидно, необходимость максимизации. В результате решения данной задачи ИИ подберет такие значения параметров, которыми компания способна управлять, чтобы

максимизировать прибыль и не выйти за ограничения. Таким образом, использование ИИ в корпоративном управлении часто представляет собой оптимизационную задачу, которая оптимизирует ту или иную функцию машинного обучения. В данной работе не рассматриваются такие аспекты применения ИИ как распознавание, например, голоса или изображений.

Таким образом, на данном уровне зрелости происходит принципиальное изменение подхода к анализу данных от описательной и прогнозной аналитики к интеллектуальному управлению бизнес-процессами.

### Анализ аналитических архитектур крупнейших вендоров

Рассмотрим, какие варианты построения аналитических архитектур предприятий предлагают крупнейшие вендоры. Для определения крупнейших вендоров обратимся к анализу компании Гартнер (Рисунок 2) [12].



Рисунок 2 – Магический квадрант Gartner для решений по управлению данными для аналитики

Figure 2 – Gartner Magic Quadrant for Data Management Solutions for Analytics

Такие решения являются кроссиндустриальными, а значит подходят для применения в любых отраслях корпоративного сектора: от торговых сетей до промышленных предприятий.

Тремя лидерами в этой области являются компании Microsoft, Oracle и Teradata. Ниже приводятся диаграммы компонентов архитектур решений вендоров (Рисунки 3-5).

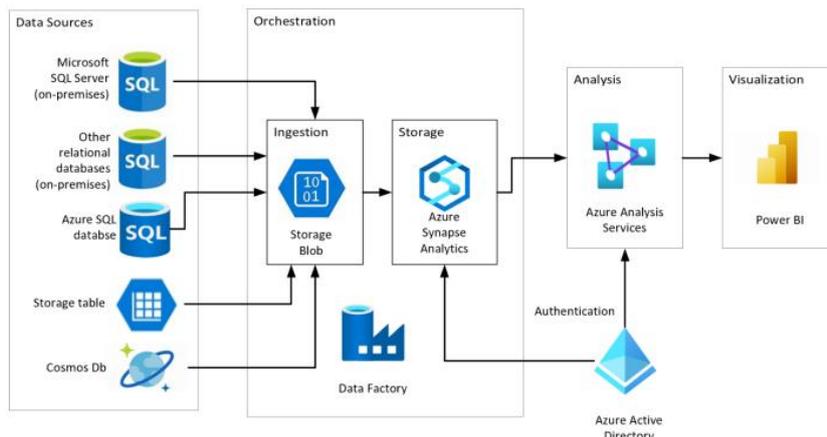


Рисунок 3 – Архитектура аналитической платформы от Microsoft  
 Figure 3 – The architecture of Microsoft analytics platform

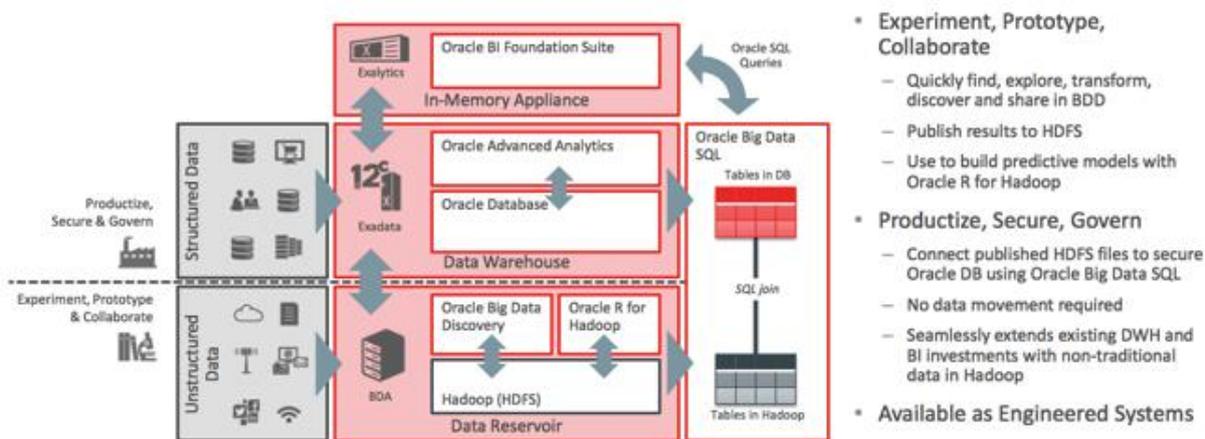


Рисунок 4 – Архитектура аналитической платформы от Oracle  
 Figure 4 – The architecture of Oracle analytics platform

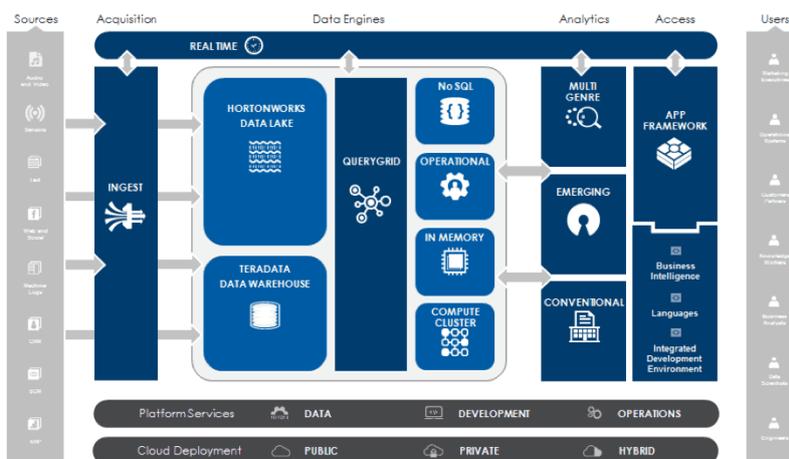


Рисунок 5 – Архитектура аналитической платформы от Teradata  
 Figure 5 – The architecture of Teradata analytics platform

Все их объединяет наличие компонентов для следующих задач:

- сбор, организация и хранение данных в реляционной форме, что соответствует уровню зрелости «Хранилище данных»;
- работа с большими объемами неструктурированных данных, что соответствует уровню зрелости «Озеро данных»;
- разработка продвинутых аналитических алгоритмов, что соответствует уровню зрелости «Машинное обучение»;
- создание приложений в том числе на базе искусственного интеллекта, что соответствует уровню зрелости «Искусственный интеллект».

Кроме того, важной особенностью их является то, что потоки данных также проходят от меньших уровней зрелости к большим, а значит при последовательной разработке аналогичных решений лучшей практикой будет придерживаться сходной стратегии.

### Результаты и обсуждение

В зависимости от использования рассмотренных технологий и методов различие уровней зрелости корпоративной архитектуры в области анализа данных состоит в том, каким образом организован доступ к данным, процесс извлечения, преобразования данных, процесс построения бизнес-отчетов, в способе хранения данных, предусмотрено ли использование машинного обучения, осуществляется ли помощь в принятии решений.

От выбора средств построения корпоративной архитектуры в области анализа данных зависит эффективность аналитической обработки данных, и, как следствие, принятие обоснованных решений для управления бизнес-процессами для успешной деятельности.

В Таблице 1 приведена сравнительная характеристика аспектов, являющихся ключевыми при аналитической обработке данных.

Таблица 1 – Систематизация уровней зрелости корпоративной архитектуры

Table 1 – Systematization of the maturity levels of the corporate architecture

Уровни зрелости	Систематизация данных	Централизованное хранение	Осуществление преобразования данных	Построение бизнес-отчетов	Работа с большими данными	Построение предиктивных моделей	Помощь в принятии решений
«Лоскутное одеяло»	Нет	Нет	Нет	Нет	Нет	Нет	Нет
Корпоративные хранилища данных	Да	Да	Да	Да	Нет	Нет	Нет
Озера данных	Да	Да	Да	Да	Да	Нет	Нет
Машинное обучение	Да	Да	Да	Да	Да	Да	Нет
Искусственный интеллект	Да	Да	Да	Да	Да	Да	Да

Важно отметить, что переход между уровнями зрелости является строго последовательным, то есть возможно движение только между соседними уровнями зрелостями. Так происходит, поскольку функциональные возможности каждого нового уровня зависят от предыдущего. Это значит, что прежде, чем приступать к разработке аналитических решений на базе искусственного интеллекта, предприятие должно разработать архитектуру анализа данных, включающую в себя элементы хранилища данных, озера данных, а также иметь инструментарий для работы с машинным обучением, но для построения бизнес-отчетности компании будет достаточно уровня зрелости на базе хранилища данных.

На Рисунке 6 представлена схема дорожной карты развития корпоративной аналитической архитектуры.



Рисунок 6 – Дорожная карта развития корпоративной аналитической архитектуры  
Figure 6 – Roadmap for Corporate Analytic Architecture Development

Основываясь на этой информации, можно понять, на каком уровне зрелости аналитической архитектуры предприятие находится в данный момент, а также, какие мероприятия следует предпринять для того, чтобы обеспечить возможность решения тех или иных аналитических задач в дальнейшем.

### Заключение

В работе были рассмотрены и систематизированы уровни зрелости корпоративной архитектуры в области анализа данных с точки зрения применения технологий и методов работы с большими объемами данных и алгоритмами машинного обучения. Были выделены критерии, на основании которых аналитическую архитектуру можно отнести к тому или иному уровню зрелости, а также определена последовательность перехода с одного уровня зрелости на другой.

Нередки ситуации, когда предпринимаются попытки приступить к разработке аналитических решений на базе искусственного интеллекта, когда уровень зрелости организации – «лоскутное одеяло» или в лучшем случае «корпоративное хранилище данных». Такие инициативы обречены на провал, если не будут выполнены мероприятия по приведению уровня зрелости аналитической архитектуры предприятия к нужному состоянию.

Информация данной работы будет полезна корпоративным ИТ-архитекторам и руководителям бизнес-подразделений для понимания жизненного цикла аналитической архитектуры и уровней ее зрелости, поможет сформировать дорожную карту развития информационной системы.

### ЛИТЕРАТУРА

1. Что такое корпоративное хранилище данных (DWH): основы BigData. Доступно по: <https://www.bigdataschool.ru/> (дата обращения 20.05.2021)
2. Сарка Д. *Microsoft SQL Server 2012. Реализация хранилищ данных*. 2014.
3. Жуковский О.И. *Хранилища данных*. 2015.

4. Озеро данных. Доступно по: <https://habr.com/ru/post/485180/> (дата обращения 20.05.2021)
5. Что такое озера данных и почему в них дешевле хранить bigdata. Доступно по: <https://mcs.mail.ru/blog/chto-takoe-ozera-dannyh-i-zachem-tam-hranyat-big-data> (дата обращения 20.05.2021)
6. Mitchell T. Machine learning. *McGraw-Hill Science/Engineering*. 1997.
7. Талачаев И. Е. Применение машинного обучения в задаче классификации с использованием языка Python. *Военный инновационный технополис «ЭРА»*. 2019.
8. Паклин Н. Б. *Бизнес-аналитика: от данных к знаниям*. 2013.
9. Ноткин Л.И. *Искусственный интеллект и проблемы обучения*. 1999.
10. Margaret A. Boden. Creativity and artificial intelligence. *Elsevier. Artificial intelligence* 1998;103:347-356
11. Шалютин С.М. *Искусственный интеллект*. 1985.
12. Magic Quadrant for Data Management Solutions for Analytics. Доступно по: <https://b2bsalescafe.files.wordpress.com/2017/07/magic-quadrant-for-data-management-solutions-for-analytics-feb-2017.pdf> (дата обращения 20.05.2021)

## REFERENCES

1. What is Enterprise Data Storage (DWH): BigData Basics. Available at: <https://www.bigdataschool.ru/> (accessed 20.05.2021)
2. Sarka D. *Microsoft SQL Server 2012. Implementation of data warehouses*. 2014.
3. Zhukovsky O. I. *Data storage*. 2015.
4. Data Lake. Available at: <https://habr.com/ru/post/485180/> (accessed 20.05.2021)
5. What are data lakes and why it is cheaper to store bigdata in them. Available at: <https://mcs.mail.ru/blog/chto-takoe-ozera-dannyh-i-zachem-tam-hranyat-big-data> (accessed 20.05.2021)
6. Mitchell T. Machine learning. *McGraw-Hill Science/Engineering*. 1997.
7. Talachayev I. E. Application of machine learning in the classification problem using the Python language. *Military innovation Technopolis "ERA"*. 2019.
8. Paklin N. B. *Business analytics: from data to knowledge*. 2013.
9. Notkin L. I. *Artificial intelligence and learning problems*. 1999.
10. Margaret A. Boden. Creativity and artificial intelligence. *Elsevier. Artificial intelligence* 1998;103:347-356
11. Shalyutin S. M. *Artificial intelligence*. 1985.
12. Magic Quadrant for Data Management Solutions for Analytics. Available at: <https://b2bsalescafe.files.wordpress.com/2017/07/magic-quadrant-for-data-management-solutions-for-analytics-feb-2017.pdf> (accessed 20.05.2021)

## ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

**Логинов Федор Геннадьевич**, аспирант,  
Воронежский государственный технический  
университет, Воронеж, Российская Федерация.  
e-mail: [lfgoff@bk.ru](mailto:lfgoff@bk.ru)

**Fedor G. Loginov**, Postgraduate, Voronezh  
state technical University, Voronezh, Russia  
Federation.

**Коваленко Сергей Александрович**, аспирант,  
Воронежский государственный технический  
университет, Воронеж, Российская Федерация.  
e-mail: [sergpc@yandex.ru](mailto:sergpc@yandex.ru)

**Sergey A. Kovalenko**, Postgraduate,  
Voronezh state technical University  
Voronezh, Russia Federation.

**Гребенникова Наталия Ивановна**, канд. техн. наук, доцент, Воронежский государственный технический университет, Воронеж, Российская Федерация.

*e-mail:* [g-naty@yandex.ru](mailto:g-naty@yandex.ru)

**Nataliya I. Grebennikova**, Cand. Sc. (Technical), Associate Professor, Voronezh State Technical University, Voronezh, Russia Federation.

**Малиновкин Владислав Алексеевич**, магистрант, Воронежский государственный технический университет, Воронеж, Российская Федерация.

*e-mail:* [malinovkin@mail.ru](mailto:malinovkin@mail.ru)

**Vladislav A. Malinovkin**, master's degree student, Voronezh State Technical University, Voronezh, Russia Federation.