

Концентратор данных (Data Hub)

Разделы: [Источники данных](#)

Концентратор данных — это подход к построению корпоративной информационной архитектуры, при котором все источники данных компании интегрируются на одной платформе. При этом обмен между источниками и потребителями производится через некоторую центральную точку. Цель такого подхода — обеспечить единое представление данных, отвечающее потребностям компании, и бесперебойный доступ к ним.

Концентратор обеспечивает единую точку доступа для всех потребителей информации, будь то приложение, аналитик или бизнес-пользователь. Он также позволяет манипулировать данными для решения различных задач, обеспечивая возможности централизованного управления их потоками.

Основными преимуществами концентраторов являются:

- Объединение разрозненных мест хранения в единый интерфейс для всех данных компании.
- Возможность создания высокоскоростных и высокопроизводительных конвейеров обработки данных.
- Хорошая видимость данных и доступность к ним.

Структурно архитектуру концентратора данных можно представить в виде нескольких уровней:

- **Уровень исходной системы** отвечает за извлечение данных из исходного источника и их интеграцию с концентратором данных.
- **Уровень интеграции данных** содержит все преобразования, необходимые для того, чтобы сделать информацию понятной для конечных пользователей. Часто это предполагает такие операции, как гармонизацию и обогащение данных.
- **Уровень хранения** обеспечивает потребности управления базами данных и моделирования бизнес-процессов.
- **Уровень доступа к данным** объединяет все точки доступа, подключенные к концентратору (OLTP-системы, платформы бизнес-аналитики, инструменты машинного обучения и т. д.).
- **Уровень управления** обеспечивает контроль над процессами интеграции, преобразования и потоками данных.

Концентраторы часто упоминаются в одном ряду с хранилищами и озерами данных как разные подходы к архитектуре информационной платформы компании. Но они не взаимозаменяемы, хотя часто используются совместно. В таблице приводится сравнение

этих архитектур.

	Хранилище данных	Озеро данных	
Область применения	Бизнес-аналитика и <u>интеллектуальный анализ данных</u>	<u>Наука о данных</u> и машинное обучение	
Структура данных	Только <u>структурированные</u>	Полуструктурированные и неструктурированные	
<u>Качество данных</u>	Высокая	Среднее и низкое	
Производительность	Средняя и низкая	Средняя и низкая	
Вид интеграции	<u>ETL</u>	<u>ELT</u>	

Поддержка и обслуживание концентраторов данных является нетривиальной задачей, требующей постоянных усилий по обеспечению их работоспособности. Однако использование этой технологии открывает перед предприятием действительно широкие возможности в плане эффективного использования данных для совершенствования бизнеса.