

•ВВЕДЕНИЕ



В этой главе представлен обзор технологий хранения больших данных, которые послужили вкладом в создание межотраслевой дорожной карты для развития технологий больших данных в ряде важных прикладных областей.

Вместо того, чтобы подробно рассматривать конкретные отдельные технологии, в этой главе представлен широкий обзор технологий хранения данных, чтобы получить общее представление о возможностях отдельных технологий и областей, требующих дальнейшего изучения.

Место хранения больших данных в общей цикле создания больших данных показано на рис.



Data	Data	Data	Data	Data
Acquisition	Analysis	Curation	Storage	Usage
Structured data Unstructured data Event processing Sensor networks Protocols Real-time Data streams Multimodality	 Stream mining Semantic analysis Machine learning Information extraction Linked Data Data discovery 'Whole world' semantics Ecosystems Community data analysis Cross-sectorial data analysis 	 Data Quality Trust / Provenance Annotation Data validation Human-Data Interaction Top-down/Bottom- up Community / Crowd Human Computation Curation at scale Incentivisation Automation Interoperability 	 In-Memory DBs NoSQL DBs NewSQL DBs Cloud storage Query Interfaces Scalability and Performance Data Models Consistency, Availability, Partition-tolerance Security and Privacy Standardization 	Decision support Prediction In-use analytics Simulation Exploration Visualisation Modeling Control Domain-specific usage

Хранилище больших данных связано

с хранением и управлением данными масштабируемым способом, удовлетворяющим потребности приложений, которым требуется доступ к данным.

Идеальная система хранения

больших данных должна позволять хранить практически неограниченный объем данных, справляться с высокими скоростями доступа на запись и чтение, гибко и эффективно работать с целым рядом различных моделей данных, поддерживать как структурированные, так и неструктурированные данные, а также соображения конфиденциальности, работают только с зашифрованными данными.



usage

Data	Data Analysis	Data	Data	Data
Acquisition		Curation	Storage	Usage
Structured data Unstructured data Event processing Sensor networks Protocols Real-time	Stream mining Semantic analysis Machine learning Information extraction Linked Data Data discovery	Data Quality Trust / Provenance Annotation Data validation Human-Data Interaction Top-down/Bottom-up Community / Crowd	In-Memory DBs NoSQL DBs NewSQL DBs Cloud storage Query Interfaces Scalability and Performance Data Models Consistency, Assistance,	Decision support Prediction In-use analytics Simulation Exploration Visualisation Modeling Control Domain-specific

Security and

Standardization

Хранение данных в общей последовательности создания больших данных

· Human

Cross-sectorial

Computation

Automation

Curation at scale
 Incentivisation

Interoperability

Технологии хранения больших данных называются (породской университ технологиями хранения, которые определенным образом решают проблему объема, скорости или структурированности и не попадают в категорию систем реляционных баз данных.

Это не означает, что системы реляционных баз данных не решают эти проблемы, но существуют альтернативные технологии хранения, такие как хранилища по столбцам и продуманные комбинации различных систем хранения, например, с использованием распределенной файловой системы Hadoop (HDFS), зачастую более эффективны и менее дороги.

Системы хранения больших данных обычно решают проблему больших объемов за счет использования распределенных архитектур без общего доступа. Это позволяет удовлетворить возросшие требования к хранилищу за счет масштабирования до новых узлов, обеспечивающих вычислительную мощность и хранилище.

Новые машины могут быть легко добавлены в кластер хранения, а система хранения заботится о прозрачном распределении данных между отдельными узлами.

Решения для хранения данных также должны справляться со скоростью и разнообразием данных. Скорость важна в смысле задержки запроса, т. е. сколько времени требуется, чтобы получить ответ на запрос? Это особенно важно в условиях высокой скорости поступающих данных. Например, произвольный доступ для записи в базу данных может значительно снизить производительность запросов, если требуется обеспечить транзакционные гарантии.

Например, графовые базы данных являются подходящими системами хранения для решения этих задач.