ACID EXPLAINED: ATOMIC, CONSISTENT, ISOLATED & DURABLE

I don't think it's an overstatement to say that data is pretty important. Data is especially important for modern organizations. In fact, The Economist went so far as to say that data has surpassed oil as the world's most valuable resource, and that was back in 2017.

One of the problems with data, though, is the massive amounts of it that need to be processed on a daily basis. There's so much data being generated across the globe these days that we have to come up with a new term just to express how much data there is: big data. Sure, it's not the most impressive-sounding term out there, but the fact remains.

With all this big data out there, organizations are seeking ways to improve how they manage it all from a practical, computational, and security standpoint. Like Spiderman's Uncle Ben once said:

"With great comes great responsibility."

The best method the IT world has created for navigating the complexities of data management is through the use of databases.

What is a database?

Databases are structured sets of data that are stored within computers. Oftentimes, databases are stored on entire server farms filled with computers that were made specifically for the purpose of handling that data and the processes necessary for making use of it.

Modern databases are such complex systems that management systems have been designed to

handle them. These database management systems (DBMS) seek to optimize and manage the storage and retrieval of data within databases.

One of the guiding stars leading organizations to successful database management is the ACID approach.

What is ACID?

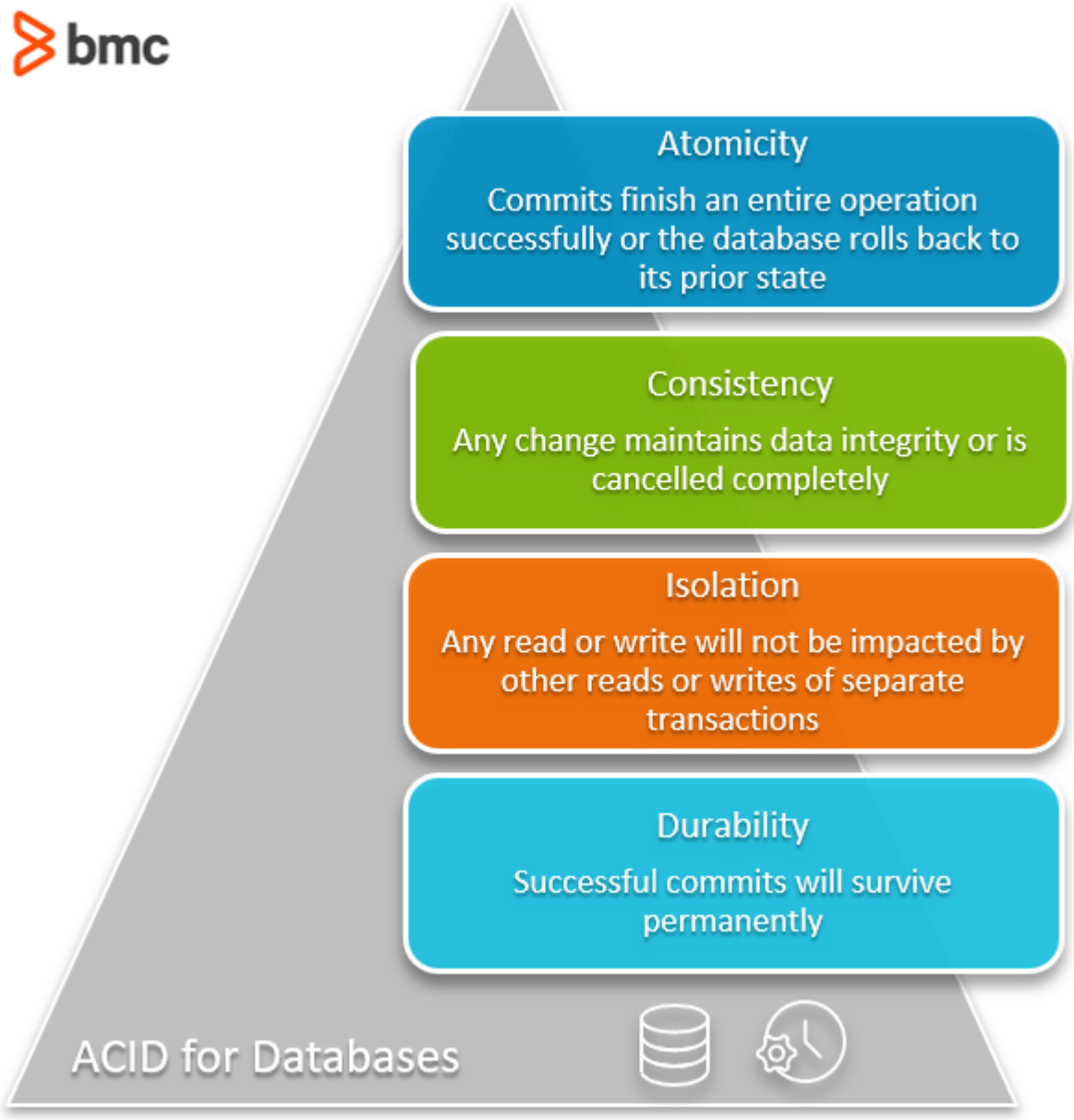
In the context of computer science, ACID stands for:

- Atomicity
- Consistency
- Isolation
- Durability

Together, ACID is a set of guiding principles that ensure database transactions are processed reliably. A database transaction is any operation performed within a database, such as creating a new record or updating data within one.

Changes made within a database need to be performed with care to ensure the data within doesn't become corrupted. Applying the ACID properties to each modification of a database is the best way to maintain the accuracy and reliability of a database.

Let's look at each component of ACID.



Atomicity

In the context of databases, atomicity means that you either:

- Commit to the entirety of the transaction occurring
- Have no transaction at all

Essentially, an atomic transaction ensures that any commit you make finishes the entire operation successfully. Or, in cases of a lost connection in the middle of an operation, the database is rolled back to its state prior to the commit being initiated.

This is important for preventing crashes or outages from creating cases where the transaction was partially finished to an unknown overall state. If a crash occurs during a transaction with no atomicity, you can't know exactly how far along the process was before the transaction was interrupted. By using atomicity, you ensure that either the entire transaction is successfully completed—or that none of it was.

Consistency

Consistency refers to maintaining data integrity constraints.

A consistent transaction will not violate integrity constraints placed on the data by the database rules. Enforcing consistency ensures that if a database enters into an illegal state (if a violation of data integrity constraints occurs) the process will be aborted and changes rolled back to their previous, legal state.

Another way of ensuring consistency within a database throughout each transaction is by also enforcing declarative constraints placed on the database.

An example of a declarative constraint might be that all customer accounts must have a positive balance. If a transaction would bring a customer account into a negative balance, that transaction would be rolled back. This ensures changes are successful at maintaining data integrity or they are canceled completely.

Isolation

Isolated transactions are considered to be “serializable”, meaning each transaction happens in a distinct order without any transactions occurring in tandem.

Any reads or writes performed on the database will not be impacted by other reads and writes of separate transactions occurring on the same database. A global order is created with each transaction queueing up in line to ensure that the transactions complete in their entirety before another one begins.

Importantly, this doesn't mean two operations can't happen at the same time. Multiple transactions can occur as long as those transactions have no possibility of impacting the other transactions occurring at the same time.

Doing this can have impacts on the speed of transactions as it may force many operations to wait before they can initiate. However, this tradeoff is worth the added data security provided by isolation.

Isolation can be accomplished through the use of a sliding scale of permissiveness that goes between what are called optimistic transactions and pessimistic transactions:

- **An optimistic transaction schema assumes that other transactions will complete without reading or writing to the same place twice.** With the optimistic schema, both transactions will be aborted and retried in the case of a transaction hitting the same place twice.
- **A pessimistic transaction schema provides less liberty and will lock down resources on the assumption that transactions will impact other ones.** This results in fewer abort and retries, but it also means that transactions are forced to wait in line for their turn more often in comparison to the optimistic transaction approach.

Finding a sweet spot between these two ideals is often where you'll find the best overall result.

Durability

The final aspect of the ACID approach to database management is durability.

Durability ensures that changes made to the database (transactions) that are successfully committed will survive permanently, even in the case of system failures. This ensures that the data within the database will not be corrupted by:

- Service outages
- Crashes
- Other cases of failure

Durability is achieved through the use of changelogs that are referenced when databases (or portions of the database) are restarted.

ACID supports data integrity & security

When every aspect of the ACID approach is brought together successfully, databases are maintained with the utmost data integrity and data security to ensure that they continuously provide value to the organization. A database with corrupted data can present costly issues due to the huge emphasis that organizations place on their data for both day-to-day operations as well as strategic analysis.

Using ACID properties with your database will ensure your database continues to deliver valuable data throughout operations.

Related reading

- BMC Machine Learning & Big Data Blog
- Data Architecture Explained: Components, Standards & Changing Architectures
- CAP Theorem for Databases: Consistency, Availability & Partition Tolerance
- Data Streaming Explained: Pros, Cons & How It Works
- Data Ethics for Companies
- 3 Simple Data Monetization Strategies for Companies