

THEME A: CONCEPTS · A3.4 · HL ONLY

Alternative databases and data warehouses — Mark scheme

33 marks · covers A3.4.1–A3.4.4

How to mark. Each creditworthy point is on its own line and ends with a semicolon (;). A line beginning with ... is an *expansion* of the line above it, so the pair is worth the marks shown. A slash (/) separates acceptable alternative wordings. Words in (brackets) are not needed for the mark. Mark positively: credit what the student has got right rather than penalising what is missing, and accept any valid alternative not listed here. Do not award more than the maximum for a part. Where a later part uses a wrong answer from an earlier part correctly, award follow-through marks (FT). HL marks parts that assess higher level only content.

Q	ANSWER	MARKS
HL 1(a) A3.4.1	Award [2] max Award [1] per point up to [2] max. A NoSQL database stores data in a form other than the fixed rows and columns of the relational model, for example documents, key-value pairs, wide columns or graphs; It does not require a rigid schema fixed in advance, so records with different fields can be stored together; It is designed to scale out horizontally across many servers rather than up on one; It often relaxes strict consistency in exchange for availability and speed;	2
HL 1(b) A3.4.1	Award [2] max Award [1] for the correct model and [1] for the justification. No mark for the model alone. A spatial database; It is designed to store geographic coordinates and shapes, and to answer questions about position, distance and area, such as which delivery zone an address falls inside or which store is nearest to a customer;	2
HL 1(c) A3.4.1	Award [2] max Award [1] for the point and [1] for the expansion. An in-memory database holds the data in RAM rather than on secondary storage; ... (so) the basket can be read and rewritten many times per second as the customer adds and removes items, without the delay of disk access; The basket is short-lived and does not have to survive indefinitely; ... (so) the loss of durability that comes with holding data in volatile memory is an acceptable trade for the speed;	2
HL 2(a) A3.4.2	Award [2] max Award [1] per point up to [2] max. The warehouse is organised around the business subjects that managers want to analyse, such as sales, customers or products; ... (rather than) around the individual applications or transactions that produced the data; Data about one subject is brought together from the tills, the website and the loyalty scheme into a single view of that subject;	2
HL 2(b)	Award [4] max Award [1] per point up to [4] max. Both properties must be addressed.	4

Q	ANSWER	MARKS
A3.4.2	Time-variant: every record carries a time dimension, so the warehouse holds the history rather than only the current position; Fifteen years of data allows managers to compare periods and identify trends and seasonal patterns; ...(whereas) the operational database usually holds only the current state of each record; Non-volatile: once data is loaded into the warehouse it is not changed or deleted by users, only appended to; ...(so) a report run today and the same report run next year return the same historical figures; This stability is what makes the warehouse trustworthy for analysis and for auditing;	
HL 3(a) A3.4.2	Award [4] max Award [1] per point up to [4] max. The two systems are optimised for different work: the tills need fast, small writes, while the warehouse needs large analytical reads; A long report scanning fifteen years of sales would slow the tills down or lock the records they need; ...(so) separating them protects the performance of the live checkout system, which cannot be allowed to fail; The warehouse can be structured and indexed purely for query speed, including denormalized structures that would be unsuitable operationally; The warehouse integrates data from the website and loyalty scheme as well as the tills, which no single operational system holds; Historical data can be kept in the warehouse indefinitely while the operational database stays small and fast;	4
HL 3(b) A3.4.2	Award [2] max Award [1] per point up to [2] max. Data arriving from different source systems is cleaned and converted into a single consistent format before it is stored; Inconsistent codes, units, date formats and naming are resolved, so the same product means the same thing everywhere; ...(so) figures from the tills and the website can be compared and combined meaningfully;	2
HL 4(a) A3.4.3	Award [3] max Award [1] per point up to [3] max. OLAP allows managers to analyse large volumes of historical data interactively; Data is organised into multiple dimensions, such as product, store, time and region; Managers can aggregate, drill down into more detail, roll up to a summary, slice by one dimension or pivot the view; ...(so) a question such as which product categories grew in the north last winter can be answered in seconds; This supports decisions on pricing, stock and store planning without needing a programmer to write a new report each time;	3
HL 4(b) A3.4.3	Award [2] max Award up to [1] per difference, [2] max. An OLAP query reads and aggregates very many historical rows, whereas a till transaction writes a few rows about one purchase; OLAP is read-only analysis of past data, whereas the till performs live updates that must be immediately durable;	2

Q	ANSWER	MARKS
	OLAP queries are complex and may run for seconds or minutes, whereas a till transaction must complete in a fraction of a second;	
HL 5(a) A3.4.3	Award [3] max Award [1] per point up to [3] max. The answer must be applied to the chain. Association rule discovery finds items that are frequently bought together in the same basket; It looks for patterns of the form: customers who buy A also tend to buy B; The chain could place associated products near each other, or on the same aisle end, to increase sales; It could target offers and website recommendations at customers who bought only one item of a pair; It could make sure associated items are stocked together, so a promotion on one does not leave the other unavailable;	3
HL 5(b) A3.4.3	Award [2] max Award [1] per point up to [2] max. Anomaly detection identifies records that differ markedly from the normal pattern in the data; These outliers are flagged for investigation rather than assumed to be errors; ...(for example) a till taking far more refunds than comparable tills, which might indicate fraud or a fault;	2
HL 6(a) A3.4.4	Award [4] max Award up to [2] per feature (point plus expansion), [4] max, for two different features. Replication: copies of the data are held in more than one region; ...(so) queries are answered from a nearby copy, reducing latency for each country; Data partitioning: the data is divided across sites, for example by region; ...(so) no single server has to hold or process everything, which improves scalability; Fault tolerance: if one site fails the others continue to serve requests; ...(so) the failure of one region does not take the whole chain offline; Location and distribution transparency: users and applications query the database as though it were one system; ...(so) they do not need to know which site physically holds the data; Concurrency control: the system manages simultaneous access across sites; ...(so) two regions cannot update the same record into an inconsistent state; Global query processing: a single query can be answered from data held at several sites;	4
HL 6(b) A3.4.4	Award [1] max Any one, [1] max. A write in one region takes time to reach the others, so copies can be briefly out of date; A network failure between regions can leave copies diverging with no way to reconcile them immediately; Maintaining the ACID guarantees across sites requires coordination that slows every transaction; Two regions may update the same record at once, creating a conflict that has to be resolved;	1

Total for this paper: 33 marks