IBM Unified Governance and Integration

Why it's time to offload your **enterprise data warehouse**

For decades, organisations have used the Enterprise Data Warehouse (EDW) for purposes that it was never initially intended, including running extraction, transformation and loading (ETL) workloads and storing large volumes of unused data.

In that time, the nature of data has changed, the practice of analytics has changed and the most efficient and cost-effective methods of storing and accessing data have changed.

EDWs continue to be an effective tool for handling trusted and clean integrated data for production analytics - such as reporting, historical comparisons, customer analysis and Key Performance Indicator (KPI) calculations. However, many EDWs are showing their age because they are:

1. Too expensive to run extraction, transformation and loading programs

- ETL workloads can consume 50 percent or more of the EDW processing capacity
- The more you run ETL in an EDW the less capacity there is for decision support applications.

2. Turning into data swamps

- As much as 50 percent of the data in an EDW is never used.[1]

3. Can't incorporate non-traditional data sources

- Social media data, click stream data, Internet of Things (IoT) sensor data require beyond row-column table structure.

4. Not for data exploration

- Requires heavy data modeling before data exploration limiting use cases such as data science discovery.

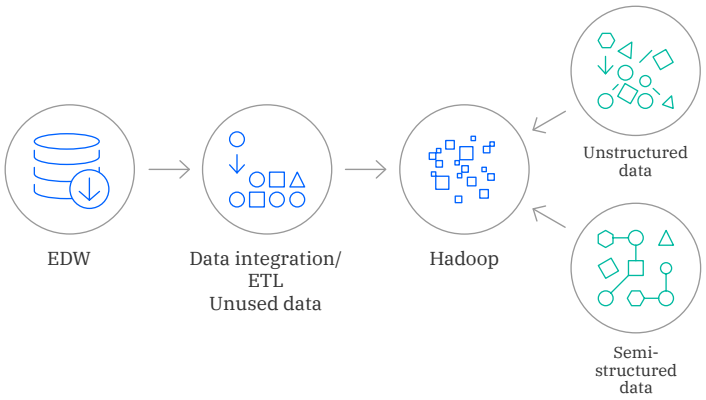
Modernise enterprise analytics

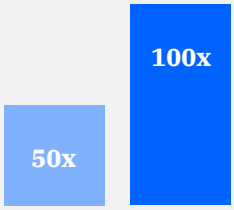
Infrastructure - whether roads, bridges or IT systems - needs to be updated to handle new demand. In the case of the data warehouse, modernising the enterprise analytics architecture represents an upgrade and dramatic shift in how data gets accessed, stored, prepared, governed and analysed.

One of the most effective modernisation approaches is to offload EDW data and ETL workloads to Apache Hadoop, which reduces costs while making it possible to effectively incorporate data integration, quality and governance.

Three separate, but related activities can be involved in EDW offloading

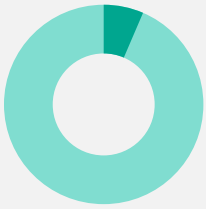
In each phase of the EDW offloading activities, organisations can ingest quality data into the data lake or Hadoop landing zone to improve data quality and help operationalise insights in the future.





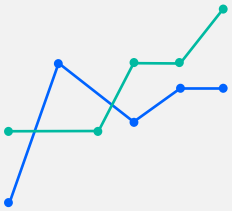
50x - 100x

reduction in processing and storage costs[2]



10x

productivity gain using tooling versus hand coding[3]

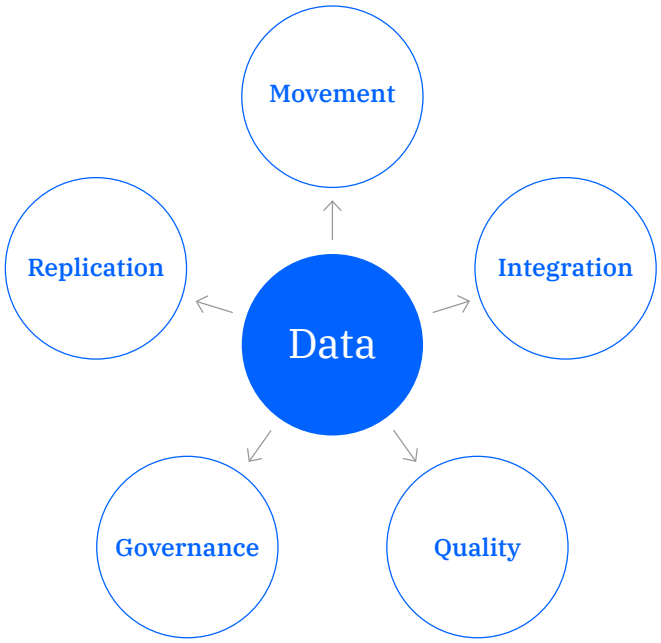


Enriched

analysis of non-traditional data

Five capabilities for effective offloading

All enterprises should consider five capabilities when creating an offloading strategy to maximise the Return on Investment (ROI) of EDW offloading.



Movement: Extract, move and ingest large volumes without having to land the data to disk and no new coding.

Replication: Achieve low latency data delivery from source systems to Hadoop targets to optimise resource utilisation while providing 'right time' updating.

Governance: Support data stewardship with stewards who: locate and retrieve information about data objects, their meaning, physical location, characteristics and usage through enterprise-adopted governance policies and a 'business vocabulary.'

Integration: Run hundreds of data integration processes that are built once and can run anywhere without modification.

Quality: Define validation rules that get applied consistently across multiple data sources; apply rules in-flight so only validated data is loaded into target systems; monitor and measure data sources to track and ensure compliance.

Read the blog

Learn about three considerations to take on your journey to a unified data platform.

Watch the webinar

Learn about clients who have modernised their analytics architecture to help solve the challenges of today's EDWs with the opportunities of offloading.

IBM United Kingdom Limited, PO Box 41, North Harbour, Portsmouth, Hampshire PO6 3AU, United Kingdom, IBM Ireland Limited, Oldbrook House, 24-32 Pembroke Road, Dublin 4.
IBM Ireland registered in Ireland under company number 16226. IBM, the IBM logo, ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at 'Copyright and trademark information' at ibm.com/legal/copytrade.shtml © Copyright IBM Corporation 2018

1 Forrester: The Next-Generation EDW Is The Big Data Warehouse (2016)
2 ITG: Business Case for Enterprise Data Integration Strategy: Comparing IBM InfoSphere Information Server and Open Source Tools (2013)
3 ITG: Business Case for Enterprise Data Integration Strategy: Comparing IBM InfoSphere Information Server and Open Source Tools (2013)