# TIME AND RESOURCE EXPENSIVE QUERIES TO EXTRACT LARGE DATA COMING FROM MULTIPLE BIG DATA SOURCES

- Extracting billions of records
- Data from multiple data sources
- Issues with interactive analysis
- Tedious query performance tuning
- Redundant summary tables

Challenges:

As enterprises turn towards data driven strategies, one of the challenges they face is to have a common business understanding of the data. Different departments in the same enterprise often establish different definitions for the same business processes or data artefacts. For example, what is "net sales"? One department might interpret it as net of invoice line-items and another department might interpret it as net of rebates. Without some level of abstraction or consistency, businesses are dependant to IT to generate and run reports or risk making big, costly and worst of all, hidden mistakes.
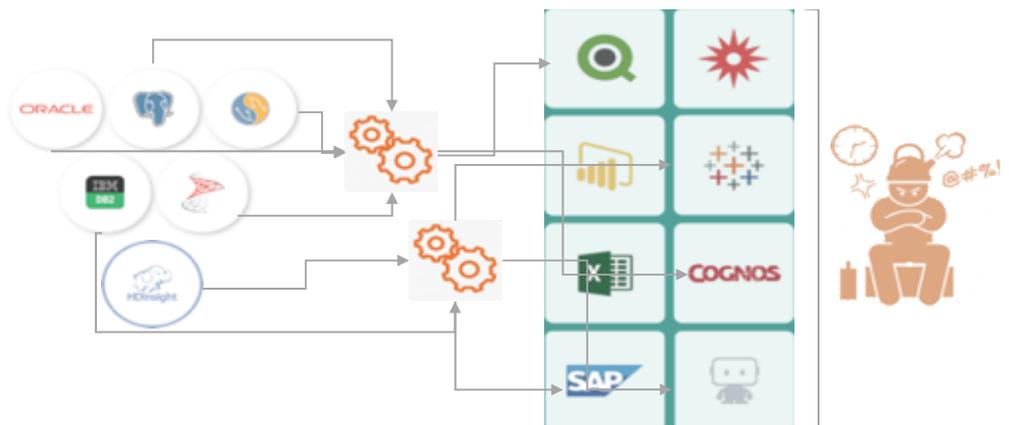
Solution:

In order to optimise query performance, enterprises should be looking to eliminate scripting and redundant summary tables. Queries that are constantly scanning raw data should be avoided to increase query throughput and concurrency. Enterprises should invest in defined strategies to manage massive data coming from multiple data sources and work with live data for business insights. Having consistent joins between tables and applying machine learning (ML) can reduce data query times from 2X to 1000X. Enterprises need a smart approach to embracing the cloud and significantly reduce data warehousing and query costs. Eliminating the need to copy or move data across the infrastructure can also significantly reduce the query cost. With query times shortened from days to hours or minutes, BI teams can deliver time-sensitive analytics effectively

- Increase query throughput and concurrency
- Autonomous data engineering approach to manage massive data sets
- Smart approach to migrate to cloud-based solution
- Reduce query cost by using aggregations

and efficiently, as an approach to data management that allows applications to use data without requiring technical details such as how it is structured, where it is physically located, or how the data is accessed.

How AtScale can help?

AtScale's autonomous data engineering ensures that, as queries are run against datasets in the adaptive analytics fabric (the virtualised data and associated tools to aid analytics speed, accuracy, and ease of use), machine learning is applied to determine what data within the larger set is needed.

AtScale's A3 platform optimises analytical workloads based on user behaviours, using artificial intelligence. AtScale monitors, analyses and optimises data queries based on how data is being used; learning what data is important and how best to shape that data into aggregates, providing conversational response time.

Because extraneous data is bypassed altogether during the query process, data is delivered between 2x and up to 1000x faster, depending on which data sets are queried. Queries that run against many terabytes of data can yield responses in tenths of a second from an AtScale aggregate.

Collectively, these aggregates are referred to as 'Accelerated Data Structures' because they constantly adapt to changes in user behaviour. With query times shortened from days to hours or minutes, BI teams can deliver time-sensitive analytics to the line of business effectively and efficiently.