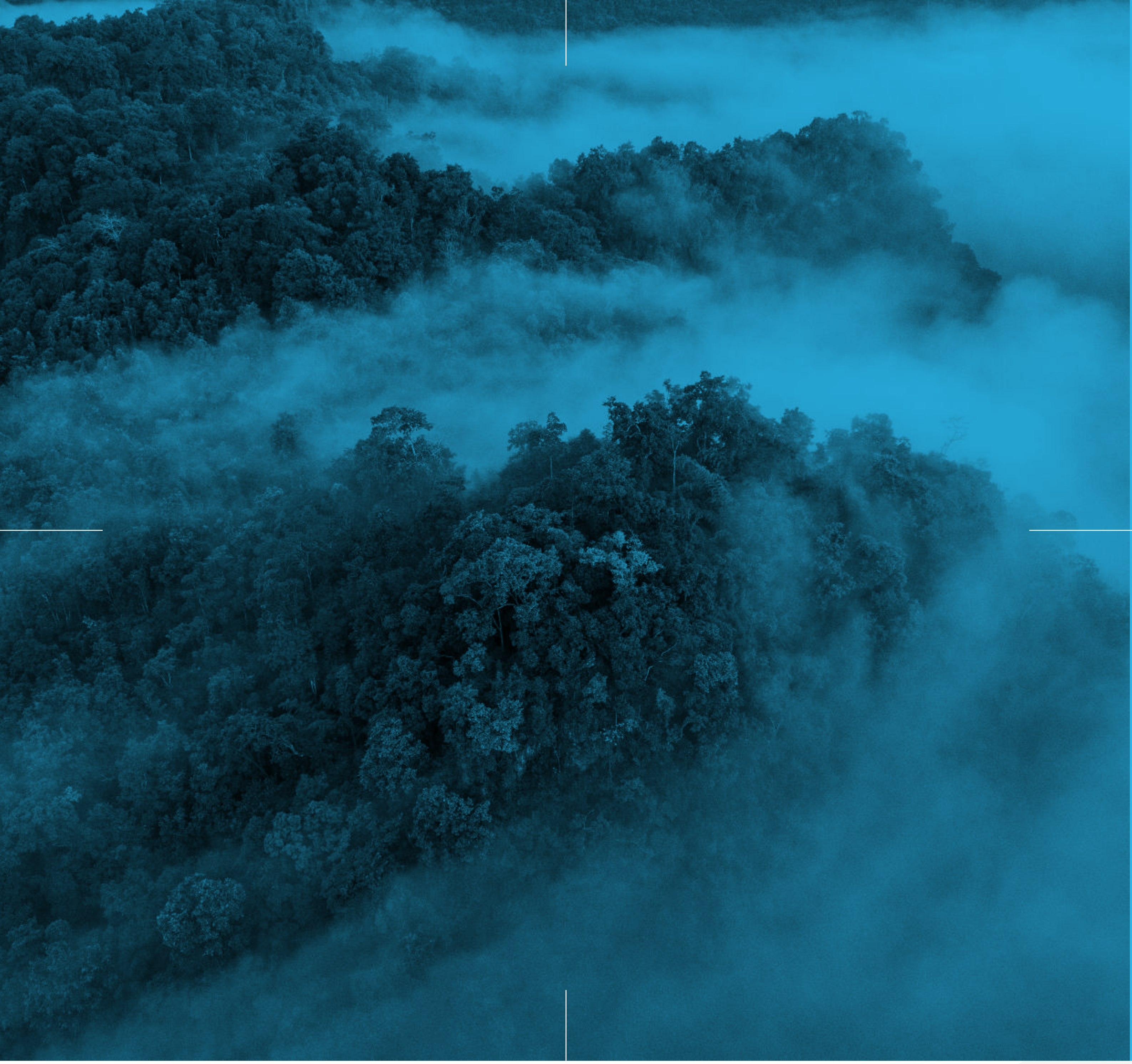


SNOWFLAKE & NET ZERO DATA

TOWARDS A RADICALLY MORE SUSTAINABLE SOLUTION FOR THE WORLD'S DATA

snowflake®

Why Net Zero matters

Boardrooms and shareholders are not only being made more acutely aware of the dangers of global climate change, but acting on it. In an annual letter to shareholders earlier this year, BlackRock CEO Larry Fink indicated that his firm would avoid investments in companies that "present a high sustainability-related risk", and put sustainability at the centre of its investment process. The following month, BlackRock publicly criticised German industrial conglomerate Siemens' management for supplying rail signalling systems to a controversial coal mine in Australia, contributing to a broader pressure which resulted in Siemens' pulling out of the agreement.

Managers of the world's largest corporations across industries are thus taking notice, and have almost uniformly made environmental sustainability a strategic priority for their firms. Certain industries have gone further than others, with energy and computing businesses having come under increasing scrutiny as a result of their energy-intensive nature and growth, respectively.

NET ZERO IN THE ENERGY AND COMPUTING INDUSTRIES

Major Oil & Gas (O&G) companies directly and indirectly account for 42% of global emissions, yet only spend 5% of their investment budgets in renewables and new-energy businesses. Shell has pledged to cut its' carbon footprint in half by 2050, and as a part of that effort has committed a $2 billion annual investment into renewables and cleantech. With the recent appointment of new CEO Bernard Looney, BP has announced its' ambition to achieve 'Net Zero' on carbon emissions in its operations on the same timescale, and is backing that up with remuneration policies which incentivise emissions reductions.

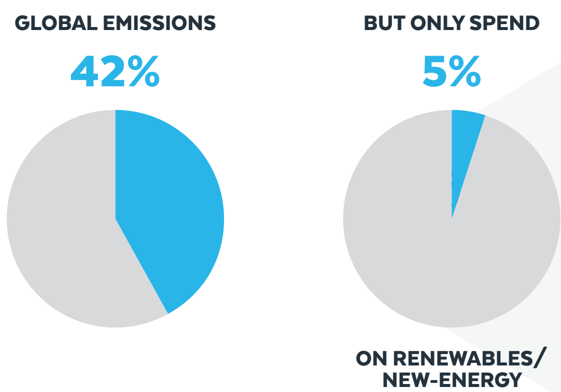
As the fastest-growing consumers of power and energy products, computing companies have also come under increased scrutiny and pressure. According to DXC, Datacentres consume an estimated 3% of global electricity supply, and generate 2% of global emissions -- more than the entire airline industry. The worst culprits are private corporate datacentres, which run very low server utilisation rates (10-15%) designed for peak capacity, and have largely not adopted the latest technologies or management practices to optimise for energy efficiency.

> "O&G Majors have an imminent opportunity to meaningfully advance their Net Zero ambitions by transforming their 'data operations'"

The major cloud service providers perform much better than corporate datacentres (typically achieving 65% server utilisation rates); AWS, Azure, and GCP are all on a path to carbon neutrality. That path has been enabled in majority by shifting energy inputs to low-carbon renewable sources, 'offsetting' emissions through investment into carbon sinks, and redesigning the management of datacentre operations to consume less energy per unit of computing power.

As the O&G Majors continue to decarbonise their core businesses, they will look not only at production and distribution operations as areas for improvement, but also at supporting functions and infrastructure. As large consumers of data storage and processing in both their core hydrocarbon and rapidly growing renewable energy businesses, O&G Majors have an imminent opportunity to meaningfully advance their Net Zero ambitions by transforming their 'data operations'-- from data processing, to data storage, to data analysis -- which support their business activities.

Major O&G companies account for



HOW NEW DATA PLATFORM ARCHITECTURAL DESIGNS CAN HELP ACHIEVE 'NET ZERO DATA'

There are meaningful energy, emissions and cost efficiencies to be gained from fully exploiting the flexibility of cloud computing. Unfortunately, the realisation of these benefits is often constrained by data platform architectures designed for fixed capacity environments. We've observed this to be the case with all of the 'native' data warehousing services offered by the large Cloud Services Providers (CSPs). AWS Redshift, for example, is built on top of what was the ParAccel Analytic Database (PAD), an on-premise database technology which AWS acquired in 2012 and retrofitted to make available as a cloud service. To this day, however, the core architectural 'shared-nothing' design choice underpinning the Redshift technology requires a proportional scaling of data warehouse units and compute clusters. The same goes for Microsoft Azure SQL Data Warehouse (DW), which was similarly built on top of the technology from Microsoft's DATAllegro acquisition in 2008, and existed as the on-premise Microsoft Analytics Platform System prior to the launch of Azure.

These design choices were bleeding-edge at the time that they were made, and enabled the use of massive parallel processing (MPP) computing techniques required to handle the proliferation of 'Big Data' datasets. In today's context, however, those very same design choices have become constraints in how efficiently services including Redshift and Azure SQL DW can be deployed with respect to the utilisation of compute resources. And a less efficient use of CPU has a knock-on effect on the energy intensity, emissions footprint, and cost of data operations. As mentioned previously, this isn't to say that the increased rate of server utilisation which comes from running on co-located resources in the cloud does not have a positive impact on emissions. Rather, there are additional meaningful gains to be had by employing a modern, multi-cluster, shared-data architecture built for the cloud. Specifically:

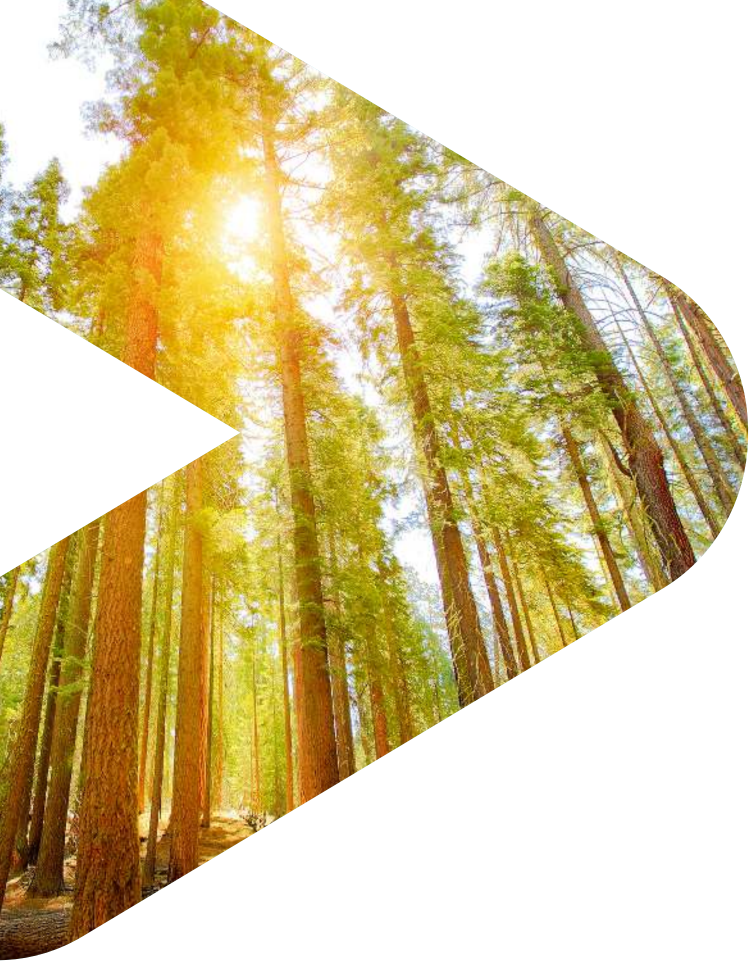
- **Eliminating the need to transform and process Big Data datasets.** The analytics databases native to the large CSP suites require the transformation of raw semi-structured data files into a traditional columnar structure to be ready for analytical workloads. But new methods for partitioning and indexing highly-compressed, cost efficient data formats -- like JSON or Avro -- now allows for the full range of traditional enterprise-grade SQL to be used to query that data such that the data structure, or schema, is determined 'on-read'. This eliminates entirely the energy, emission, and cost associated with the compute power needed to transform semi-structured, machine-generated datasets which are so common in the energy industry; for example, seismic data in subsurface exploration, IoT in upstream fossil fuel production, refinery, and renewable electricity generation, time-series or other market data in energy trading, and transaction data in forecourt and convenience retail.

- **Eliminating the need to store multiple forms and copies of the same data.** By eliminating the need to transform all of the aforementioned semi-structured data to be ready for analytics, we simultaneously eliminate the need to store pre- and post-processed forms of those datasets. Indeed, the concept of pre- and post-processing is replaced entirely by a single, unified data environment across not only semi-structured, but structured data as well. By creating software-defined, functional 'views' on top of that single data environment, we can additionally eliminate the need to make copies of the same data for different use cases; for example, data engineering, data science, business intelligence reporting, financial and regulatory compliance reporting, or ad hoc analysis. And fewer copies of the same data means less disk space required to store that data, thus reducing the energy, emissions, and cost requirements of data operations.

"a less efficient use of CPU has a knock-on effect on the energy intensity, emissions footprint, and cost of data operations"

- **Radically reducing the CPU capacity required to run a global, enterprise-grade analytics platform.** The 'shared-nothing' architecture underpinning the CSP's native analytics database offerings requires a one-to-one scaling of compute clusters and database instances, where the number of nodes in any single cluster has a fixed upper bound. Delivering the 24x7 speed and performance of analytical workloads required by a large, complex global organisation thus requires full-time availability of 'peak capacity' CPU resources to avoid, for example, latency associated with boot-up, or concurrent users. Through a 'shared-data' architecture, however, we can achieve a complete logical decoupling of storage and compute to provision resources per second of actual CPU usage, and eliminate the need to keep machines idling on stand-by while realising improved performance. No machines on stand-by means radically less energy, emissions, and cost attributed to data.

- **Compounding the effects of more efficient datacentre design and management.** There are large energy and emissions gains to be had from other, non-software-related improvements in the design and management of datacentres. AWS is contracting more and more power from renewable energy producers, including from Lightsource BP's solar energy plants. Microsoft is experimenting with sinking datacentres into the ocean to reduce the energy associated with cooling servers. In contrast to many on-premise corporate datacentres, CSPs run their cloud operations as profit centres, and the profit motive has naturally accelerated the development of structural datacentre designs which yield a step-change in energy and cost efficiency. Optimisations made at the database architecture level are likely to compound the effects of more efficient datacentre designs, although further research in this area is required to understand to what order of magnitude this holds true.



MAPPING THE IMMEDIATE PRACTICAL APPLICATIONS FOR 'NET ZERO DATA': THE CASE AT BP

In exploring how to initiate a practical exploration of the impact that an optimal analytical database architecture can have on BP's Net Zero strategy, let's consider a few particularly data-intensive areas of BP's operations:

- **Subsurface imaging and upstream exploration analysis.** BP employs the world's largest commercial supercomputer and Microsoft Azure's AI suite to locate new hydrocarbon reservoirs with subsurface imaging, estimate how much can be retrieved from those reservoirs, and then calculate the optimal path to extraction by simulating different drilling scenarios. The exploration process leverages enormous geospatial datasets and intensive analytical methods which are likely targets for energy and emissions reduction.

- **Digital twinning of upstream production sites.** Broadly, the O&G industry has adopted the concept of 'digital twinning' to monitor well-site production facilities, equipment, and operations and prevent unplanned downtime. BP pioneered it's Plant Operations Advisor in collaboration with Baker Hughes GE in the Gulf of Mexico and has been rolling it out to upstream assets worldwide, with the expectation that it will yield a 2% increase in operational efficiency. The digital twin system is enabled by large amounts of operational data coming off of well-site sensors -- the capture, transformation, movement, and analysis of which are all potential targets for cleaner computing.

- **Methane emissions leak monitoring of production, refinery, and shipping operations.** From rigs to refineries, and ships to forecourts, the O&G industry is increasingly adopting advanced analytical techniques to monitor and address undesirable emissions leaks across the value chain, especially of methane. BP has designed a comprehensive programme for methane leak detection using Gas Cloud Imaging enabled by drones mounted with on-board sensors, and spectral radiometry to monitor the emissions consumption efficiency of flaring. The various aspects of the methane detection programme are underpinned by the accurate, timely streaming of data from the field, and are additional target areas for reducing the emissions impact of computing operations.

- **Supply and trading of energy products and related instruments.** Most O&G Majors operate trading businesses specialised in buying and selling of energy products, their derivatives, and related financial instruments which help with hedging against fluctuations in currencies, interest rates, and other sources of volatility. BP's energy trading business is to forecast market supply and demand dynamics for energy products, and can be enabled by the aggregation of large internal datasets -- from operational data coming off of well-sites, to refineries, to forecourts -- as well as external datasets representing everything from weather patterns, to government-issued gas and power market data, to maritime and trade port information. The acquisition and analysis of this diverse set of data used to support trading decisions represent meaningful areas to reduce the energy and emissions footprint of BP's computing operations.

> "the capture, transformation, movement, and analysis of data are all potential targets for cleaner computing"

- **Marketing and distribution of retail and commercial end-products.** BP's portfolio of fuels, lubricants, electricity and convenience products, for both B2B and B2C customers, necessitates a complex distribution network and marketing infrastructure to ensure that the right product is offered to the right customer at the right time. The BPme digital rewards programme, for example, collects data about customers as they transact at forecourts which is used to both reward customer loyalty and personalise marketing promotions. A similar opportunity exists to drive customer loyalty and personalise offers from data collected through the Polar Plus app at BP's EV network operator, BP Chargemaster. Collecting, storing, and analysing these large customer transaction and loyalty datasets represents yet another opportunity to optimise data and analytical operations for energy and emissions efficiency.
- **Optimising the asset and operations management practices of solar and wind power plants.** As with any complex infrastructure, the management of gas and renewable power plants alike are increasingly being fitted out with large amounts of sensor-based IoT monitoring equipment to collect accurate, timely information on hardware performance and ambient conditions. That data is then used to reduce downtime and extend the lifetime value of power plant assets through an operational management practice increasingly known as 'predictive maintenance'. The core of the practice is focused on preempting failures by automating the requisition of replacement parts and engineering labour. Data volumes can be unprecedentedly large, with tens of millions of sensors passing information every second, and thus presents an opportunity to reduce the energy, emissions, and cost of data operations.

WHAT DOES THIS MEAN FOR THE WORLD'S DATA?

59M METRIC TONNES OF CARBON

(POTENTIAL CO2 EMISSIONS REDUCTION FROM ADOPTING PUBLIC CLOUD)

"there are meaningful energy, emissions and cost efficiencies to be gained from fully exploiting the flexibility of cloud computing"

SNOWFLAKE IS HERE TO HELP

In partnership with AWS and Microsoft Azure, Snowflake has developed an analytical database technology and management practice which we believe can enable the world's largest organisations to realise the full potential that cloud computing can have on the energy intensity and carbon footprint of it's enterprise data platforms.

Moreover, with the recent contraction of global economic activity putting unprecedented cost pressure on many traditional industries, we believe that Snowflake can simultaneously enable many to realise necessary cost savings on technology infrastructure, as well as cost avoidance on labour required to maintain IT systems' complexity.

There is an imminent and deeply meaningful conversation to initiate with Snowflake on the sustainability of business today. We look forward to having it.

ZAC AGHION

Sales Director, **Snowflake**

Zac advises leaders of Fortune 500 companies on strategies at the intersection of growth, data, and sustainability.

ABOUT SNOWFLAKE

Snowflake delivers the Data Cloud — a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency, and performance. Inside the Data Cloud, organizations unite their siloed data, easily discover and securely share governed data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single and seamless experience across multiple public clouds. Snowflake's platform is the engine that powers and provides access to the Data Cloud, creating a solution for data warehousing, data lakes, data engineering, data science, data application development, and data sharing. Join Snowflake customers, partners, and data providers already taking their businesses to new frontiers in the Data Cloud.

snowflake.com