We’re a people led data and analytics strategy company
Our mission is to contribute to a better future for all through the open and positive use of data | We do this through 4 service pillars

- **Advise**: Craft and shape the future through strategic advise and leadership
- **Design and Build**: Apply cutting edge technology solutions to solve business problems
- **Maximise**: Drive business outcomes and support business change
- **Grow**: Help individuals realize their personal and professional potential
We’ve collaborated with over 60 market leading, forward thinking organisations and delivered strong business outcomes across industries.
An introduction to data platforms.
Analysts and data scientists don’t work in isolation.
Cross functional teams

From this

- Capability aligned
- Centres of Excellence
- Reporting lines to capability lead
- Sitting in different departments
- Hand offs between capabilities

To this

- Outcome aligned
- Centres of Outcome
- Reporting lines to business owner
- Acting as virtual team to drive outcome
- Mix of capabilities working together
What is a data platform?

- Data Warehouse
- Data management platform
- Data Lake
- Data Mart
- Datastore
- CRM
- Customer insights platform
- Big Data
- Enterprise data platform
Data platforms

Ingest and process

Sources

Store

Factory

Lab

RDBMS
Distributed systems
NoSQL
Geospatial
Graph
Timeseries

Provision
Integrate
Report
Analyse

DataOps and Governance
Data lives in lots of source systems across the business, in various shapes and quality. Data ingestion brings that data into the data platform, without touching it or changing it. The data lake holds that raw data, whatever its format, building an immutable archive of data. Data transformation cleans and processes the data into formats that are easy to use by the business. The data warehouse is the store for that output, becoming the s|three's best view of its data. The data warehouse is the store for that output, becoming the s|three's best view of its data.

**DataOps**

DataOps tools support the delivery of data and are used by the data team to make sure that data quality is preserved as data flows through the stages of the data pipeline.

**Data management**

Data management tools help the s|three govern and control its data. This can range from helping manage master lists of data to documenting the meaning of different data and where it can be found.
Data lakes

- All your raw data held in cheap, scalable storage
- Distributed file store
- Separate processing and compute
- Parquet / Avro / ORC / Sequence files
- Schema on read vs schema on write
- Common pitfalls of data lakes

Azure Data Lake Storage
Amazon S3
Hadoop
Google Cloud Storage
Data warehouses

- Centralised, cleaned and modelled data
- Data marts
- OLAP vs OLTP
- RDBMS & Distributed processing
- Columnar storage & partitioning
- Software as a service & infrastructure as a service
- Virtualisation

Data Lake → Data warehouse
Data warehouse models

Star schemas

Data Vault

Conceptual

Logical

Physical
ETL vs ELT

ETL Process
- Extract
- Transform
- Load
- Analyze

ELT Process
- Extract
- Load
- Transform
- Analyze

Fivetran, dbt, talend, Azure Data Factory, databricks
Real-time streaming

- Event driven architectures
- Micro-batch
- Message queues and message brokers
- Real time processing
Reporting and BI

- Reports
- Self service reporting
- Dashboards
DataOps

- Data products
- Infrastructure as code (IaC)
- Version control
- Automated testing
- CI/CD
- Orchestration
- Serverless and containers
Data science

- Auto ML
- MLOps
- Commodity models
- Model serving