

## REA Data Platform



**Surj Rangi** Architect



**Piyush Dhir** Senior Tech Lead



Matt Ellis Tech Lead



# Agenda

- **1.** REA Group history
- 2. Who are Tech & Data
- 3. Data Platform our approach
- 4. Data Management our approach
- **5.** Q/A



## Who is REA Group?

We're a leading global digital business working to change the way the world experiences property

3000+ people over 3 continents

REALTAIR

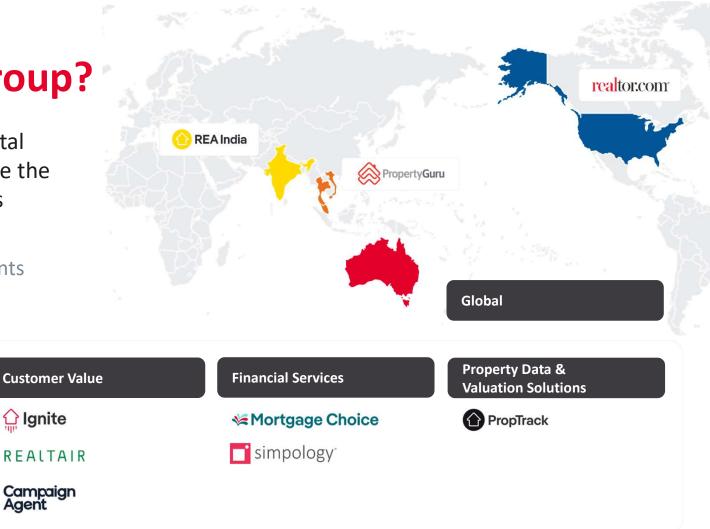
**Consumer Experiences** 

nealestate.com.au

Property /

nealcommercial.com.au

**AUSTRALIA** 







Who are Tech & Data



## Tech & Data – making it easy to understand, protect and leverage REA data

Mature data platform, scaled adoption org-wide, driving high quality real-time

data ready for the next wave of growth



#### **Centralised Platform**

Batch and Real-time data processing across multi-cloud platforms (AWS/GCP)



#### **Team Topology**

Business Partners, Federation, Forums and Governance to drive adoption



#### **Strategic Demand**

Partner with LOB, with clear strategies and initiative alignment



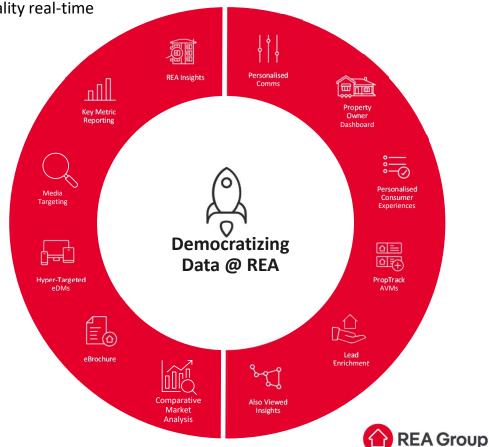
#### **Tech Trends**

Explore, "practice what we preach", evaluate and embracing modern tech and Al for the right purpose



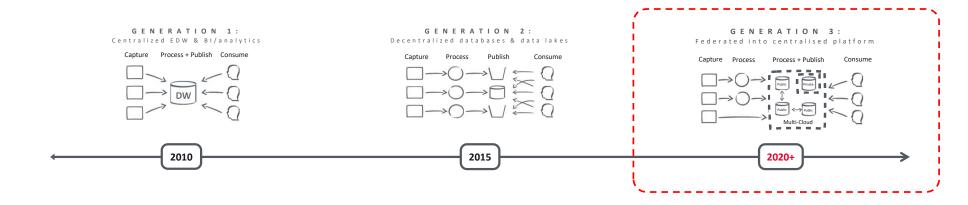
#### **Data Health**

Survey and listen to our platform users



## Tech & Data – the journey so far

Data Platform Evolution over the last decade+

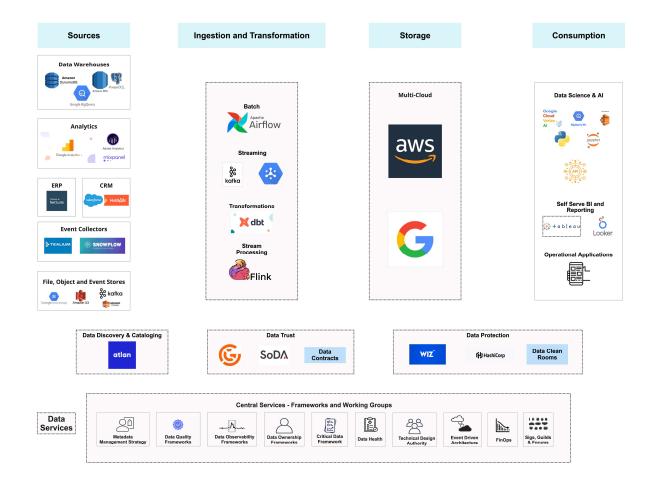








## **REA – Conceptual Data Platform Architecture**





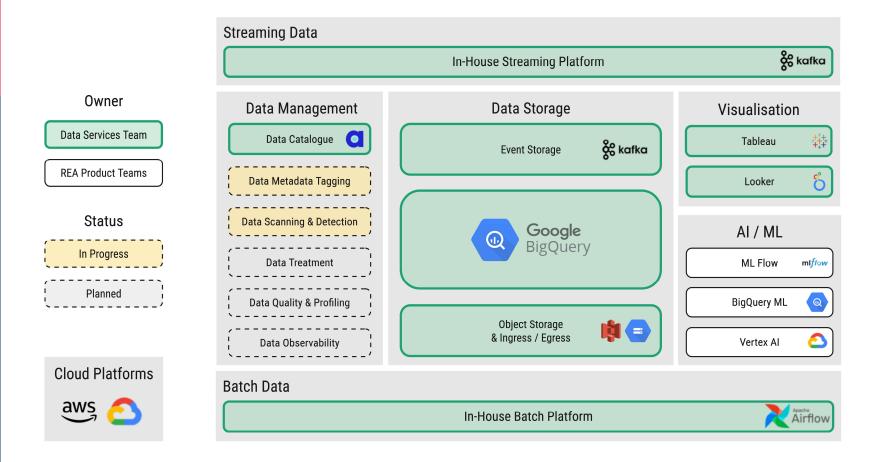


Data Platform – our approach



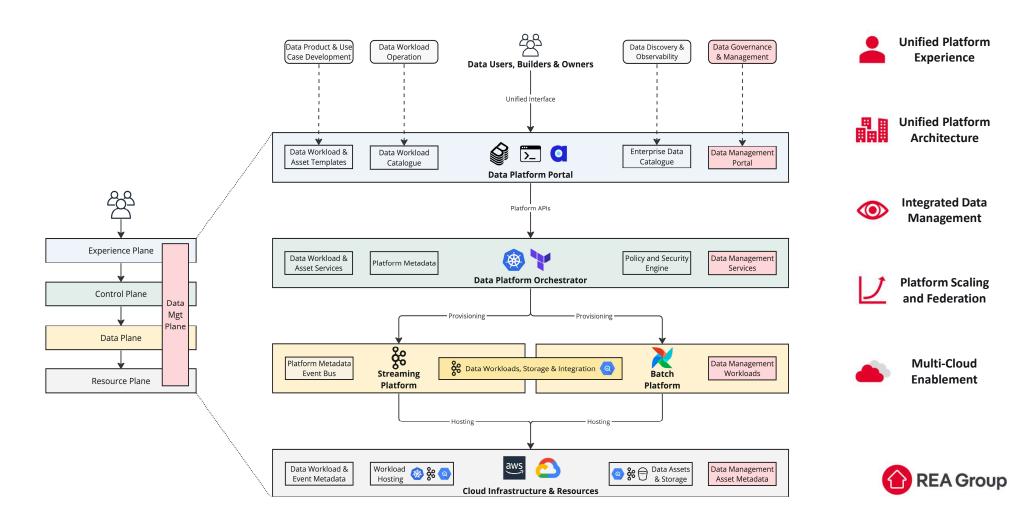
#### **Data Platform – Current Estate**

**In-House Platform Products for Batch, Streaming & Data Management** 





#### **Data Platform - Unified Architecture**



**Case Study** 

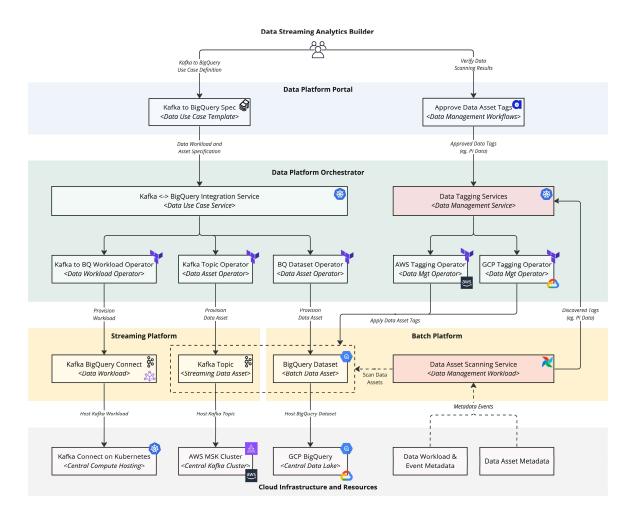
Experience Plane

Control Plane

Data Plane

Resource Plane

Data Management Plane





Streaming event data stored in Kafka topic

Streaming analytics in GCP BigQuery

Data may contain personal information



**Case Study** 

Data & workload use case templates

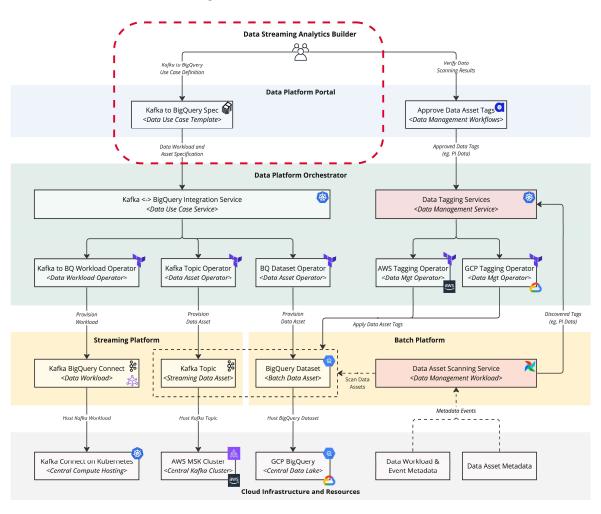
Experience Plane

Control Plane

Data Plane

Resource Plane

Data Management Plane





**Self-service** resource definition and ownership

Golden paths for data use cases

**Democratised** access via platform portals



**Case Study Data Streaming Analytics Builder** Verify Data Kafka to BigQuery Data Platform Portal Approve Data Asset Tags Kafka to BigQuery Spec Data & workload <Data Use Case Template> <Data Management Workflows> Approved Data Tags Data Workload and Asset Specification **Data Platform Orchestrator** (\*) Kafka <-> BigQuery Integration Service Data Tagging Services High level workload & data abstractions <Data Use Case Service> <Data Management Service> Experience Plane Control Plane Kafka to BO Workload Operator Kafka Topic Operator BO Dataset Operator AWS Tagging Operator GCP Tagging Operator Data Plane <Data Workload Operator> <Data Asset Operator> <Data Asset Operator> <Data Mgt Operator>\_ <Data Mgt Operator> Resource Plane Discovered Tags Data Management Plane Data Asset Data Asset Streaming Platform **Batch Platform** Kafka Topic Kafka BigQuery Connect Data Asset Scanning Service eployed workloads & infrastructure BigQuery Dataset <Data Workload> <Streaming Data Asset> <Batch Data Asset> <Data Management Workload> Scan Data Metadata Events Host Kafka Workload Host Kafka Topic Host BigQuery Dataset Kafka Connect on Kubernetes AWS MSK Cluster GCP BigQuery Data Workload & Data Asset Metadata orkloads & storage <Central Compute Hosting> Central Kafka Cluster> <Central Data Lake> Event Metadata Cloud Infrastructure and Resources



**Orchestration** via REA's Platform as a Service

**Encapsulation** of data use cases and resources

**Composable** services, assembled into experiences



**Case Study Data Streaming Analytics Builder** Verify Data Kafka to BigQuery **Data Platform Portal** Approve Data Asset Tags Kafka to BigQuery Spec Data & workload use case template <Data Use Case Template> <Data Management Workflows> **Integrated Data** Data Workload and Approved Data Tags Asset Specification Management **Data Platform Orchestrator Cross-cutting services** High level workload & data abstractions Kafka <-> BigQuery Integration Service **Data Tagging Services** Shared Data <Data Use Case Service> <Data Management Service> and enforcement Experience Plane Common Control Plane metadata layer Data Plane Kafka to BQ Workload Operator Kafka Topic Operator BQ Dataset Operator AWS Tagging Operator GCP Tagging Operator <Data Workload Operator> <Data Asset Operator> <Data Asset Operator> <Data Mgt Operator> <Data Mgt Operator> Data management process orchestration Resource Plane Discovered Tags (eg. Pl Data) Data mgt workloads writing to Prod State Data Management Plane Data Asset Data Asset Apply Data Asset Tags Streaming Platform **Batch Platform** Kafka Topic Kafka BigQuery Connect BigQuery Dataset Data Asset Scanning Service eployed workloads & infrastructure <Data Workload> <Streaming Data Asset> <Batch Data Asset> <Data Management Workload> Scan Data Metadata Events Host Kafka Workload Host Kafka Topic Host BigOuery Dataset Kafka Connect on Kubernetes AWS MSK Cluster Data Workload & GCP BigQuery Platform metadata layer at source Data Asset Metadata <Central Compute Hosting> <Central Kafka Cluster> Event Metadata <Central Data Lake> Cloud Infrastructure and Resources





Data
Management &
Governance – our
approach



#### **Driving Excellence in Data Management & Governance**

"At REA, we are constantly investing and prioritizing data management and governance as part of our privacy-first approach within the organization."



#### **Data Quality Standards**

- o Establish robust *data quality frameworks*
- o Enable enterprise tooling



#### **Data Management Processes and Frameworks**

- o Metadata *tagging strategy* for the enterprise
- o Developing *PII framework and Data Treatment process* to ensure privacy compliance



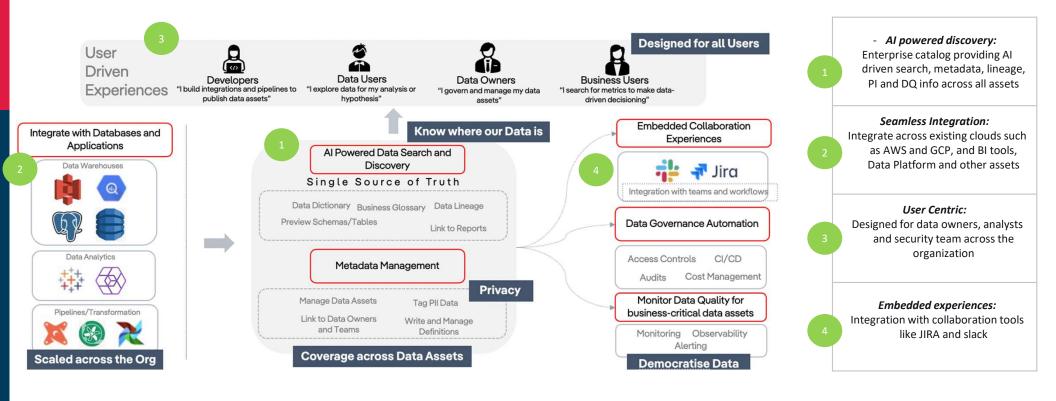
#### **Data Ownership and Accountability**

- o Creating *accountability framework* for data owners
- o Implementing *FinOps strategies* for effective data management



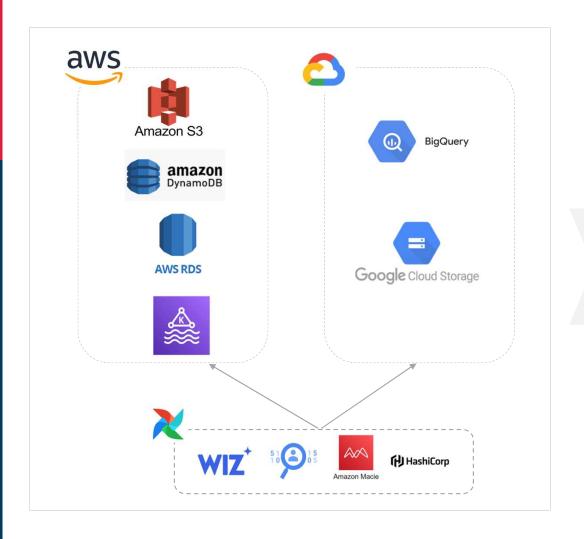
### **Enterprise Data Catalogue**







### **Data Scanning and Treatment**



#### **Overview**

Strategic investments to build and mature *data security and privacy platform capabilities* across the enterprise

Focused on discovering *Personal sensitive information and treating it appropriately* based on the frameworks/processes and methodologies set up by security and governance

Striking the right balance of compliance with *data privacy and security while minimizing operational costs* 

Data Treatment & Scanning capability will be integrated on the *existing data platform*, this capability will be available to *all users across the enterprise*.



## **Lasting thoughts**







+



+





Unified Data
Platform capability

Commitment to Security and Data Privacy Support for Analytics, ML and AI use cases High grade Data Products High Value Platform & Data Landscape

"Our focus on technology and innovation has been constant and the driving force behind our success"





Q&A