Data Governance isn't a _Dirty Word_

DataCon Africa 2022

Dr. Sizwe Gwala

Head: Data Governance

Absa Group Compliance

Data, a valuable business asset

Data-Heavy Era

- More that **2 quintillion** of data is generated by businesses, daily.

- Unstructured & semi-structured data accounts for 80%;

  - *Mobile devices; Applications; Wearables* and the *Internet of Things (IoT)*.

- Approx.. 81% of companies have a **mix of legacy** and **modern cloud technologies**.

- Despite the importance of data, **only 14% of** companies have it widely & easily accessible to employees.

The V⁴ Data Headache

- **Volume** makes finding trusted data difficult

- **Variety** makes lineage & transformation tough to track

- **Veracity** (or accuracy) is time-consuming to verify

- **Velocity,** the high rate at which data is accumulated

Professionally manage data as a strategic asset

Drive data-centric & data-driven businesses

Move towards value-driven data governance

Embrace a Data Citizenship culture

Adopt flexible, & agile data management solutions

It's estimated that the global datasphere will grow to 175 zettabytes by 2025.

Existing Status Quo

Is Data Governance really a dirty word?

Predictive Systems to **AI-driven automation** applied, most problems still solved through **traditional approaches.**

Though the proliferation of data is driven by **unstructured or semi-structured data,** most **usable data** is still organized in a **structured fashion** using r**elational database** tools.

Data privacy & security are perceived as **compliance issues,** driven by **nascent** regulatory data-protection mandates

Data engineers **manually exploring datasets**, establishing relationships among them, and joining them together.

- Data Teams managing data using **top-down standards, rules, and controls**.
- Data **has no true ownership.**

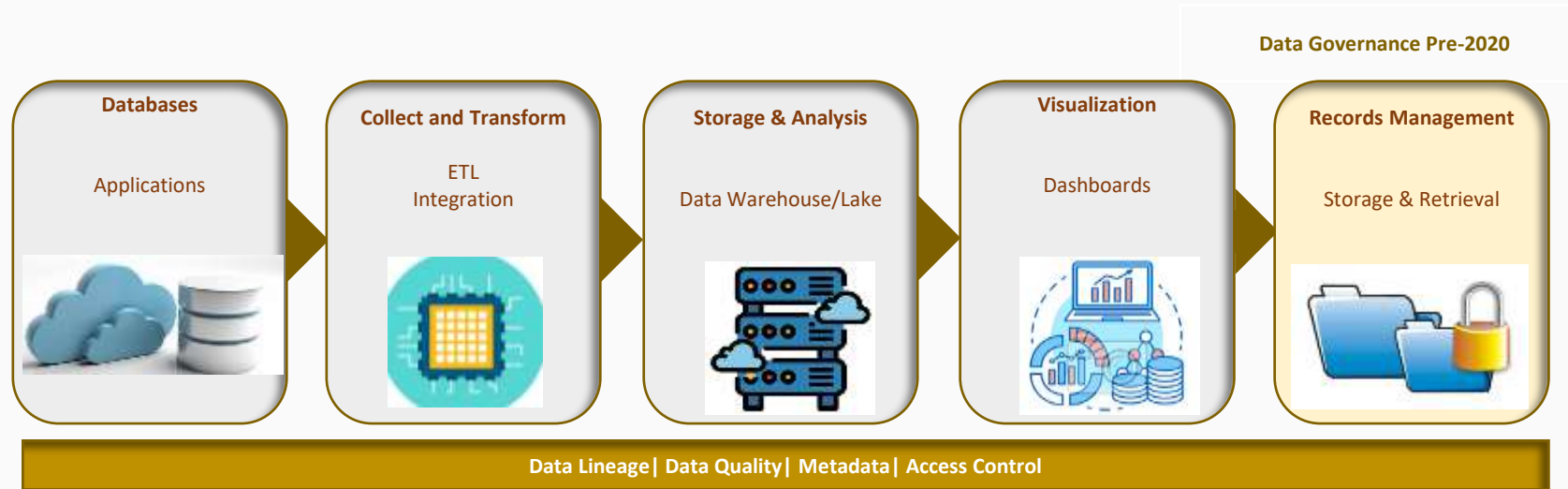Data is stored in **duplication, across siloed**, and often costly environments.
Making it **difficult for users** to quickly find, access, and integrate the data they need.

Only a fraction of data ingested, processed and analyzed in real time due to;
- Limits of **legacy technology** structures
- Hindrances of **adopting more modern architectural** designs.
- **High computational demands** of intensive, real-time processing jobs.

Data Governance & Management

DATA GOVERNANCE		DATA MANAGEMENT
*Data governance (DG) is the process of managing the **availability, usability, integrity and security** of data in enterprise systems, in line with internal data standards and policies.*		*Data Management (DM) is the practice of **collecting, organizing, protecting, and storing** an organization's data so it can be analyzed for business decisions.*
DG refers to policies, rules and controls to governing data and managing data quality.		DM refers to actual methods on how data is organized, stored and handled.
Refers to strategic practices, process and theories.		Refers to collecting, organizing, protecting, processing, sorting and maintaining data.
Focuses on the processes for ensuring data trustworthiness and integrity.		Focuses on ensuring that high data quality is achieved and preserved.
It is philosophical and focuses on an overall business strategy.		It is logistical and focuses on data management technologies.
Defines how data is accessed and treated within a broader data strategy.		The implementation of architectures, tools and processes to achieve data governance objectives.

Data Pipelines & Ecosystems: Pre 2020

Data Governance Pre-2020

Databases	Collect and Transform	Storage & Analysis	Visualization	Records Management
Applications	ETL Integration	Data Warehouse/Lake	Dashboards	Storage & Retrieval

Data Lineage | Data Quality | Metadata | Access Control

- Basic data pipelines- Manual interfaces & unclear accountabilities

- Manual report vetting, setting up custom rules, and comparing numbers side by side.

- *Common business requirements*: Access to static historical financial reports for analysis purposes.

- While technologies enabling data to become more accessible and easier to use were innovating at an unprecedented pace, tools and processes for ensuring data was easily discoverable and reliable were not catching up.

  - DG&M was through manual, outdated, and ad hoc tooling.

- While data catalogs emerged as effective governance approaches, most lacked rudimentary aspects when it comes to manual controls.

- As the amount of data sources increased and tech stacks become more complex, this approach was neither scalable nor efficient.

- As companies migrated to more distributed architectures (i.e., the data mesh), the need for ubiquitous and end-to-end governance has never been greater.
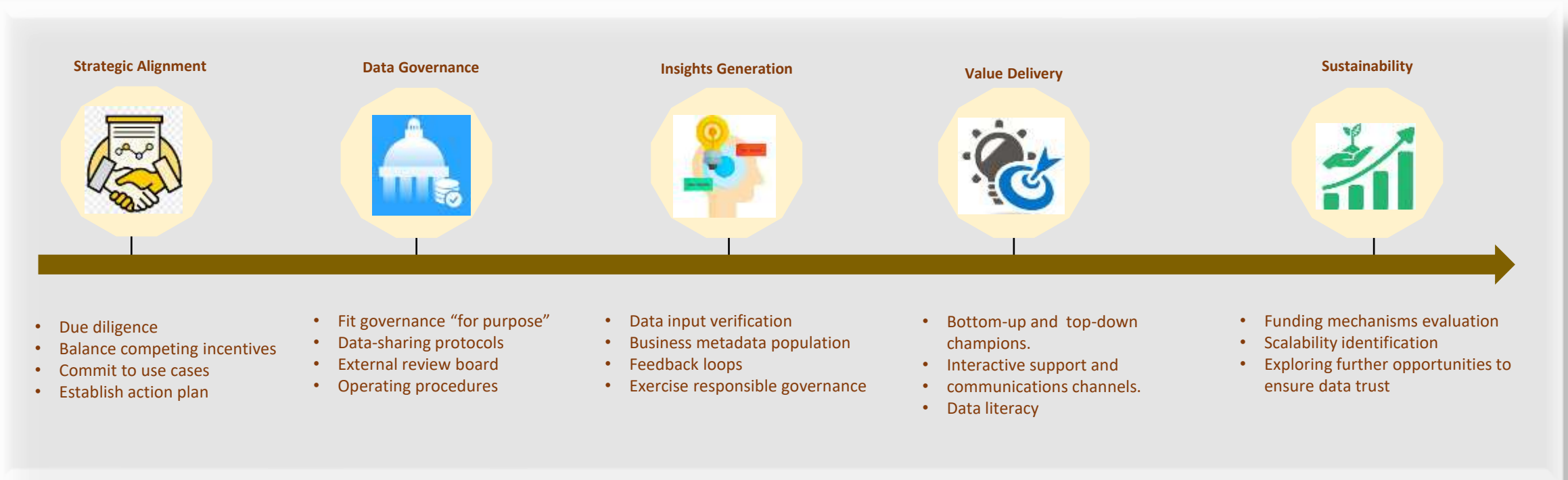
Cause for change

- Data Tech-stack complexities called for a change in DG approach, industry-wide.

- While data infrastructure and BI tools advanced, DataOps lagged far behind with most *DM solutions remaining manual, one-dimensional, and unscalable.*

- With most people using data and building reports/dashboards/new datasets, most organizations ended up with **misaligned numbers** across different departments.

- Most business teams were confronted with **data access complexities** alongside ensuring that data was used in the right manner.

- Data pipelines ran with a combination of **complex tools,** and as the number of interconnected parts increased, so did the **risk of pipeline failure**.

- Most businesses increasingly engaged in *Operational Analytics* – an approach of making data accessible to "operational" teams, for operational use cases (*sales, marketing*).

- Organizations are still increasingly democratizing data, ensuring everyone can access the data they need, whenever they need it.

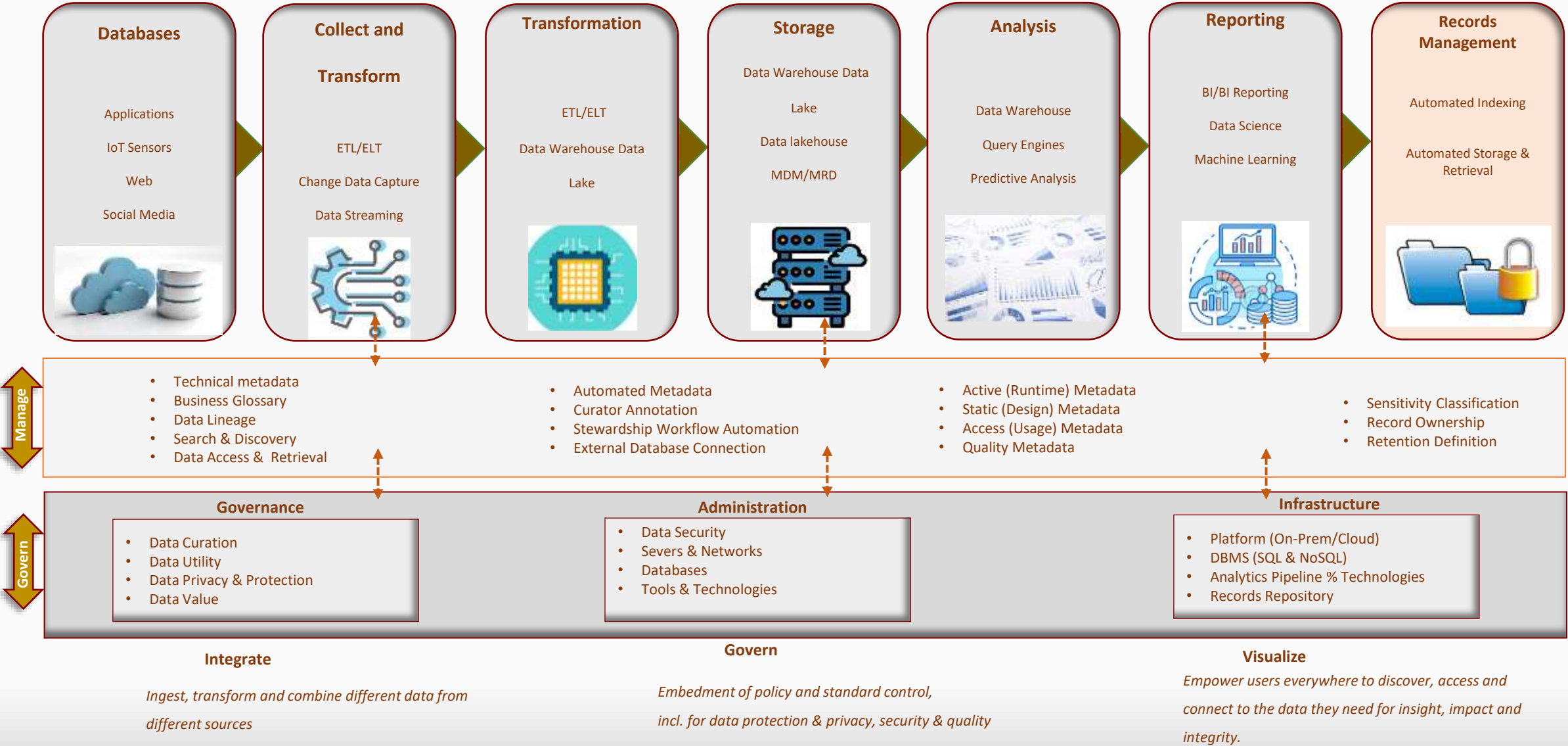*More modernized DG approaches were and are still required by businesses, industry-wide*

Minimum criteria for value-driven DG

Strategic Alignment	Data Governance	Insights Generation	Value Delivery	Sustainability

- Due diligence
- Balance competing incentives
- Commit to use cases
- Establish action plan

- Fit governance "for purpose"
- Data-sharing protocols
- External review board
- Operating procedures

- Data input verification
- Business metadata population
- Feedback loops
- Exercise responsible governance

- Bottom-up and top-down champions.
- Interactive support and communications channels.
- Data literacy

- Funding mechanisms evaluation
- Scalability identification
- Exploring further opportunities to ensure data trust

- Modern day businesses require governance approaches that believe in treating **data as a strategic asset** – & are focused on adding value onto data for colleagues, partners and beneficiaries, in a responsible manner.

- This means **orchestrating people, processes & technology** for ensuring users manage data based on defined best practice principles.

- Success lies on **DG focusing solely on stakeholder value creation** , in ways that with respect for rights, security, accountability & data quality.

Modern Data Pipelines & Ecosystems

Data Governance Post-2020

Databases	Collect and Transform	Transformation	Storage	Analysis	Reporting	Records Management
Applications	ETL/ELT	ETL/ELT	Data Warehouse Data	Data Warehouse	BI/BI Reporting	Automated Indexing
IoT Sensors	Change Data Capture	Data Warehouse Data	Lake	Query Engines	Data Science	
Web	Data Streaming	Lake	Data lakehouse	Predictive Analysis	Machine Learning	Automated Storage & Retrieval
Social Media			MDM/MRD			

Manage

- Technical metadata
- Business Glossary
- Data Lineage
- Search & Discovery
- Data Access & Retrieval

- Automated Metadata
- Curator Annotation
- Stewardship Workflow Automation
- External Database Connection

- Active (Runtime) Metadata
- Static (Design) Metadata
- Access (Usage) Metadata
- Quality Metadata

- Sensitivity Classification
- Record Ownership
- Retention Definition

Govern

Governance
- Data Curation
- Data Utility
- Data Privacy & Protection
- Data Value

Administration
- Data Security
- Severs & Networks
- Databases
- Tools & Technologies

Infrastructure
- Platform (On-Prem/Cloud)
- DBMS (SQL & NoSQL)
- Analytics Pipeline % Technologies
- Records Repository

Integrate

Ingest, transform and combine different data from different sources

Govern

Embedment of policy and standard control, incl. for data protection & privacy, security & quality

Visualize

Empower users everywhere to discover, access and connect to the data they need for insight, impact and integrity.

Value-driven Data Governance

Business Strategic Objectives	Strategic Alignment DG Objectives	Required Capabilities	Business Value-add
Leveraging data assets	Source integration: Internal & External	Data Integration & Interoperability	**Accelerating access to source data** • Improved access to trusted data. • Reeal time insights and reduction of key-man dependencies,by connecting through an approved data source.
Ease of access of data access	Data Cataloging	Business-friendly data catalog	**Data Democratization** • Federated data access, subscription-based model. • Enriched and comprehensive datasets. • End-to-end data lineage & audit trail tracking. • Stewardship collaboration engine.
Faster decision-making & improved insights	Business Process Automation	A stable DQ management environment.	**Improved data quality & integrity** • Automated DQ Management processes. • Direct external 3rd Party data (Enrichment) connection, through standardized ETL, cleansing and transformation. • Cost reduction due to automated controls.
Standardized data access control	Data security governance	Role-Based Access Control	**Secure data in transit & at rest** • Secure and governed data access. • End to end encrtyptions & Secure API's.
Central command & Control	Data Governance Framework	Data Governance Framework Embedment	**Control and Compliance** • Centarlised data governing forums. • Central discision making layer. • Formalised accountability Mechanism. • Data Risk Management

Measurable DG Benefits

- Accelerate the delivery of the data platform to **enable the democratization of data assets** for use across within Group Compliance, starting with prioritized business use cases.

- Comprehensive (*360-degree view*) customer data to help business in;

  - Optimizing marketing campaigns

  - Increasing sales

  - Upgrading customer service offerings

- Rationalize the management and provisioning of data to allow for **accelerated access to high quality data** for in-process use, analytics and change delivery.

- Faster decision-making and improved forecasting due to trusted data and automated processes.

- Optimizing product pricing and increasing operational efficiency.

- Automated, business-driven DQ management processes.

- Streamline delivery and recovery model towards using a **service-based approach** for clear business engagement. Improving customer acquisition and retention, enabled by better understanding of customer needs and preferences.

- Developing personalization and recommendation systems for;

  - Corporate websites

  - Streaming services

  - Online advertising

Measurable DG Benefits

- Tapping into **unstructured and semi-structured data** which has emerged as the biggest dataset in recent years.
- Analyzing data from the following sources to help understand customer sentiment, spot patterns and match content to advertising
  - Text
  - Images
  - Videos and Audio
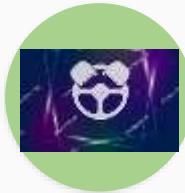
- Strengthening **fraud prevention and cybersecurity protections** by better identifying suspicious transactions and security threats.
- Identifying and mitigating potential risks in key business processes.

- Automated Disaster Recovery **(DR)** and Business Continuity Management **(BCM)** controls.
- Enabling **preventive maintenance** to minimize system failures and downtime in customer facing and support functions deemed business critical.

Data Governance Success Stories

Starbuck Data Success Story:

- Uses big data and AI to help direct marketing, sales and business decisions from 90 million transactions a week in 25,000 stores worldwide.

Starbucks Rewards and Mobile App

- The mobile app has more than 17 million and the reward program has 13 million active users.
- These users create data about what, where and when they buy coffee and complementary products that can be overlaid on other data including weather, holidays and special promotions.
- Starbuck in tun uses the data to offer the following customized services;
    1. **Personalizing the Starbucks experience:** Based on the customers ordering patterns, the app suggests new products (and treats) customers might be interested in trying.
    2. **Targeted and personalized marketing**: Gathered data helps Starbucks suggest new products for to try also helps the company send personalized offers and discounts that go far beyond a special birthday discount.
    3. **Virtual barista:** My Starbucks Barista through the Starbucks mobile app, allows you to place an order through voice command or messaging to a virtual barista using artificial intelligence algorithms behind the scenes.
    4. Etc..

Self introspection

- What purpose does data serve in your business?

- Which data elements are considered "Essential" within your business & why?

- Where is data governance most important? What domains and parts thereof does the business most need right now?

- What data governance archetype best fits the organization?
- Are current efforts aligned to that level of need?

What is the opportunity cost of not getting data governance right in terms of;

- Missed business opportunities?
- Extensive time lost in manually cleaning data, or incorrect and suboptimal business decisions?

Thank You

DataCon Africa 2022

Dr. Sizwe Gwala
Head: Data Governance
Absa Group Compliance