# DATA ON DEMAND AT THE BANK OF CANADA

BANK OF CANADA
BANQUE DU CANADA

## Streamlining Infrastructure for Continuous Improvement

CDAO Public Sector | June 19th, 2024 | Ottawa, ON

**Erik Balodis**
Director – Analytic Solutions and Data Science
Policy and Research Division
Bank of Canada


# A little bit about me

# Speaker's Disclaimer

*The views and opinions I am about to express in this presentation are solely my own and do not necessarily reflect the official policy or position of the Bank of Canada.*

*Any content provided is for informational purposes only and is not intended to represent the Bank of Canada's stance or strategy.*


Setting our data context…

# A little bit about the Bank

- ~2000 employees
- Established via the Bank of Canada Act to *"promote the economic and financial welfare of Canada"*
  - develop and implement monetary policy;
  - oversee the financial system;
  - develop and issue currency;
  - Act as the fiscal agent for the GoC;
  - *new* provide supervision over payment service providers

# How we typically think of data support across the organization…

Economic Policy and Research

- Assessing the Canadian economy and financial system to inform policy decisions

Operations

- Conducting the Bank's operational mandates for Currency, Banking, etc.

Corporate Administration

- Administering the Bank of Canada's business (e.g. HR, Finance, Legal, etc.)

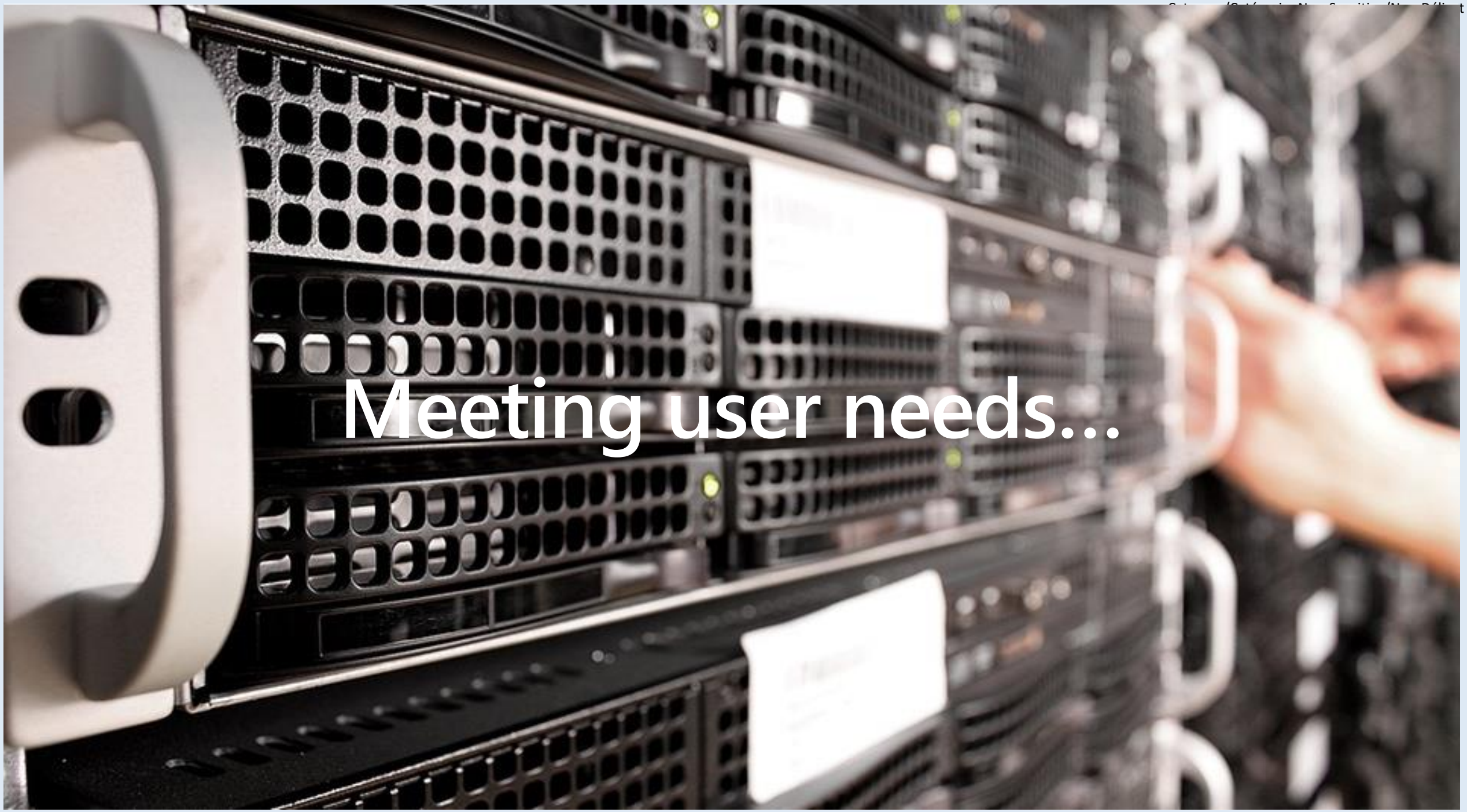
**…Why is this important… ?**

# Understanding the Bank of Canada's Use-Cases

| | Economic and Financial System Policy and Research | Operations | Corporate Administration |
|---|---|---|---|
| Use-case complexity | High | Mixed | Mixed |
| Population of data users | High | Medium | Growing |
| Storage and compute needs | High | High | Standard |
| Tooling needs | Varied | Varied | Simpler |
| Data governance | Centralized | Central/Federated | Federated |
| Data source | Mostly 3rd party | Mixed | Internal |

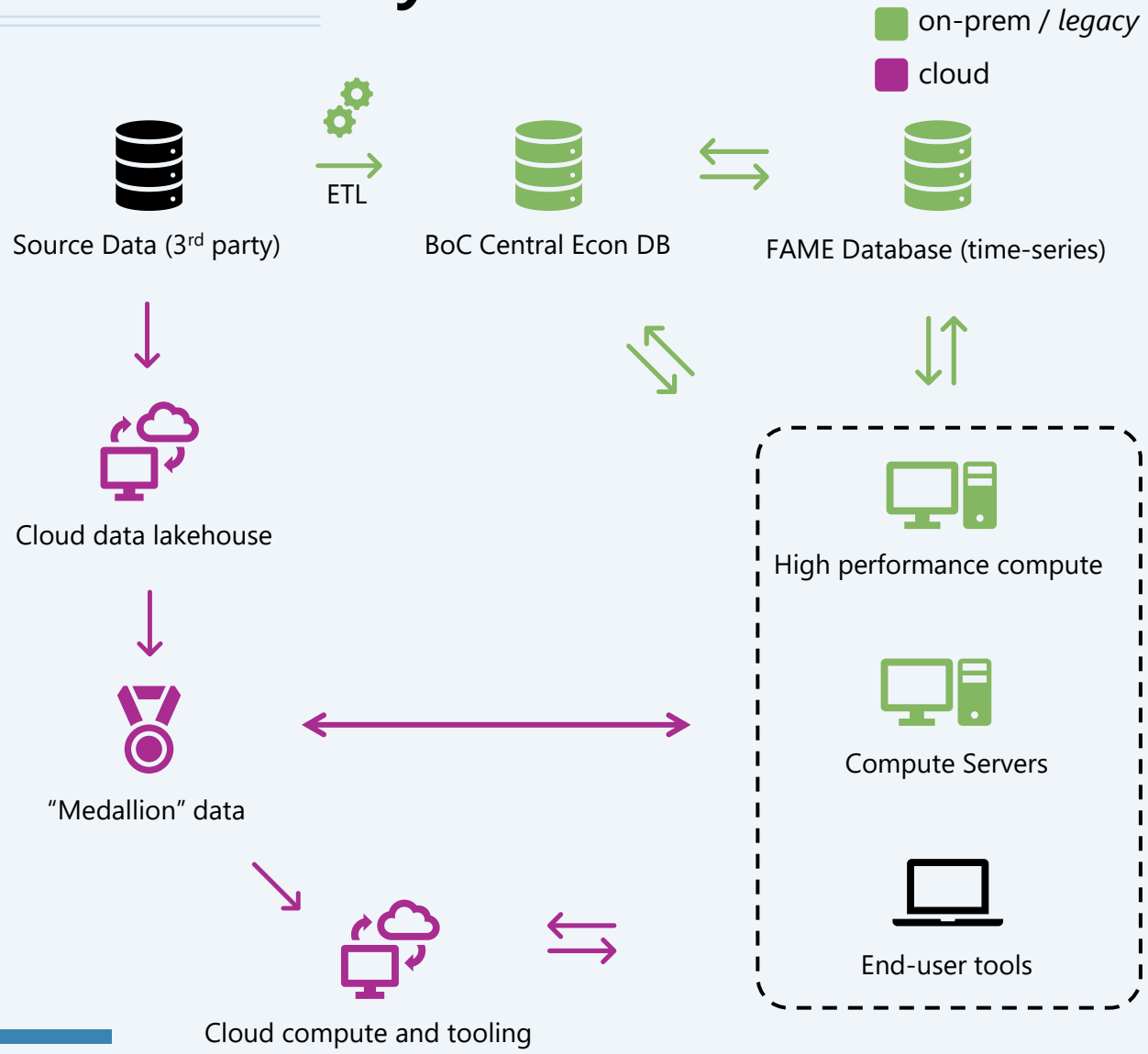
# How our data strategy shapes our decisions

- The Bank of Canada's data strategy focuses on **enablement** and **empowerment** of end-users
  - Implementing capabilities for self-service
  - Enabling and expanding access to data
  - Building knowledge and communities of practice
  - Supporting skills development across the Bank
- Our business needs have influenced our approach to building a data infrastructure and support service
- As needs continue to evolve, so too does the nature of our data support
- **Finally, we are also responsive to broader trends in data technology (e.g. cloud)**


Meeting user needs…

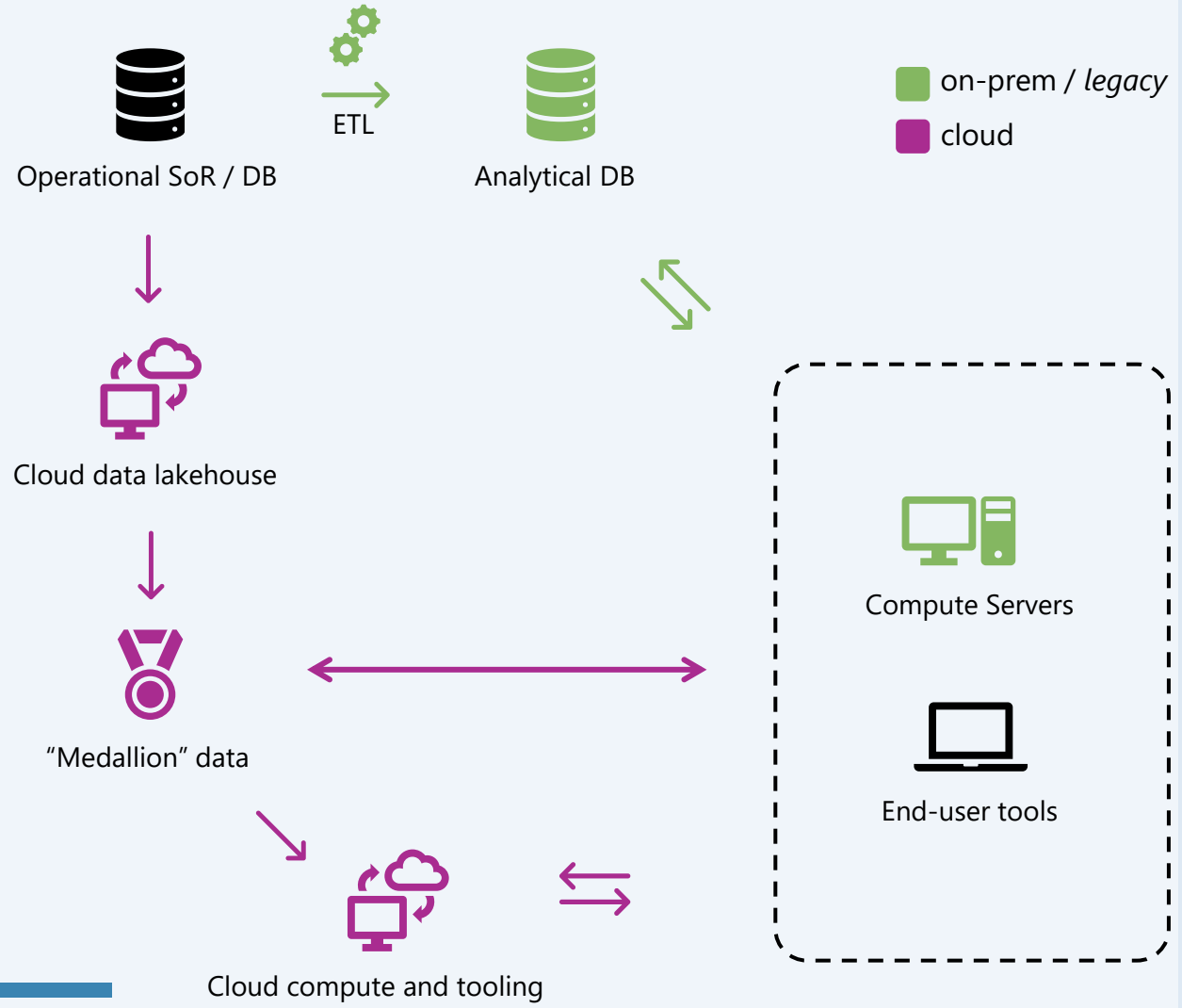
# Economics, Research and Financial System Work

- Primary objectives (use-case driven):
  - Ingest 3rd party data onto platform
  - Clean and standardize data
  - Make data available to economists in a business-ready format
  - Develop and support data-reliant processes and use-cases:
    - Time-sensitive data
    - "Big data"
    - Data science
    - Visualizations and dashboards
    - Data dissemination



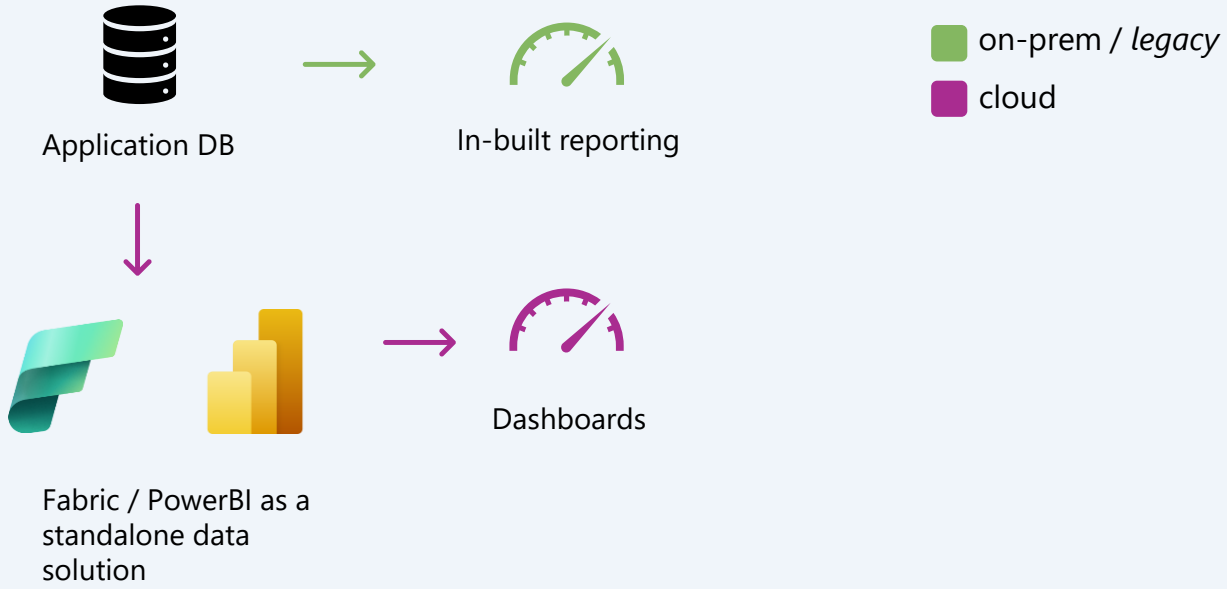
# Operations

- Primary objectives (use-case driven):
  - Platform operational data for analysis
  - Clean and standardize data
  - Make data available to operational analysts in a business-ready format
  - Develop and support data-reliant processes and use-cases:
    - "Big data"
    - Data science
    - Visualizations and dashboards



# Corporate Administration

- Primary objectives (use-case driven):
  - Unlock application data
  - Make it useable by a wide range of users
  - Beachhead some key reports/dashboards
  - Develop and support data-reliant processes and use-cases:
    - Visualizations and dashboards
    - Cross-system integrations
- Why does this look so different?
  - No centralized data repository
  - Needs were historically met by in-built reporting
  - Federated governance




Case study: cloud migration for policy and research work

# Cloud Migration Objectives

1. **Enhance** our ability to onboard and manage a wide range of data on a consolidated platform
   - Structured, unstructured, semi-structured
   - Relational, non-relational
2. **Modernize** our ability to manipulate and analyze data
   - Best-in-class ETL (Databricks, Azure Data Factory) and user-facing tooling
   - Cloud-based data science solutions (e.g. MLFlow)
3. **Centralize** our approach to using data into fewer, more centrally managed “patterns”
   - Reduce our reliance on local data solutions (including spreadsheets)
   - Strengthen central data management, cataloguing, and governance
4. **Future-proof** by reducing dependency on legacy data processes and technology
   - Move towards open-source languages (e.g. Python instead of FAME 4GL)
   - Reduce legacy footprint and reliance on high-cost legacy data stack components

# Workstreams

## Platform and Architecture

- Develop a flexible data lake platform based on open-source languages that meets business needs
- Select, test, and implement an appropriate architecture (storage format and governance)
- Develop user-facing functionality for accessing and working with data on the lake

## Data Migration

- Identify and migrate priority data feeds for key economic data domains:
  - Macroeconomic time-series analysis and survey data
  - Financial Markets
  - Regulatory (Financial Institutions)
  - Research and *ad hoc* data
- Establish robust and repeatable patterns for ingesting data

## Use-case Migration

- Identify and evaluate end-user processes and business requirements
- Develop and integrate approaches based on new technology platform
- Establish a robust operational support model that aligns with business process criticality

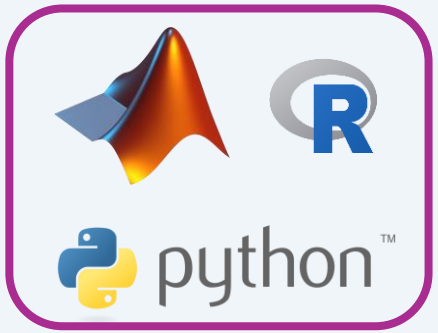
# Bank of Canada Data Lake Architecture

## Foundation and Infrastructure

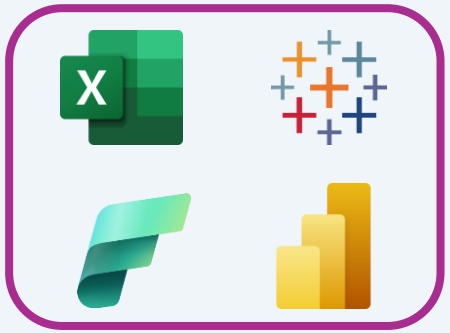
- Azure data lake environment for storage and orchestration (ADLS, Azure Data Factory)
- Parquet/Delta Lake as storage format (open-source)
- Databricks for notebooks, compute, and governance (Unity Catalog – presents as SQLDB to end-users)

## User-facing Access and Tooling

Code-based access




Low-code / No-code



# Lessons Learned

- **There are a lot of "Unknown Unknowns"**
  - Consistent w/any change in technology, unforeseen challenges arise
- **Tech evolves faster than your migration**
  - You need to anchor on an architecture and start building – we iterated 2-3 times and had a lot of refactoring work to do. We're holding onto our current architecture as our "target state"
- **Change management**
  - We're just starting up the slope of change management, and anticipate that there's going to be a lot of work to do – breaking and reforming processes, access patterns, re-skilling staff, etc. etc.
- **Pivoting from "innovation" project to operational project is hard**
  - May require different skillsets, teams, frameworks, etc.
  - Managing the capacity through the transition and down-scaling of project => operations


*all images are licensed for use royalty-free and are used with permission from Pixabay