

Using Snowflake to Unify Data, People and Technology

Dan Christian

Sr. Director, Data Warehouse Development

About Penguin Random House LLC

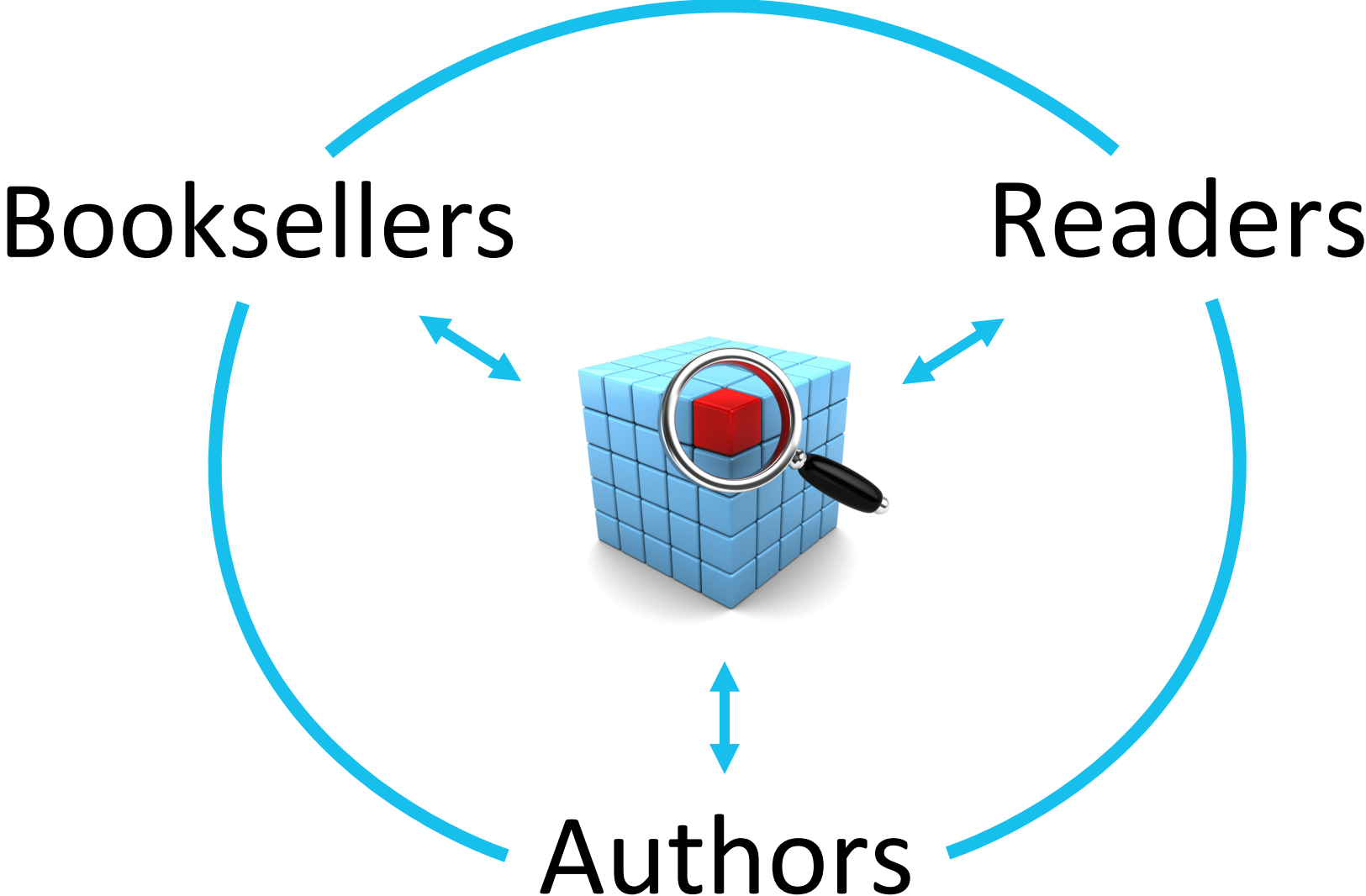
- World's largest trade book publisher
- Owned by Bertelsmann and Pearson (75%/25% equity stakes, respectively)
- Multi-local global company, with entities in UK, Australia, Spain and others
- 250 editorially independent publishing imprints
- Over 15,000 new titles published annually
- 461 titles on the NY Times bestseller list in 2017



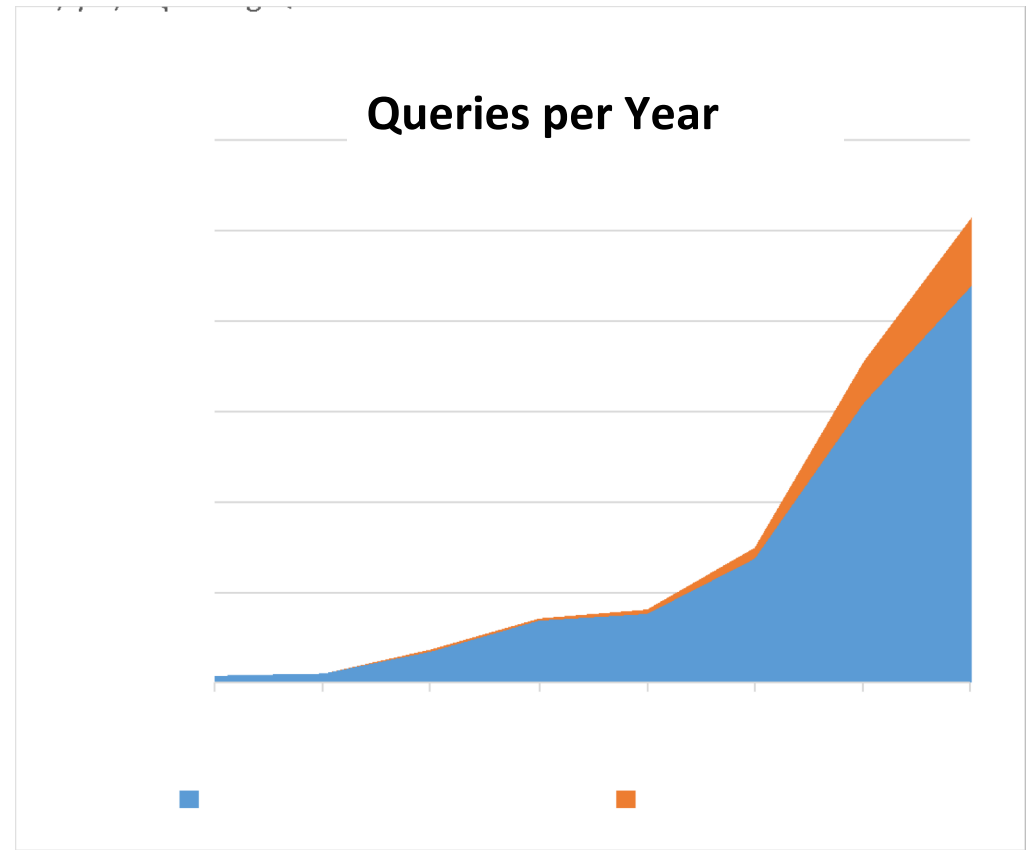
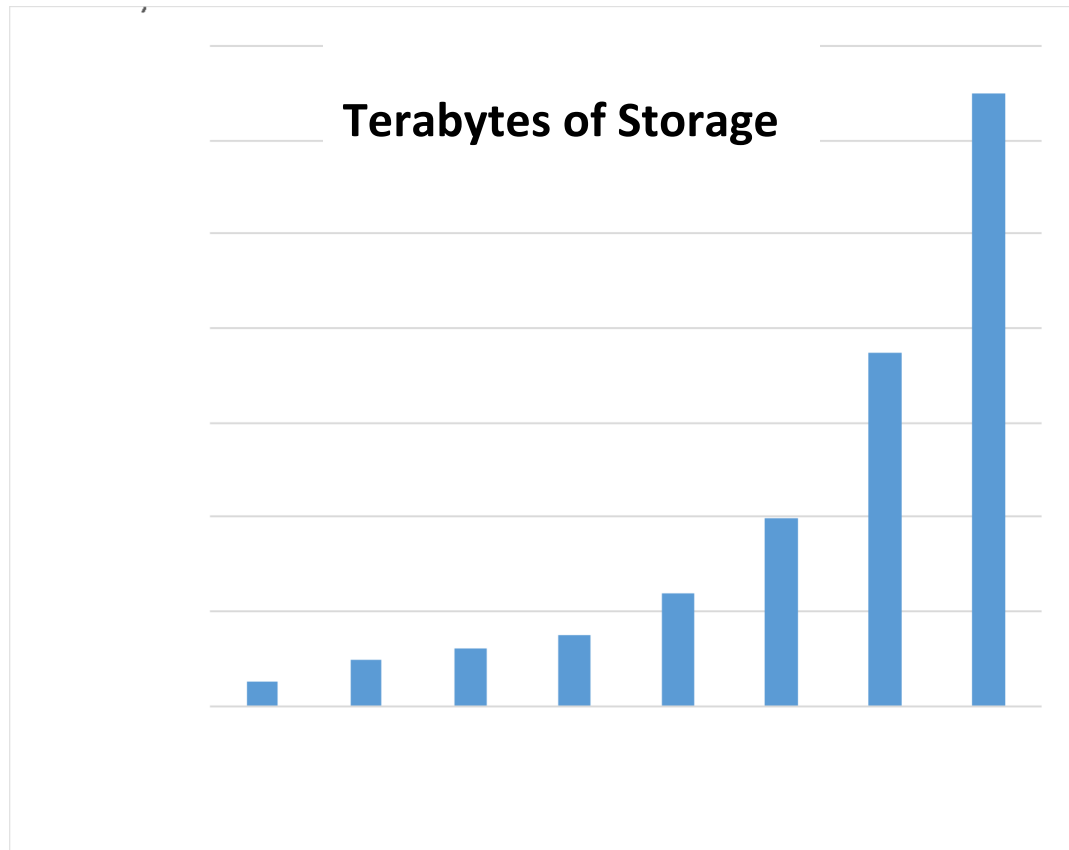
Business Challenges in Publishing Today

- A world where people love reading books, but where helping readers to discover new books becomes increasingly difficult.
- The need to react quickly when we spot trends and opportunities through social channels.
- Extending our reach and our brand globally on behalf of our authors.
- Supporting our authors and helping them to connect with readers in innovative ways.

Data Drives Our Success



Challenge: Rapid, Persistent Growth in Scale



Our Journey to Snowflake Started With a Drip



Four Needs that Drove our POC Process

- Speed
- Data Unity
- User Sophistication
- Globalization

The 'Tortoise and Hare' Problem: Can Both Win?



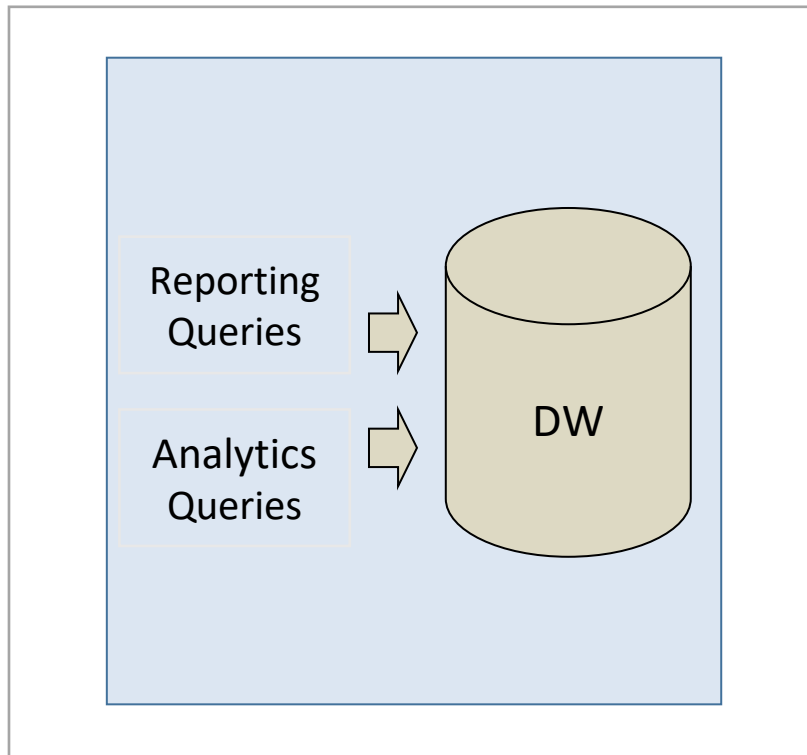
- Deep-dive analytics questions
- Huge amounts of data
- Queries are ad-hoc and unpredictable
- User can afford to wait for the query (well, a little while!)



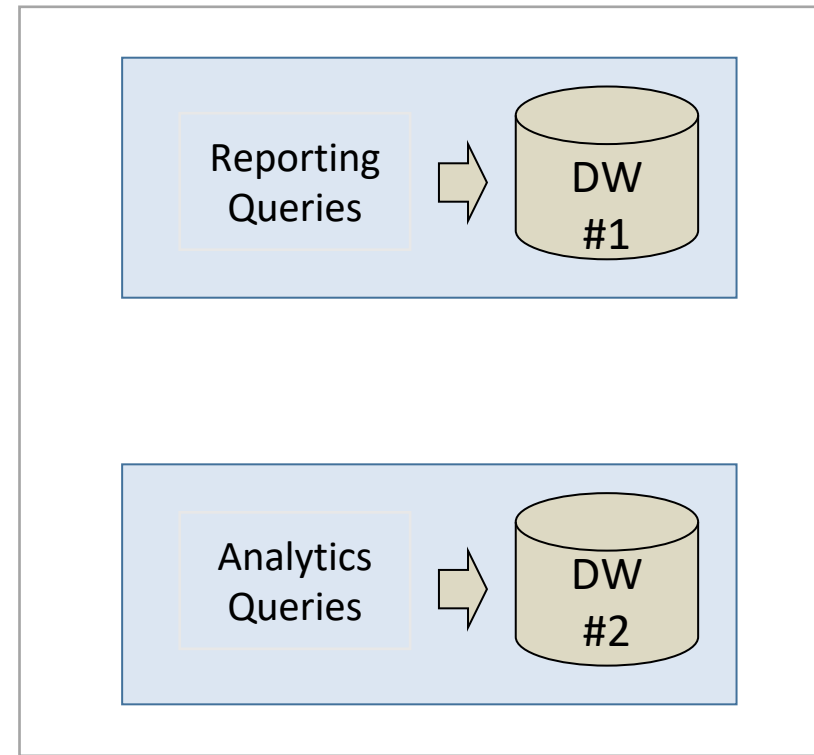
- Common questions for company-wide reporting
- Small to intermediate amounts of data
- Queries are predictable
- User wants an answer fast (clicking and drilling at the speed of thought)

Two Solutions We Resisted

(1) Buy a Much Larger System?



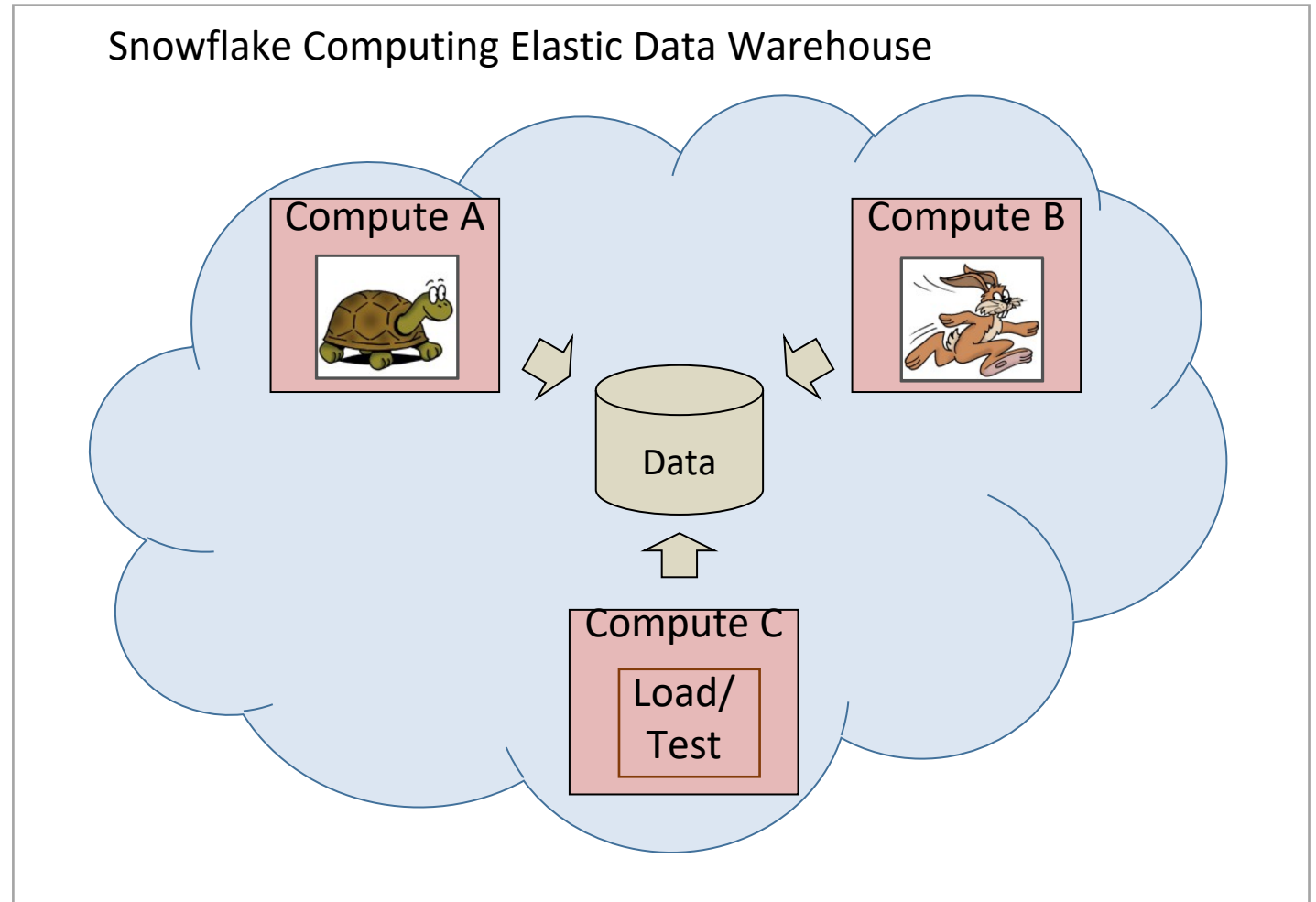
(2) Split the Data Warehouse in Two?



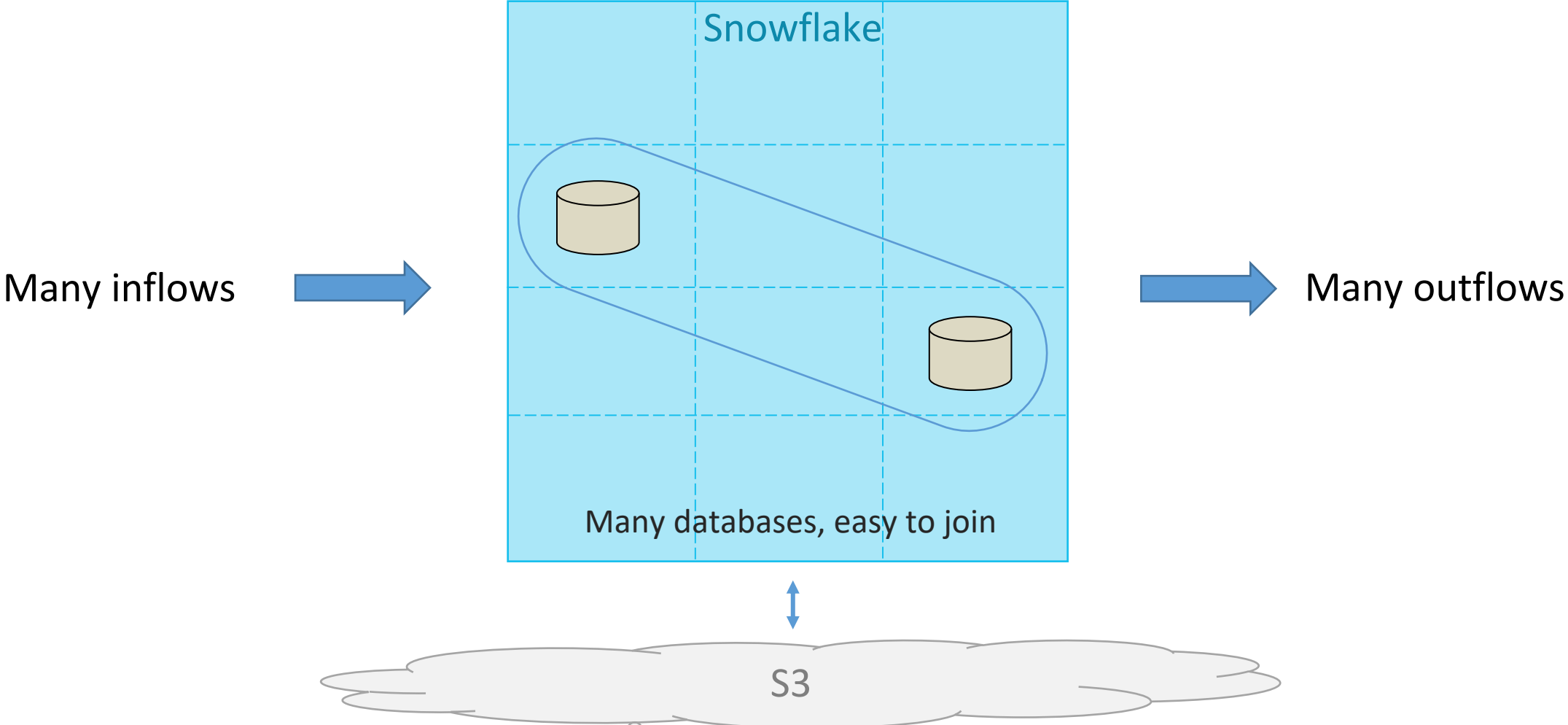
Speed and Scaling Problem – Solved

Snowflake's Big Idea:

- Separation of Compute and Storage
- Multiple Virtual Warehouses, spun up as needed
- All VW's point to the same data (one version of truth)
- All components are elastic



Data Unity is Easy on Snowflake



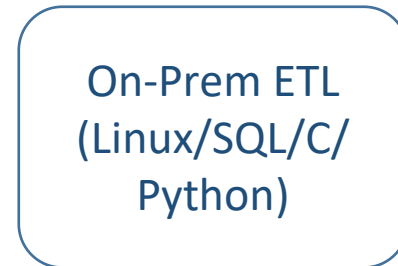
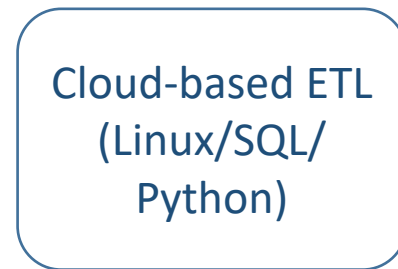
User Sophistication – Bring it On!

Many of our users are now using both BI and lower-level tools:

- Python
- R
- Data Science packages
- SQL queries via Snowflake UI
- Snowflake command line (SnowSQL)
- Business Objects, Tableau and Microsoft Power BI

Analytics Architecture

Data Sources



Data Consumers



Globalization of Data at Penguin Random House

Globalization Needs:

- Easy access for our multi-local global companies
- Easy access for our authors worldwide
- 24 x 7 uptime
- Seamless information, data integrity



Penguin
Random House
Canada



Penguin
Random House
UK



Penguin
Random House
Australia

Penguin
Random
House

Penguin
Random House
Grupo Editorial



PENGUIN
RANDOM HOUSE
INDIA PVT. LTD.



Penguin
Random House
PENGUIN BOOKS CHINA

Seven Things I Like About Working With Snowflake

- Strong, sensible product vision
- Huge queries run fast
- No indexes, no distribution clauses
- No more red zones (workload contention)
- No more daily downtime for load cutovers
- Hands-off software updates
- Great customer service