

# Using Snowflake to Unify Data, People and Technology

Avinash Deshpande, Chief Software Architect  
Logitech

**WE CONNECT YOU TO THE  
DIGITAL EXPERIENCES YOU LOVE**



**WE DESIGN PRODUCTS THAT BRING PEOPLE TOGETHER**





Creativity and  
Productivity



Music



Gaming



Video  
Collaboration



Home



### **The Big Picture**

Confidently make — and defend — critical decisions using the results of your data analysis.



### **Skills, Tools, & Strategies**

Use Excel, SQL, and Tableau to spot trends and drive business decisions with real-world data.



### **Analytics in Practice**

Share your insights and tell compelling stories using data visualizations and dashboards.

# DATA DRIVEN ENTERPRISE

---

# Monetize The Data In Your Business

Democratizing data

Enhanced cognitive solutions

The power of hybrid-data sources

Get the best out of unstructured data



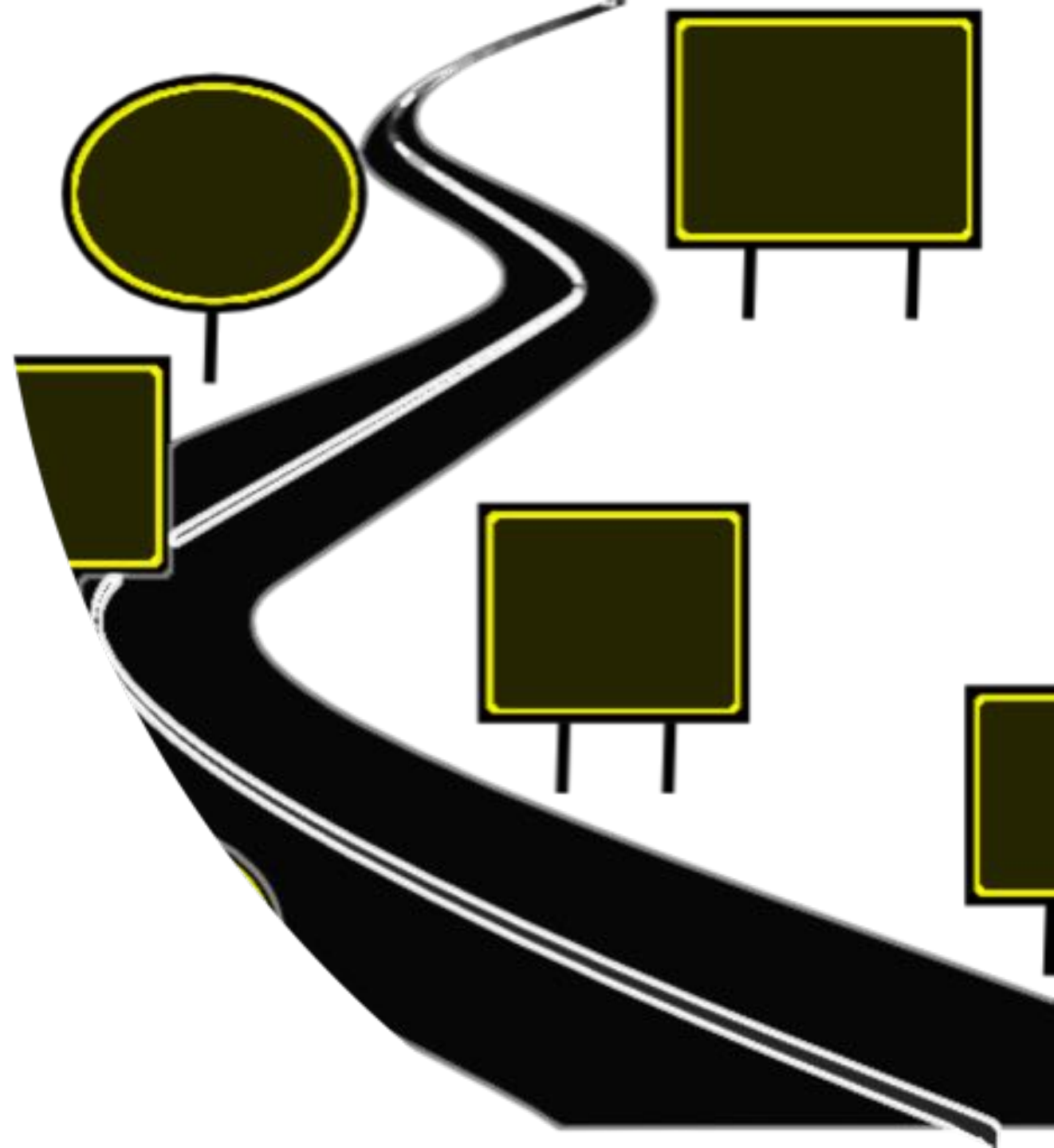


**ANALYTICS AT SCALE**  
**SUPPORTING OUR GROWING BUSINESS**

Cloud empowers IT organizations to redefine the way data services are produced and delivered

---

Scalable	Elastic Infrastructure simple, secure, robust, and scalable
Efficient	<b>Pay as use</b>
Reliable	<b>Managed services</b>
Govern	<b>Transparency</b> on usage patterns <b>Breadth of services</b>



# NEED FOR DATA VIRTUALIZATION

Abstract access to disparate data sources

A single semantic repository

Optimized data availability in real-time to consumers

Centralized, governed and secured data layer



WAIT BUT WHY

## REAL-TIME ON DEMAND delivery to your PHONE and DESKTOP and DASHBOARD

- Executive Summaries
- Customer by Product
- Product by Customer
- Demand / Supply updates
- Market Analytics / Market Share
- Marketing Reports
- Competitive Analysis
- Sentiment
- ...

*NLP is a scalable self-service environment, meaning we can open it to business users (self-service) and allow them to improve and drive business impact and adoption. It is language agnostic, meaning we can publish reports in the language they are written.*

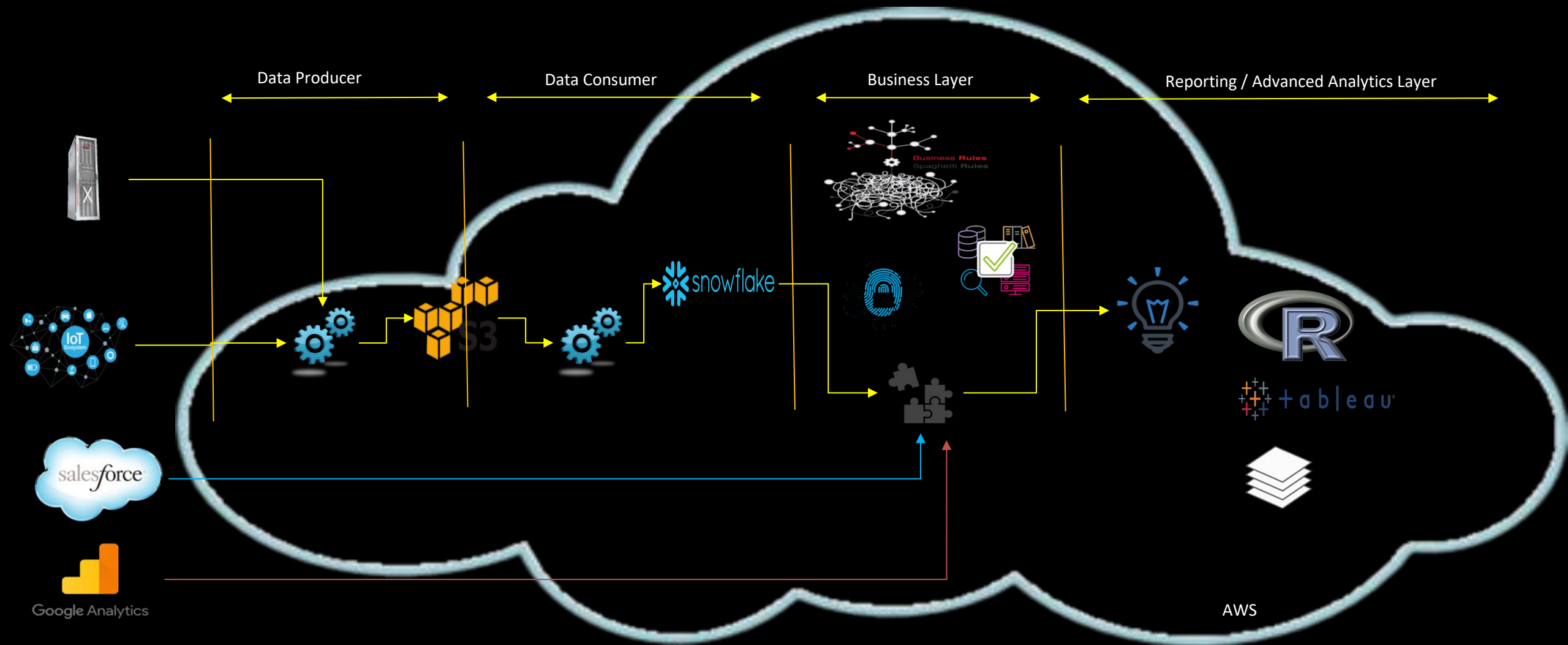


# Improve the User Experience



- ✓ **User Pain:** “Tableau is always slower when I want to use it (peak business hours).”
  - Snowflake is able to flex-up compute power in seconds
  - Business users can have their own isolated instance of right sized compute so that performance is always consistent for the work *they do*, and not impacted by what others are doing.
- ✓ **User Pain:** “I want access to more historical data than I have today.”
  - Snowflake’s low cost, fast, infinitely scalable storage layer removes the limitations of adding and keeping more historical data than typical data warehouse solutions allow.
- ✓ **User Pain:** “Commonly used reports always seem to be slow.”
  - Snowflake has the unique ability to globally cache commonly used queries that are sent via Tableau. This means that commonly used workbooks are almost always cached and end users experience extremely fast performance regardless of how many people are running the same workbook.
- ✓ **User Pain:** “I want to explore non-traditional data sets that aren’t currently available.”
  - Unlike other traditional DW solutions, Snowflake treats non-traditional data types like JSON/AVRO/XML as first class citizens (direct SQL access and fast performance). This allows the data to be immediately available without complex ETL.
- ✓ **User Pain:** “I’m tired of waiting for new data to be loaded into the system.”
  - Snowflake’s unique architecture allows customers to implement new data ingestion processes like 24/7 loading. This lets end users “see” their data in near real-time vs the traditional nightly batch. Use Tableau Live Connection rather than Extract.

# Solution Architecture



## Other Solutions

- **Architecture/Storage:** traditional shared nothing architecture. Data lives on EC2 nodes, requiring costly 24/7 uptime, even when not in use.
- **Data Types:** Requires use of additional tools (Hadoop, Mongo, etc.) to ingest and make semi-structured data available.
- **Scalability:** Extended process to resize compute resources to accommodate additional demand.
- **Concurrency:** Published limits of 50 concurrent users/queries, but generally slows down around 15.
- **Administration/Design:** Need to continually manage: vacuuming, distribution/sort keys, compression, metadata, indexing, backups, etc. Need to understand data model in advance.

## Snowflake

- **Architecture/Storage:** Multi-cluster shared data architecture. Data stored in S3, allowing multiple EC2 compute clusters to access simultaneously without contention.
- **Data Types:** Ability to ingest and query raw JSON, XML, Avro, Parquet without prior transformation.
- **Scalability:** Data not coupled to compute, allowing the ability to resize instantly and shut down when not in use.
- **Concurrency:** Ability to isolate users on separate compute resources to avoid contention. Auto-scale feature scales compute resources horizontally to support concurrent workloads.
- **Administration/Design:** ZERO; free up your DBA team for other tasks. Load data in real time without need for model.

Compare

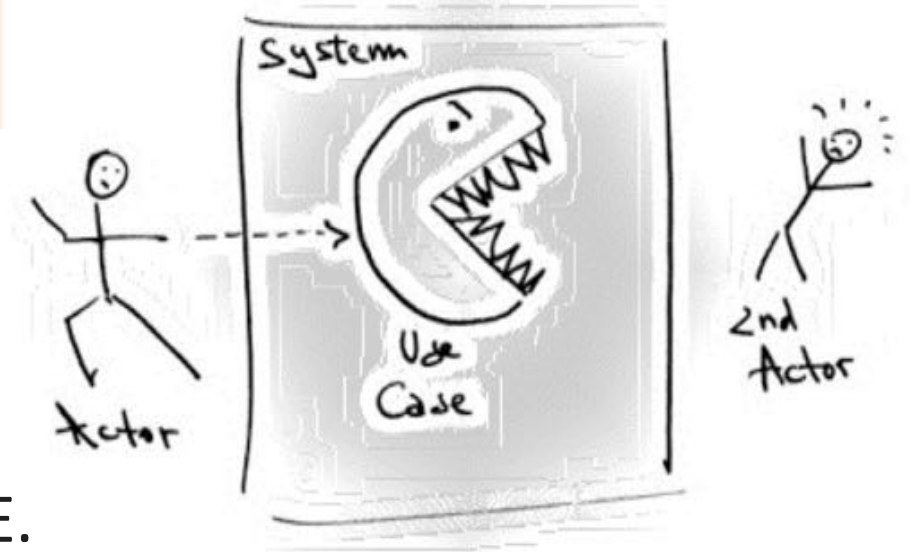
# Unique Snowflake features

- **JSON:** ingest raw JSON without transformation. Query JSON with SQL and correlate against relational data
- **Cloning:** instant dev/test environments or point in time snapshots.
- **Time Travel:** Query data as of any point in time within the past 90 days
- **Query Caching:** instant results for Executive dashboards and commonly run reports.
- **Backups:** automatic cross data center replication
- **Data Sharing:** publish or consume data sets to or from external clients without direct system access
- **Auto-Scaling:** dynamic horizontal scaling for concurrency to deliver consistent SLAs
- **Central Data Store:** Get everyone under one platform
- **Upgrades:** weekly system updates with zero downtime
- **Security:** encryption by default
- **Charge Back:** monitor business usage to understand how much each user costs you

# IOT

## Why Big Data?

DEVICES PUBLISH BILLIONS OF DATA RECORDS ...  
WITHOUT IT, YOU'RE NOT SEEING A FULL PICTURE.



We aggregate data about sessions and individual behaviors from across devices & build machine learning models to anticipate key issues from data such as:

INSPECTION  
RESULTS

CRASH  
REPORTS

IOT  
SENSORS

CONNECTED  
DEVICES

PROACTIVE  
MONITORING



# DATA SCIENCE

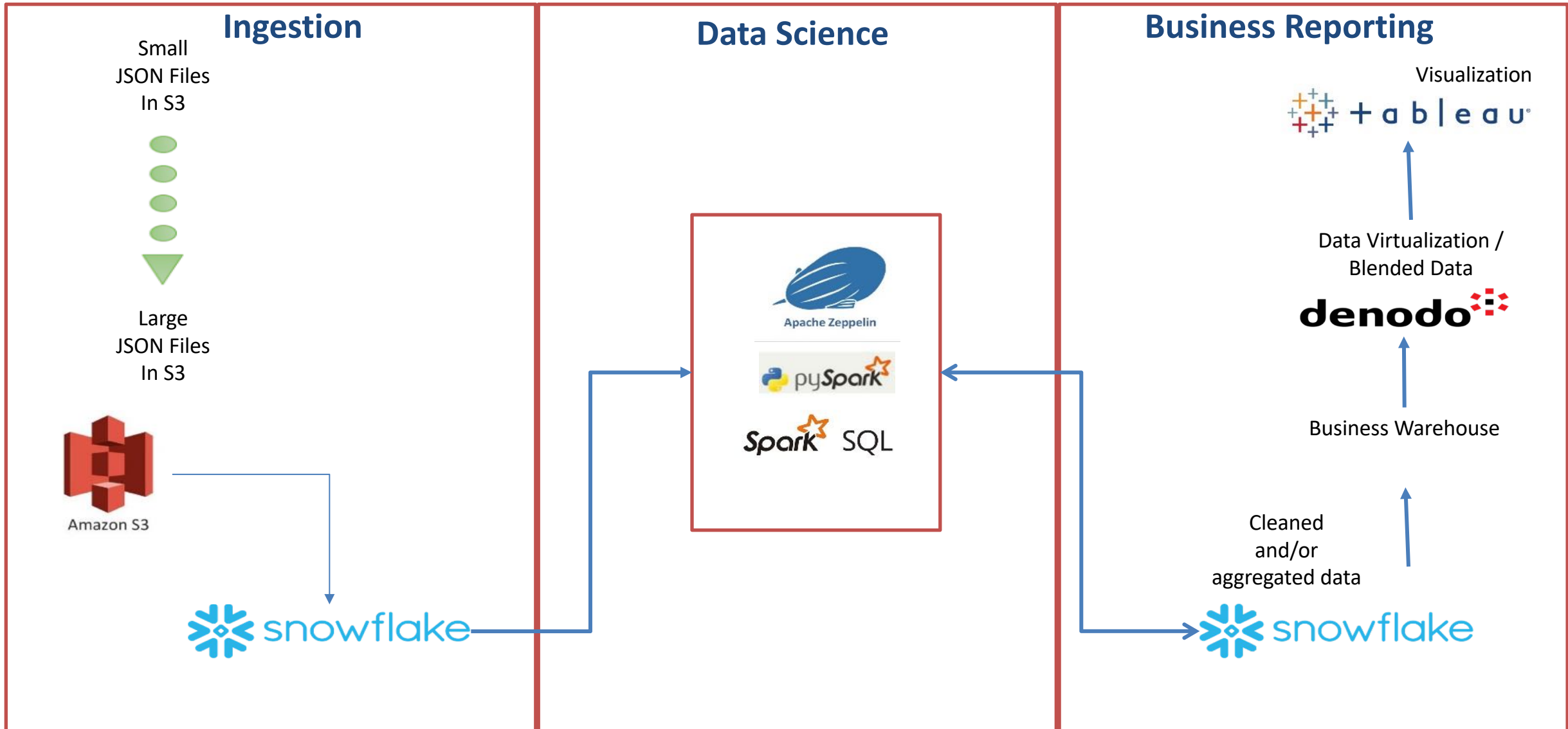
## **DON'T GET BURIED IN DATA.**

- Get purpose-driven data feeds and modular machine learning tools that deliver insights about where key issues and risks will emerge in real-time.

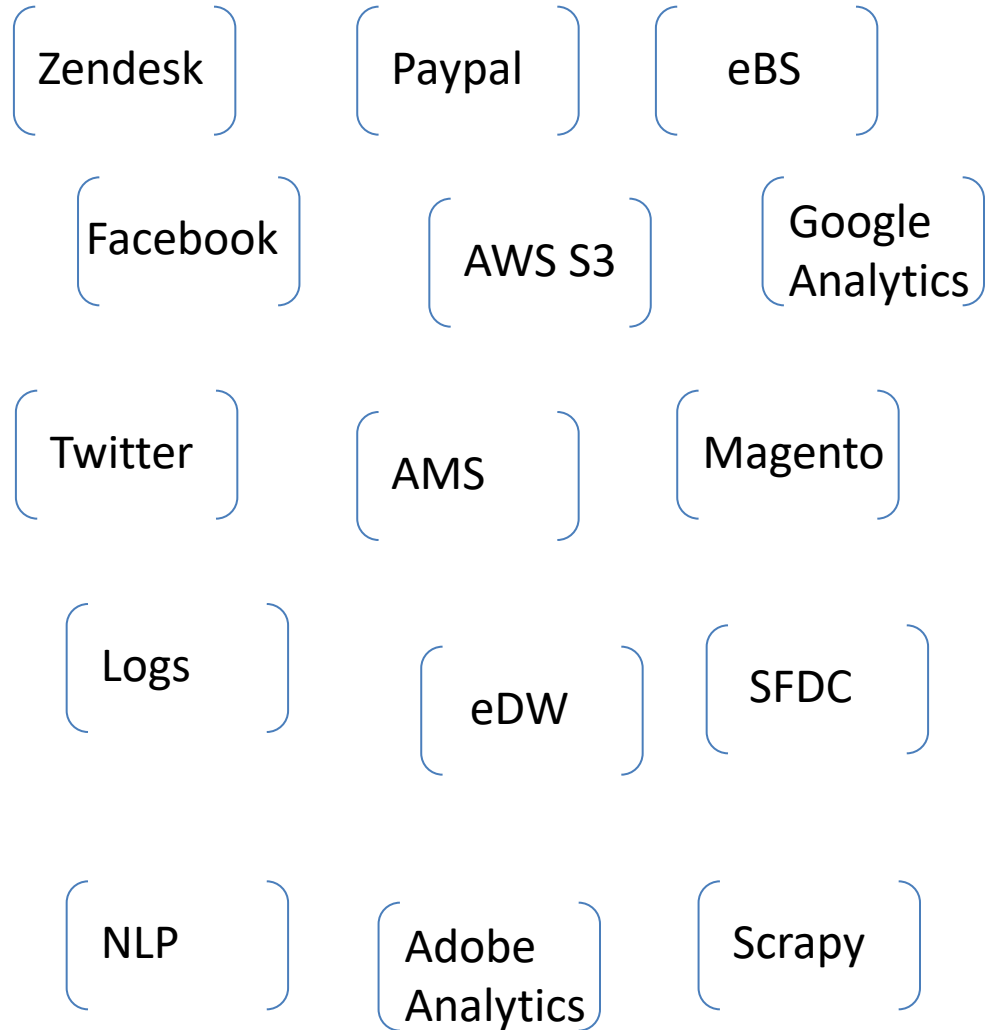
## **DATA CURATION**

- Make data more available and useful for those engaged in data discovery and analysis.
- Data socialization helps augment the power of human-powered data curation by fostering a culture of data-driven decision-making.

# IOT ARCHITECTURE



# BIG DATA FABRIC



Data Virtualization



**logitech<sup>®</sup>**