Many Faces of Real-time Analytics

Dunith Danushka

Redpanda

About the presenter

Dunith Dhanushka

Senior Developer Advocate, Redpanda Data

- Event streaming, real-time analytics, and stream processing enthusiast

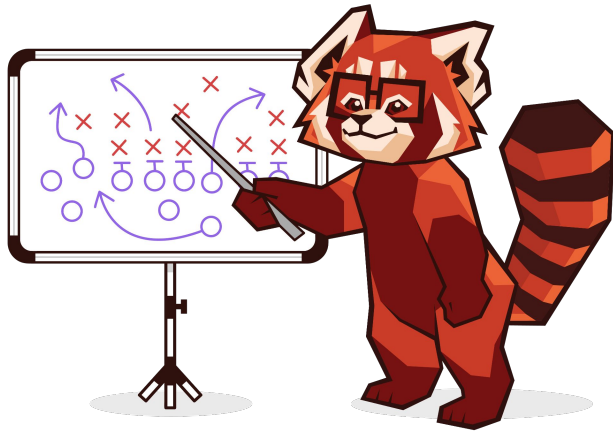- Frequent blogger, speaker, and an educator

twitter.com/dunithd

Medium
medium.com/event-driven-utopia

linkedin.com/in/dunithd/

Redpanda University

Free, self-paced online learning
https://university.redpanda.com

- Learn the fundamentals of data streaming and Redpanda

- Install Redpanda and use the rpk CLI to configure it

- Create producers and consumers in Java, Python and NodeJS

- Sign up today for free!

Get a Free Copy!

Compliments of Redpanda

The State of Streaming Data

2023-24 — Independent analysis of the size, shape and scope of modern streaming data deployments.

Commissioned by
Redpanda

Real-time Analytics

Deriving meaningful and actionable insights from moving data.

Not all real-time analytics systems are made equal!

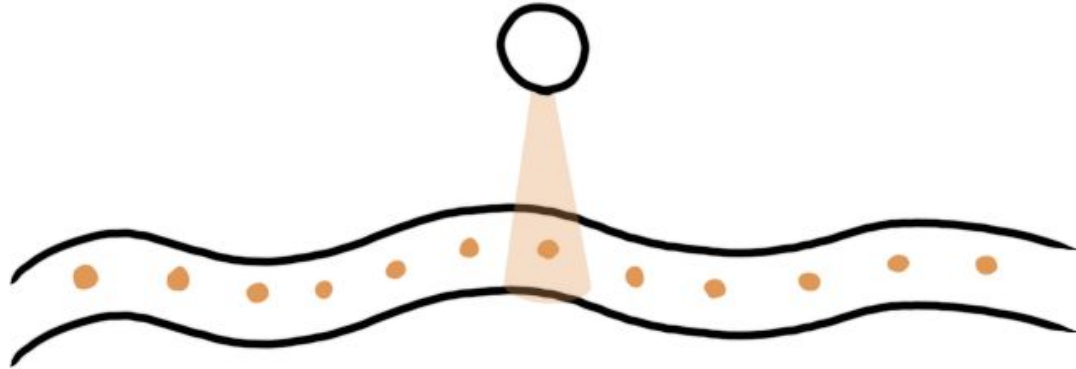# Five characteristics to classify them...

Query latency

How fast they can generate insights?

Data freshness

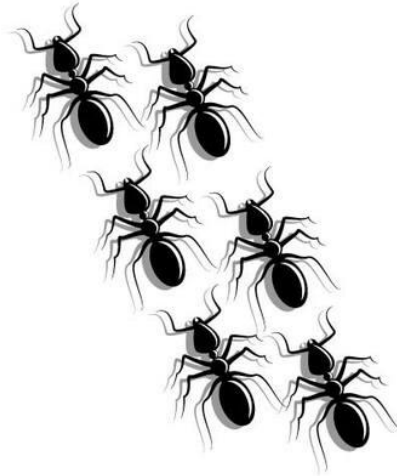How fresh are the insights they generated?

Query complexity

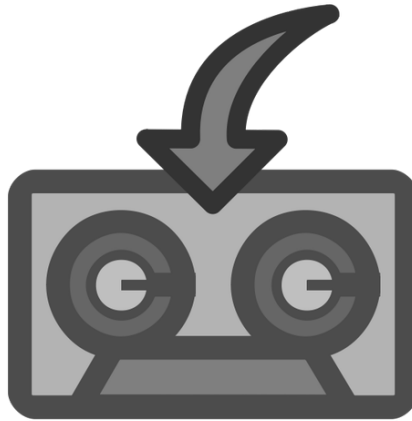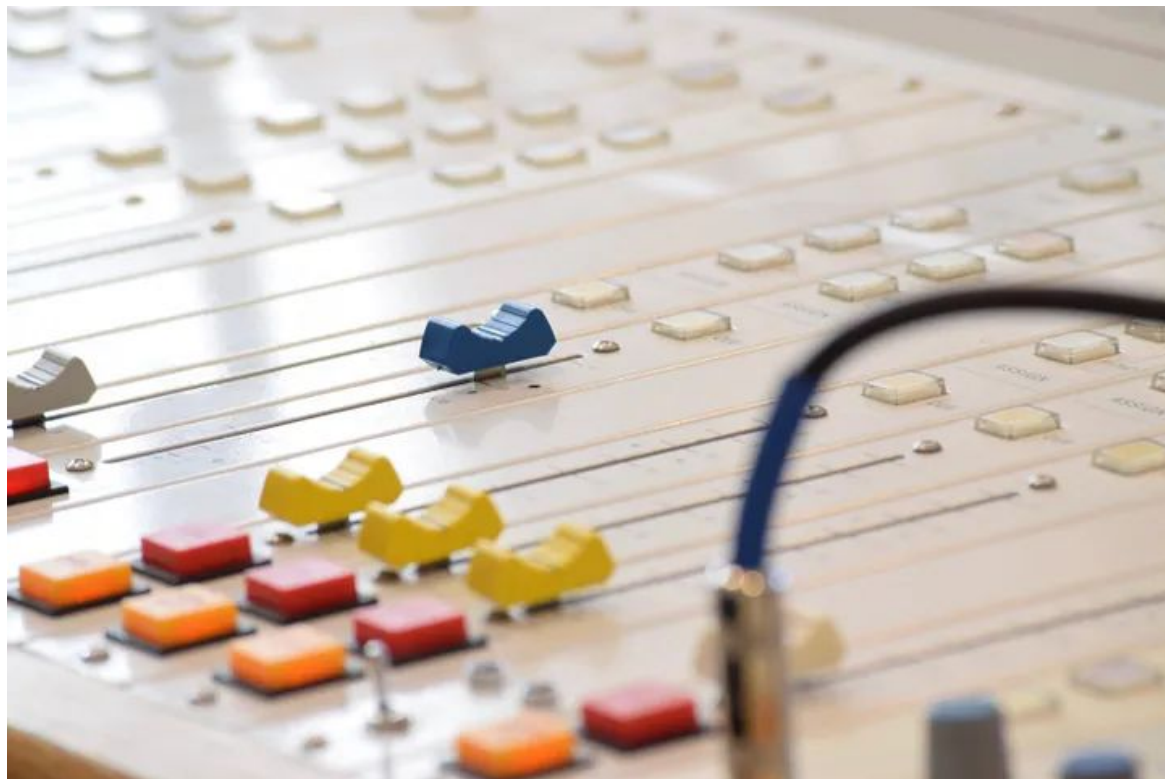How complicated would the queries be?

Concurrency

How many queries can be answered at a time?

Access to historical data

How much data do they need to generate insights?

Four Categories

1. Real-time automated decision-making
2. User-facing analytics
3. Operational monitoring
4. Ad-hoc querying of large data sets

Redpanda

Real-time automated decision-making

Redpanda

Machines making quick decisions

Fast decisions made without human intervention

- These systems spontaneously make decisions or predict outcomes after sifting through fresh data streams in real time.

- Insights are generated from constantly moving data with ultra-low latency, followed by quick reactions, organized as automated event-driven workflows.

- No human interactions since machines can make faster and more accurate decisions than humans.
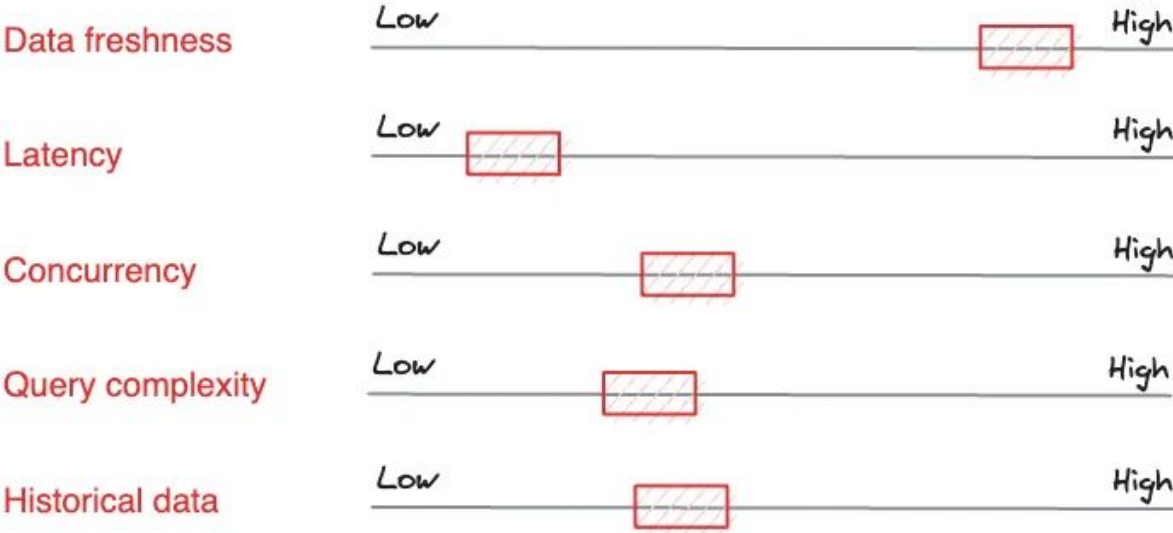
Examples

- Pattern matching
- Real-time inferencing of machine learning models.
  - Recommendations
  - Anomaly detection
  - Dynamic pricing

Must haves:

- ❏ Data freshness
- ❏ Low latency

Automated real-time decisioning

	Low		High
Data freshness	Low		High
Latency	Low		High
Concurrency	Low		High
Query complexity	Low		High
Historical data	Low		High

Technology choices

- Stateful stream processors. Eg. Apache Flink, Apache Beam, etc.
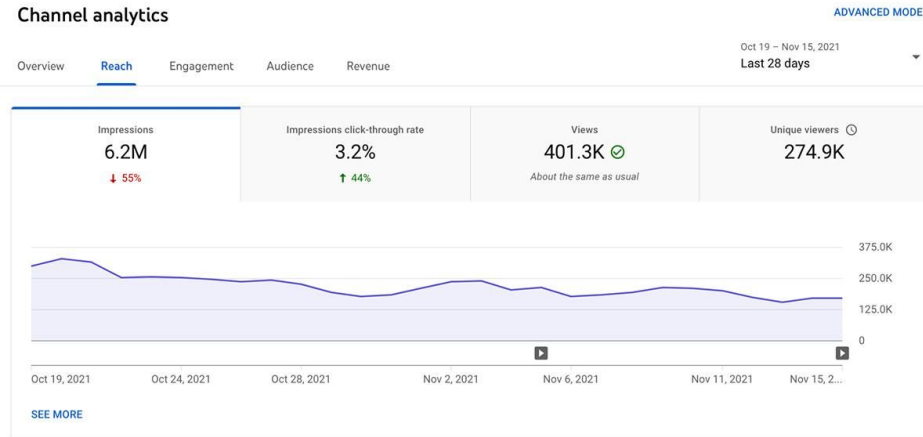
User-facing analytics

Analytics for end users

Let users slice and dice fresh data

- Directly involve end-users and customers in data exploration and decision-making.

- Users expect fresh analytics on dashboards with faster page loading times, demanding these systems to query fresh data and deliver results faster.
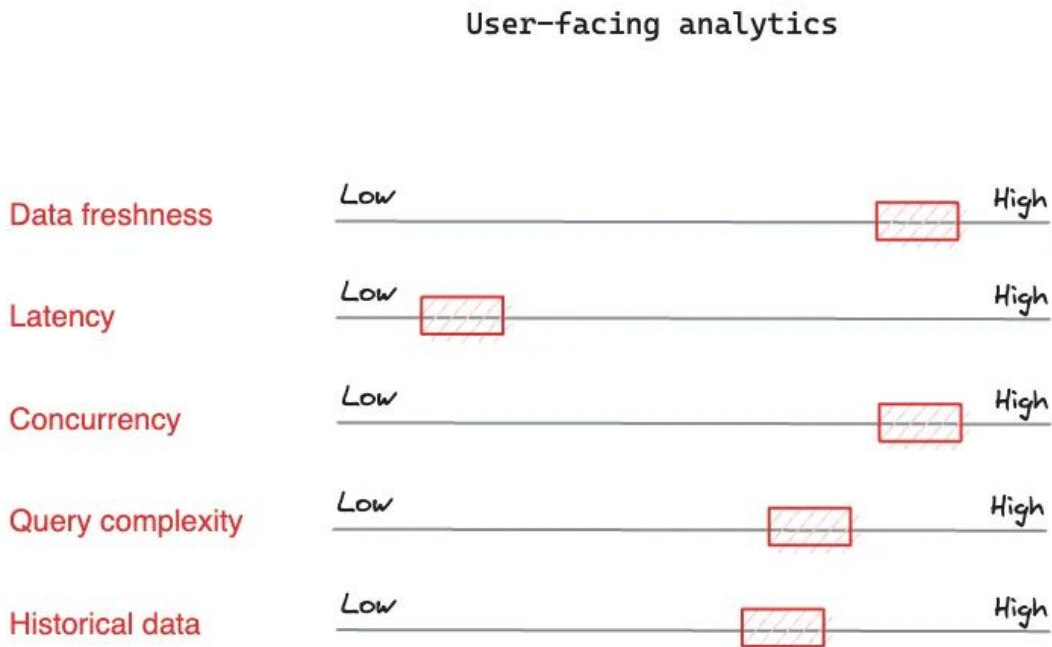
Examples

- User-facing analytics applications

Must haves:

- ❏ Data freshness
- ❏ Low latency
- ❏ Complex query support
- ❏ High query throughput

User-facing analytics

	Low		High
Data freshness			▦
Latency	▦		
Concurrency			▦
Query complexity		▦	
Historical data		▦	

Technology choices

Real-time OLAP databases
E.g. Apache Pinot, ClickHouse, Apache Druid, etc

Operational monitoring

Decision support for humans

Compute and visualize metrics

- Utilized by **internal** operational teams, such as SREs, network admins, etc.

- Assists human decision making by displaying real-time dashboards mounted on operations centers.

- Not only operational metrics, but also business KPIs can be tracked in real time.
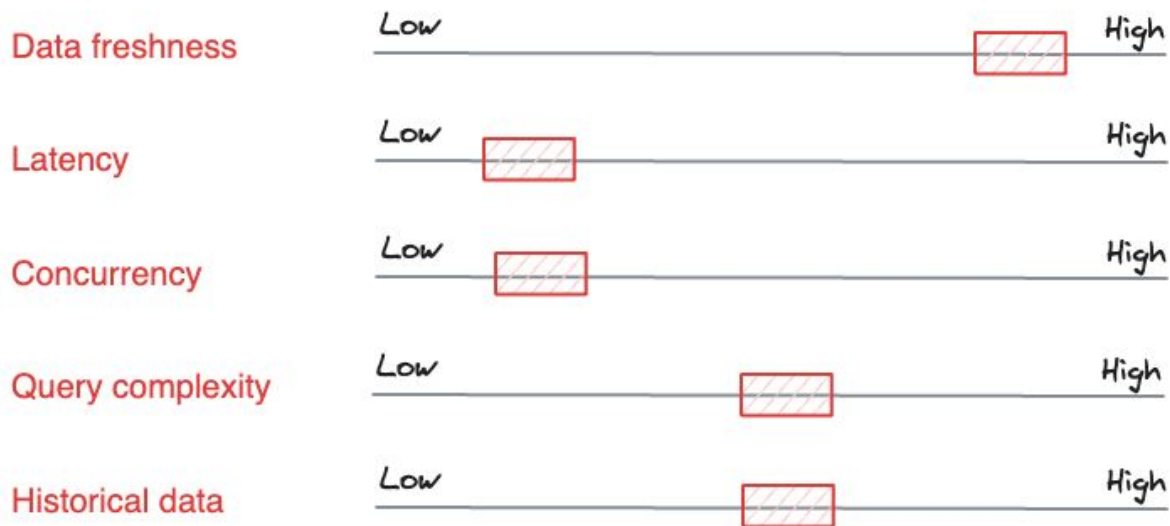
Redpanda

Examples

- Infrastructure monitoring dashboards
  - E.g Grafana
- IoT dashboards
- KPI dashboards

Must haves:

❏ Data freshness
❏ Low latency

	Low		High
Data freshness			▨
Latency	▨		
Concurrency	▨		
Query complexity		▨	
Historical data		▨	

Technology choices

Time series databases and full-text search engines.
E.g Prometheus, Elasticsearch, OpenSearch, InfluxDB

Ad-hoc querying large data sets

Run quick experimentations

On moving data

- Explores large amounts of data without a specific end goal.

- Ad-hoc querying to pose on-the-fly questions and retrieve immediate answers from vast datasets.

- Users are provided with interactive dashboards and visualizations to enable quick slicing, dicing, and drilling down on data.

Redpanda

Examples

- Rapid experimentation on large data sets.
  - BI and reporting
  - Ad-hoc SQL

- Troubleshooting and RCA
  - E.g. An SRE is combing through last-hour data to determine the cause of an incident.

Must haves:

- ❏ Historical data
- ❏ Support for complex queries

	Low	High
Data freshness		
Latency		
Concurrency		
Query complexity		
Historical data		
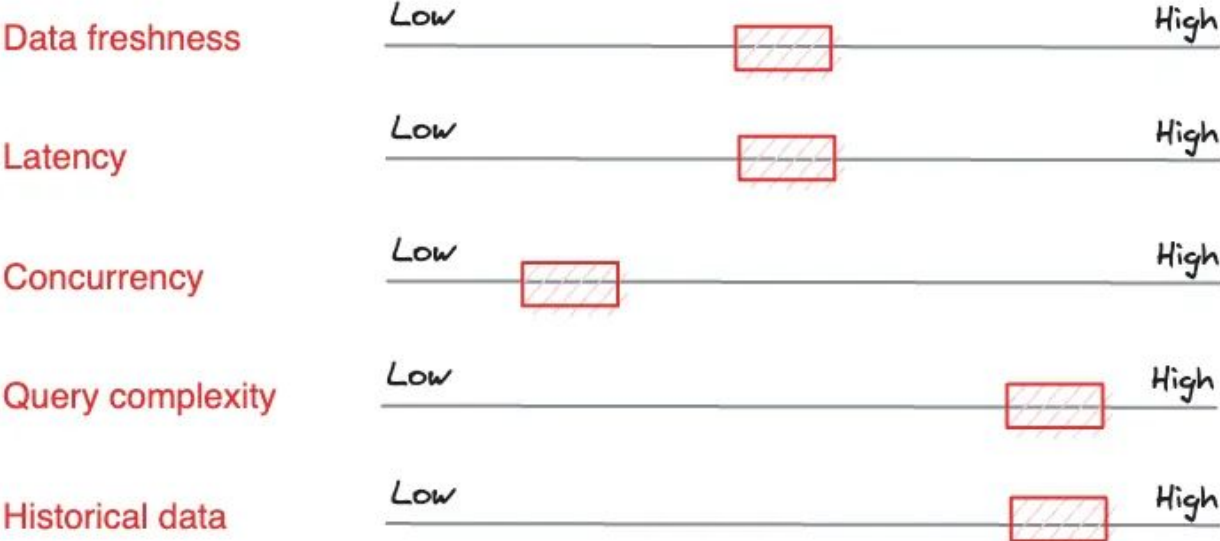
Technology choices

- Federated query engines
  - E.g Presto, Trino

- Data Lakehouse platforms

Streaming databases

- Combines stream processing + indexing of databases
- Appears to play nicely on all sliders

Summary

When you have a new use case for real-time analytics:

1. Use the five sliders to assess the use case and calculate weights for each characteristic.

2. Choose the most appropriate technology choice based on that.

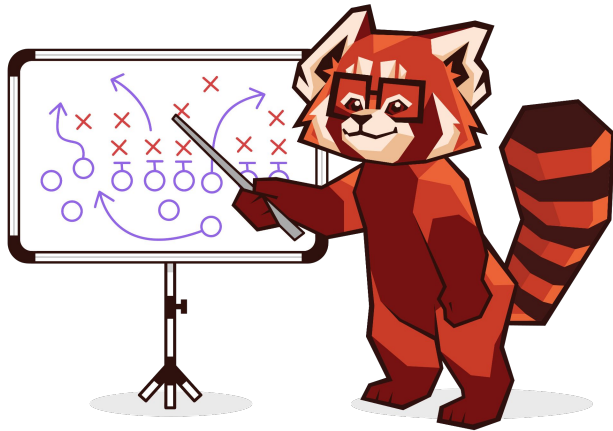   It's okay to have some overlapping.

Get a Free Copy!

Compliments of Redpanda

The State of
Streaming Data

2023-24   Independent analysis of the size, shape and
scope of modern streaming data deployments.

Commissioned by
Redpanda

Redpanda University

Free, self-paced online learning
https://university.redpanda.com

- Learn the fundamentals of data streaming and Redpanda

- Install Redpanda and use the rpk CLI to configure it

- Create producers and consumers in Java, Python and NodeJS

- Sign up today for free!

Keep Learning

Docs
Get a peak under the hood.
https://docs.redpanda.com/

Blogs
Keep up to date with Redpanda.
https://redpanda.com/blog

Slack
Engage with our community.
https://redpanda.com/slack

Code
Check out the source.
https://github.com/redpanda-data

This is the way

Thanks for joining!

Let's keep in touch

𝕏 @redpandadata

⊙ redpanda-data

in redpanda-data

✉ hello@redpanda.com

Redpanda