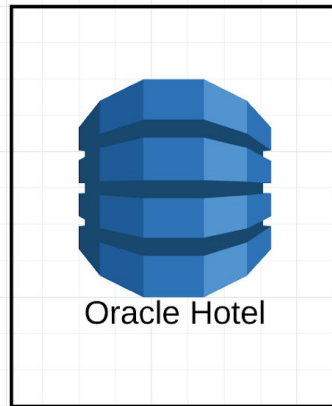
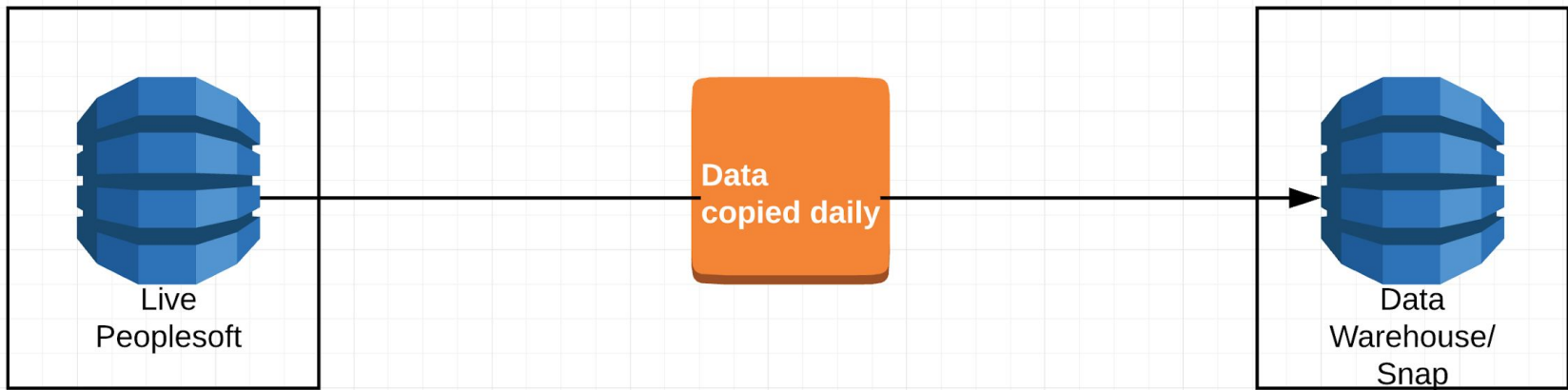


External Data and Events

Ian Whitney - ASR Custom Solutions
@ian_whitney tech-people-umn.slack.com
https://z.umn.edu/external_data

ASR and External Data

- Mostly PeopleSoft
- Web Applications



Databases within a ☐ can share data easily.

Sharing data with a database in a different ☐ requires a Database Link.

More on them later.

Example Problem



**Show a student
their classes
&**

**Alert their advisor
if they are off track**

State

The data as it is now

What we normally think of when we think of databases. We run a query, we get some data back. That is **State**.

Events

An alteration of the State

- Inserting a new row
- Updating values in a row
- Deleting a row

Each is an **Event** that changes our **State**

The problem

Student Enrollment - State

Class	Grading Method
Small Group Communication	A-F
Creativity and Intelligence	Pass/Fail

The problem

Student Enrollment - Events

- Student registered for Small Group Communication, A-F
- Student registered for Creativity and Intelligence, A-F
- Student registered for Organic Chemistry, A-F
- Student changes grading for Organic Chemistry to Pass/Fail
- Student changes grading for Creativity and Intelligence to Pass/Fail
- Student drops Organic Chemistry

Metrics for Success



What

Accurate Data

When the student looks at the web page, they see their current classes.

How

Event Driven Actions

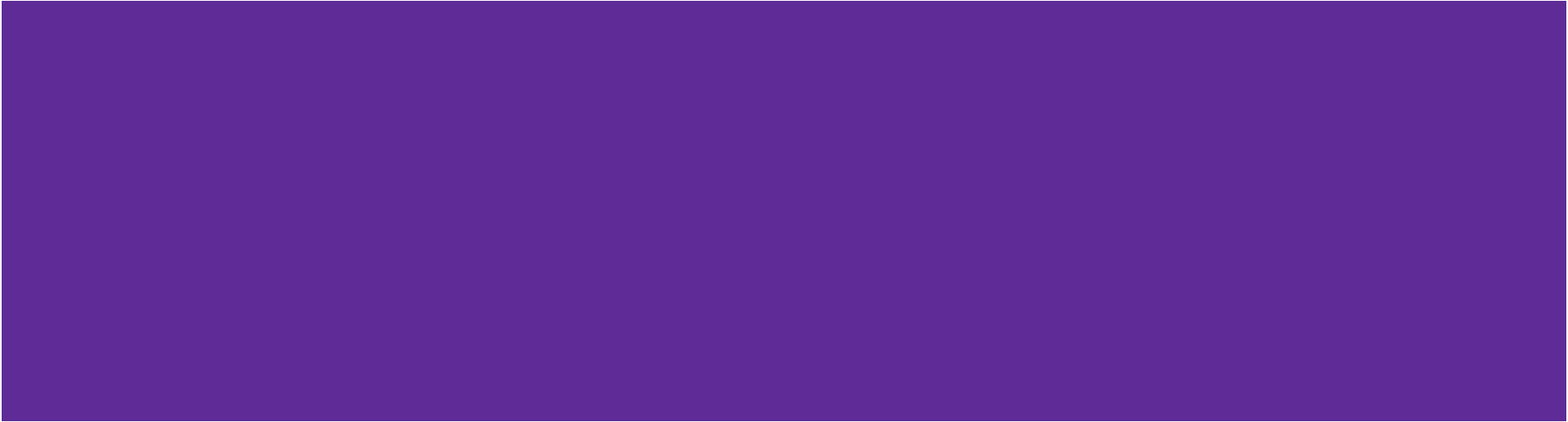
When the student changes their classes, their advisor is notified.

Resiliency

PeopleSoft is Down!

If PeopleSoft is down our application can continue to work.

Solutions



All of these work!

ETL - Extract/Transform/Load



ETL

- Select data from the DW/Snap
- Transform the data if needed
- Insert that data into new tables

ETL

Pros

- Simple mechanism, “Select Data, Insert It”
- Lets you transform the data in to the shape you need
- You’re not restricted to Oracle and can Load data in to any RDBMS you want

Cons

- Can be slow, depending on the amount of data you’re ingesting and the complexity of the ‘transform’
- Failures can lead to missing or no data

Metrics for Success

What

Accurate Data

When the student looks at the web page, they see their current classes.

Kinda. We can see yesterday's data.

How

Event Driven Actions

When the student changes their classes, their advisor is notified.

Kinda. We can keep the data we loaded yesterday and compare it to the data we load today.

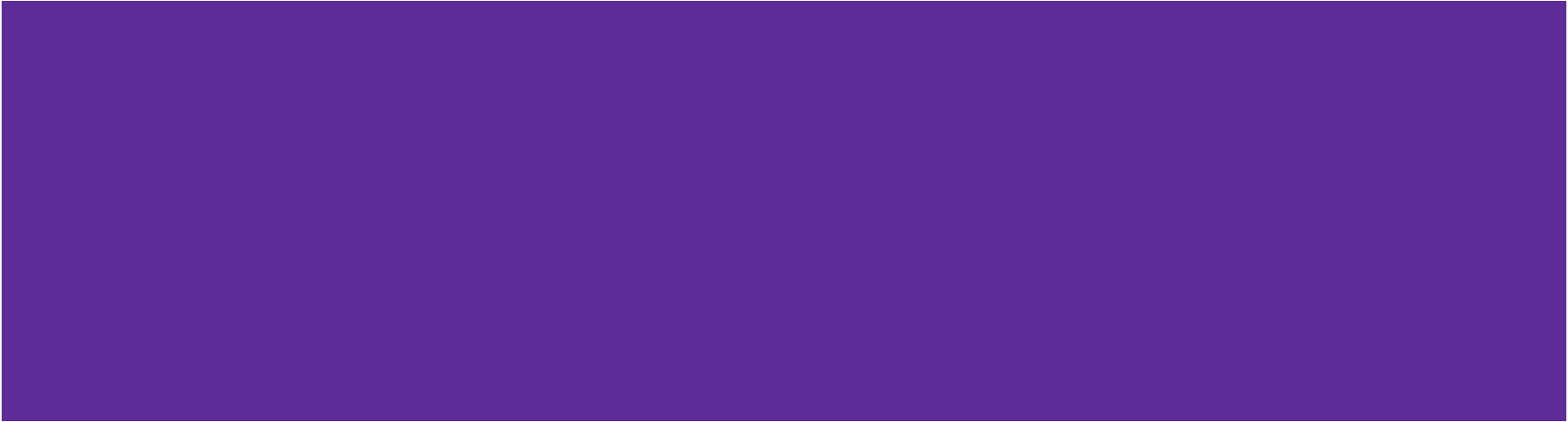
Resiliency

PeopleSoft is Down!

If PeopleSoft is down our application can continue to work.

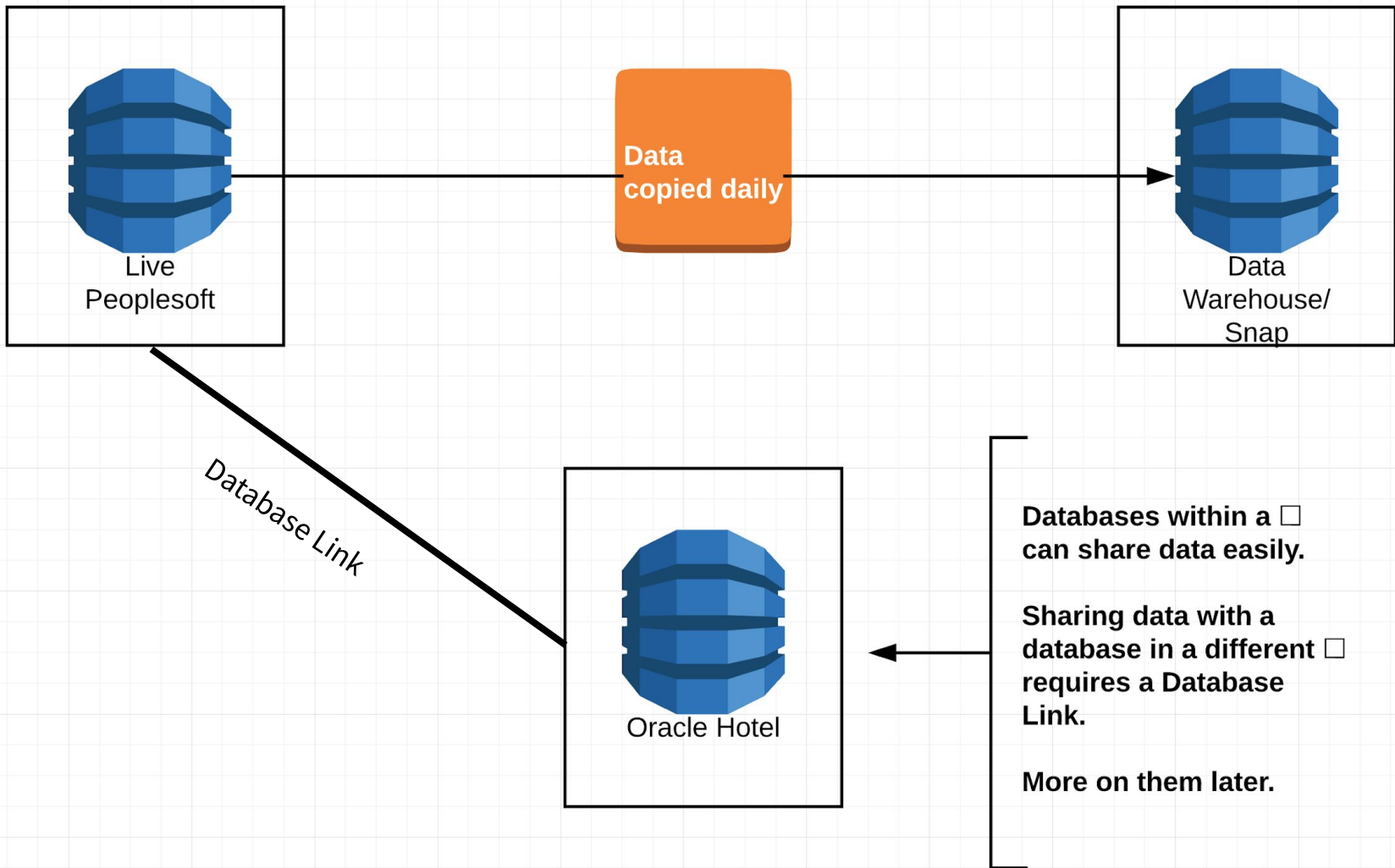
Yes! Works fine unless PeopleSoft goes down doing Extraction

Live Querying



Live Querying

- Query Live PeopleSoft directly via Database Link



Live Querying

Pros

- We know that the data is up to date

Cons

- DW pre-joins some data for you. In Live PeopleSoft you have to join it yourself
- Non-Production Live PeopleSoft goes down for maintenance regularly
- Database Links

Metrics for Success

What

Accurate Data

When the student looks at the web page, they see their current classes.

Yes!

How

Event Driven Actions

When the student changes their classes, their advisor is notified.

No*

Resiliency

PeopleSoft is Down!

If PeopleSoft is down our application can continue to work.

No

Live Querying & ETL



Live Querying & ETL

- ETL code that queries Live PeopleSoft and writes to local tables
 - Materialized Views
- Run the ETL as frequently as the application needs
 - 5 minutes to daily

Live Querying & ETL

Pros

- Removes need for Database Links
- Data is 'Near Live'

Cons

- DW pre-joins some data for you. In Live PeopleSoft you have to join it yourself
- Data topography and freshness can be hard to reason about

Metrics for Success

What

Accurate Data

When the student looks at the web page, they see their current classes.

Yes! (depending on ETL frequency)

How

Event Driven Actions

When the student changes their classes, their advisor is notified.

No*

Resiliency

PeopleSoft is Down!

If PeopleSoft is down our application can continue to work.

Yes!

No*

[Hello, Flashback]

Oracle Flashback

Lets you query State at a point in
time

```
SELECT *  
  
FROM student_enrollment  
  
AS OF TIMESTAMP CURRENT_TIMESTAMP -  
5  
  
WHERE emplid = '2411242';
```

Flashback: State

Student Enrollment - Events

- Student registered for Small Group Communication, A-F
- Student registered for Creativity and Intelligence, A-F
- Student registered for Organic Chemistry, A-F
- Student changes grading for Organic Chemistry to Pass/Fail
- Student changes grading for Creativity and Intelligence to Pass/Fail
- Student drops Organic Chemistry

Oracle Flashback

Lets you query Events over time

```
SELECT *  
  
FROM student_enrollment  
  
VERSIONS BETWEEN  
  
SCN MINVALUE AND MAXVALUE  
  
WHERE emplid = '2411242';
```

Flashback Events

Student Enrollment - Events

Action	Class	Grading Method	Started	Ended
I	Small Group Communication	A-F	2019-10-03 00:00	
I	Creativity and Intelligence	A-F	2019-10-03 00:10	2019-10-03 00:40
I	Organic Chem	A-F	2019-10-03 00:20	2019-10-03 00:30
U	Organic Chem	Pass/Fail	2019-10-03 00:30	2019-10-03 00:50
U	Creativity and Intelligence	Pass/Fail	2019-10-03 00:40	
D	Organic Chem	A-F	2019-10-03 00:50	

Oracle Flashback

Pros

- Events and State from any time! This is basically magic.
- Straightforward to write queries that tell you when the state of a column (such as course grading) is changed

Cons

- Can only go as far back as data is kept
 - Currently 2 hours
- Flashback will tell you that a row has been updated, but not what changed
- Oracle only

Event Streams



Event Streams

An immutable log of changes to
your data's state

Event Streams

Student Enrollment - Events

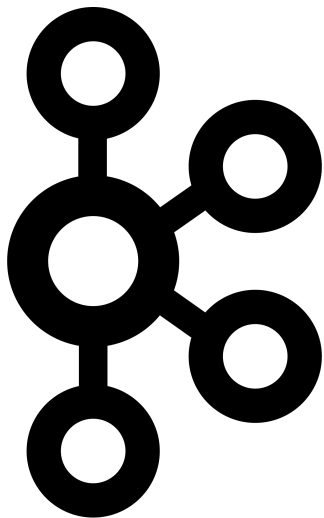
- Student registered for Small Group Communication, A-F
- Student registered for Creativity and Intelligence, A-F
- Student registered for Organic Chemistry, A-F
- Student changes grading for Organic Chemistry to Pass/Fail
- Student changes grading for Creativity and Intelligence to Pass/Fail
- Student drops Organic Chemistry

Flashback Events

Student Enrollment - Events

Action	Class	Grading Method	Started	Ended
I	Small Group Communication	A-F	2019-10-03 00:00	
I	Creativity and Intelligence	A-F	2019-10-03 00:10	2019-10-03 00:40
I	Organic Chem	A-F	2019-10-03 00:20	2019-10-03 00:30
U	Organic Chem	Pass/Fail	2019-10-03 00:30	2019-10-03 00:50
U	Creativity and Intelligence	Pass/Fail	2019-10-03 00:40	
D	Organic Chem	A-F	2019-10-03 00:50	

Event Streams and Flashback



- Use Flashback to build an Event Stream
- Write the Event Stream to Apache Kafka
- Consumers read the data in Apache Kafka and do...whatever

Demo [later!]

Event Streams

Pros

- Live State and Events
- Retention allows you to go back in time as far back as you'd like

Cons

- Operationally complex

Metrics for Success

What

Accurate Data

When the student looks at the web page, they see their current classes.

Yes!

How

Event Driven Actions

When the student changes their classes, their advisor is notified.

Yes!

Resiliency

PeopleSoft is Down!

If PeopleSoft is down our application can continue to work.

Yes!

Overview

What

ETL: **Kinda**

Live: **Yes**

Live + Etl: **Yes**

Event Streams: **Yes**

How

ETL: **Kinda**

Live: **No***

Live + Etl: **No***

Event Streams: **Yes**

Resiliency

ETL: **Yes**

Live: **No**

Live + Etl: **Yes**

Event Streams: **Yes**

Further Resources

<https://dataintensive.net/>

O'REILLY®

Designing Data-Intensive Applications

THE BIG IDEAS BEHIND RELIABLE, SCALABLE,
AND MAINTAINABLE SYSTEMS



Martin Kleppmann

Even Further Resources

- The Simplest Useful Kafka Connect Data Pipeline in the World
 - <https://www.confluent.io/blog/simplest-useful-kafka-connect-data-pipeline-world-thereabouts-part-1/>
- Streaming Data from Oracle to Kafka
 - <https://rmoff.net/2018/12/12/streaming-data-from-oracle-into-kafka-december-2018/>
- Ideas for Event Sourcing in Oracle
 - <https://medium.com/@FranckPachot/ideas-for-event-sourcing-in-oracle-d4e016e90af6>
- Confluent/Kafka Docker Quickstart
 - <https://docs.confluent.io/current/quickstart/ce-docker-quickstart.html>

Questions

Ian Whitney - ASR Custom Solutions
@ian_whitney tech-people-umn.slack.com
https://z.umn.edu/external_data