

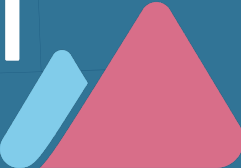


COMMUNITY DAY

AOTEAROA

*Implementing Customer Statements
at scale with AWS Lambda*

Mantel group



brad.jacques@mantelgroup.co.nz

Principal Consultant

Perth

Magnetic Island

Brisbane

Sydney

Melbourne

Hobart

Auckland

Queenstown

✓ 900+ team members

✓ Best Place to Work in 2021 and 2022

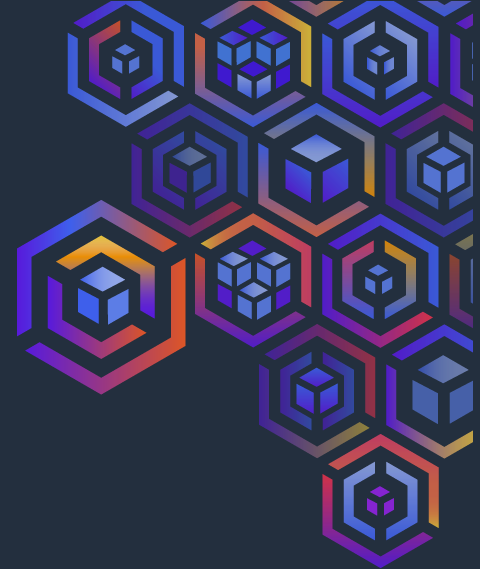
✓ Operating at over 140 clients

The Discovery Phase

- Define the problem statement
- Uncover the unknowns
- Start capturing requirements
- Agree the ways of working
- Who are the key decision makers
- AGILE isn't agile when there is no agility

The Team

- 1 x Delivery
- 2 x UX Design
- 2 x Digital
- 2 x Data



The Problem Statement

- Regulatory requirement to deliver customer statements
- Fixed deadline
- Historical data for trending and comparison
- Customer must have direct access without authentication
- Statements accessible by all

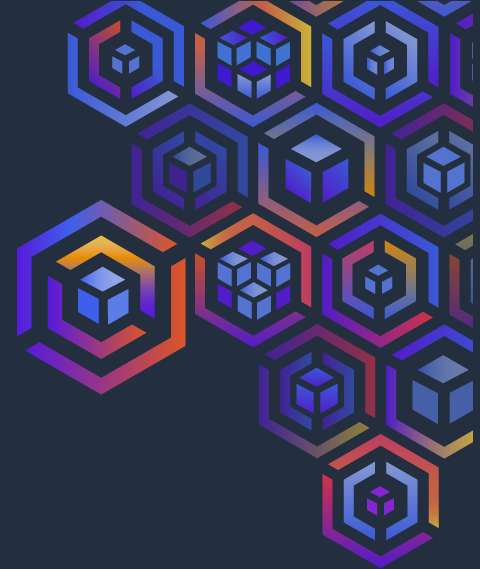


Technical Spikes

- Figma to iterate UX designs
- PDF generation
- Convert HTML/CSS/Fonts to PDF
- Snowflake connectivity
- Infrastructure (IaC)

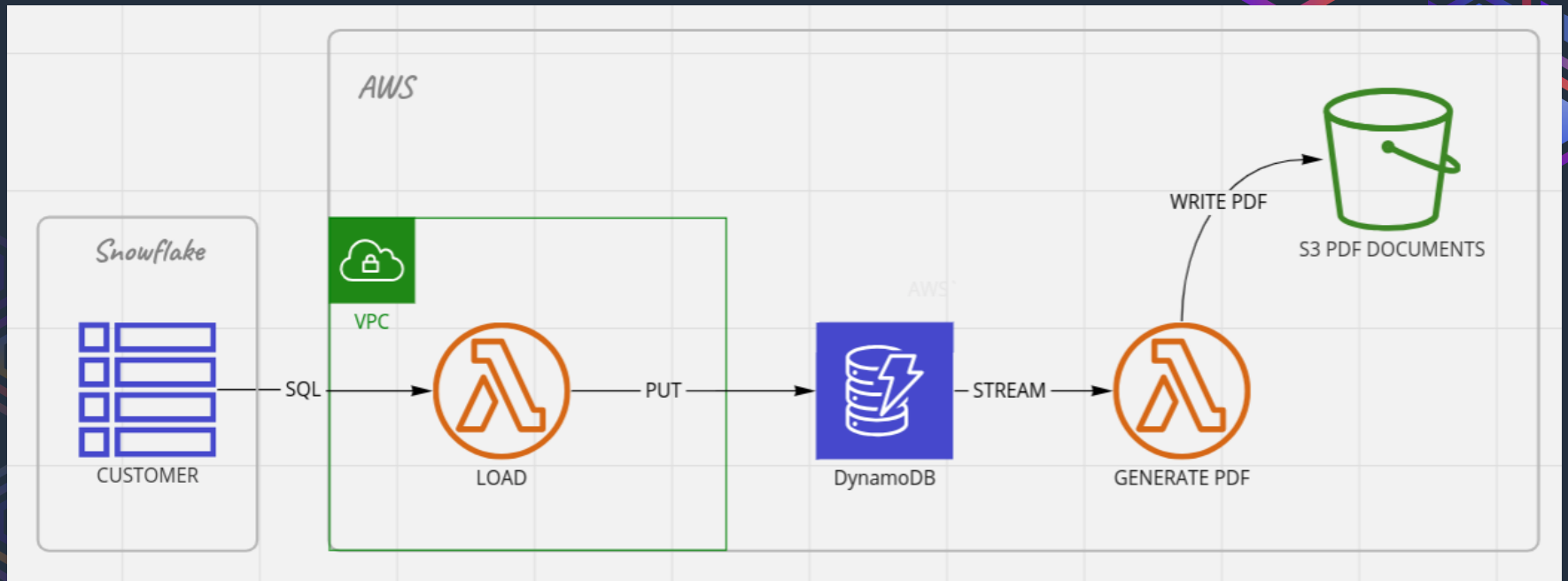
Data Analysis

- Snowflake data analysis looking for peak loads
- Stress test early to validate design



aws
COMMUNITY DAY
AOTEAROA

Just in time Architecture



Validate the Design

- Stress testing
- Find historical peak monthly volumes
- Forecast and execute at $N \times$ times peak
- Customer profiling
- Ensure data is available for frequent testing

Datasets

- One customer for each profile
- Largest month
- Forecast $N \times$ largest month

PDF Generation & S3 Upload

Time for one lambda execution to process N accounts

Number of accounts	Exec time seconds	Forecast 100,000 docs
1	2	55.5 hours
10	7	19.4 hours
100	<i>TBA</i>	

Querying Snowflake

How many records can be returned by a single query in one lambda execution

Query number of rows	Exec time seconds	Memory used MB
1	1	121
100,000	16	349
500,000	15	1,161
1,000,000	32	2,024
20,000,000	failed	Exceeded 10,000 limit



COMMUNITY DAY
AOTEAROA

Customer Data Analysis

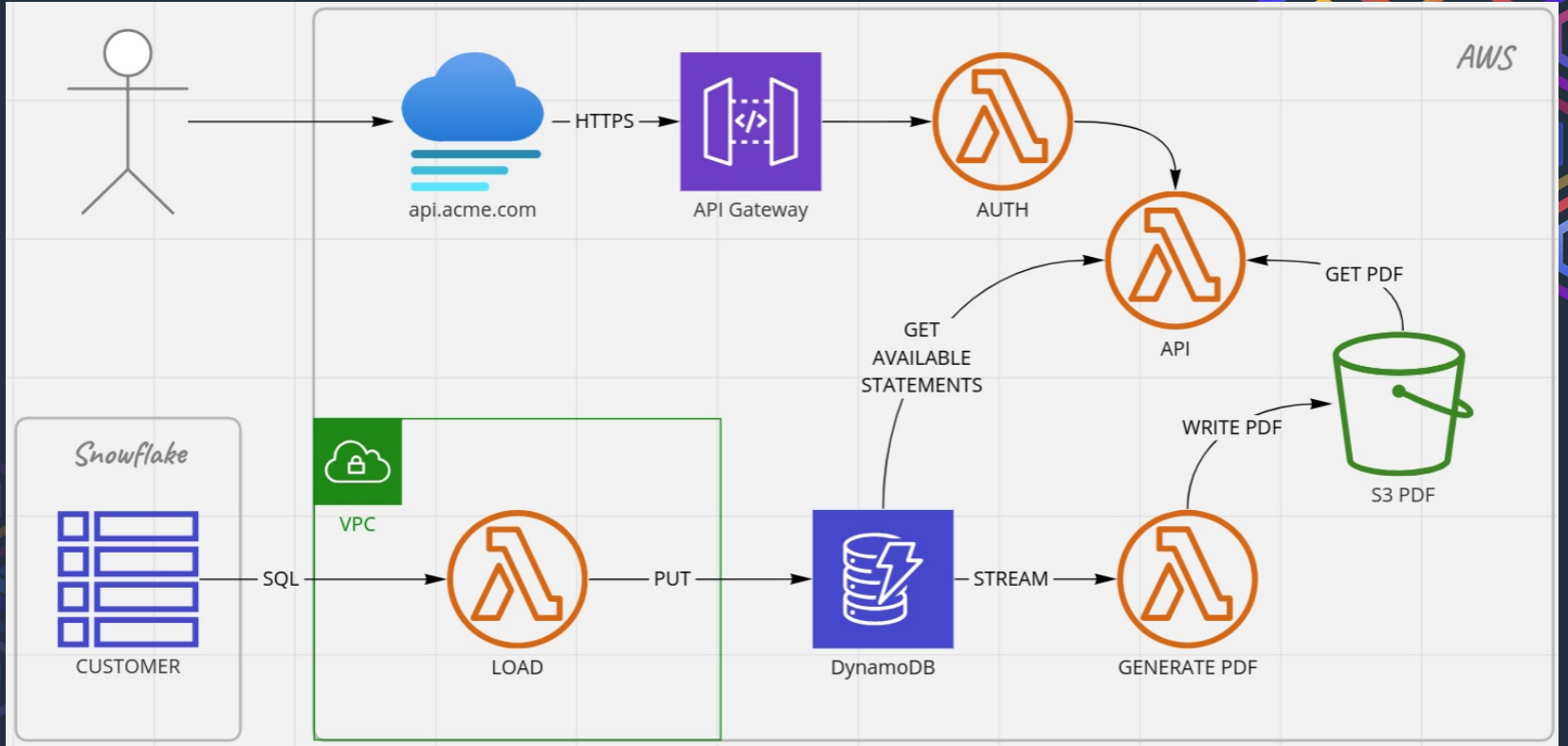
Within each Month we can see data skew

Transactions per account	% population
Up to 20,000	99.9
Over 20,000	0.1



COMMUNITY DAY

AOTEAROA



Render design to PDF

- Puppeteer provides an API to control a headless Chromium browser
- Deployed as a Lambda layer
- Header / footer page numbers
- Different layout CSS for cover page and transactions pages
- Requires 2 PDF's to be generated then merged

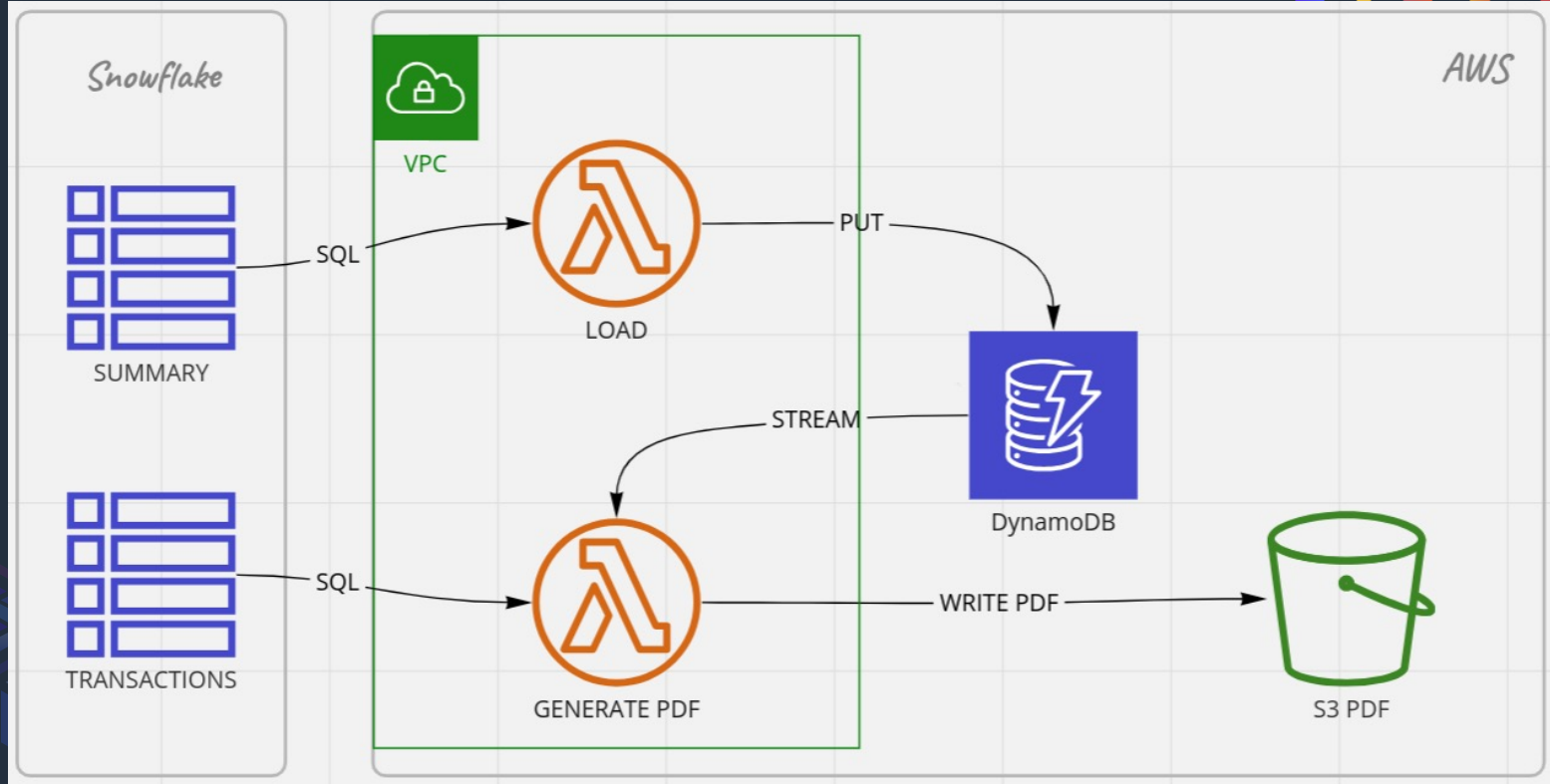
npm

- `chrome-aws-lambda`
- `pdf-lib`
- `handlebars`
- `chart-js`



COMMUNITY DAY

AOTEAROA



Tuning AWS Lambda

- How many DDB events are processed per execution
- Memory consumption per execution
- Number of CPU cores
- How many lambda to execute in parallel
- Log and drop the outliers < 0.1% population

event stream

batch_size = 10

parallelization_factor = 10

bisect_batch_on_function_error = true

maximum_retry_attempts = 5

Tuning Puppeteer

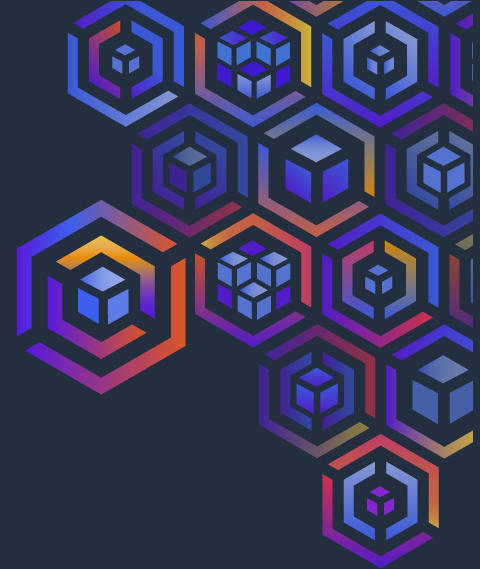
- Go-Live ~38 PDF documents per second uploaded to S3
- Disabled 34 chromium features on start up
- No file I/O use in-memory Buffers
- In-line CSS
- In-line Fonts (base64)
- In-line SVG (base64)
- Use image sprites (filters change hue)

npm

- `html-minifier-terser`

Reconciling the system

- Reconcile using Month ID (e.g. 202309)
- Separate lambda is responsible for
 - Query snowflake by monthId (index)
 - Query DDB by monthId (pk)
 - Query S3 by monthId (path)





COMMUNITY DAY

AOTEAROA

Summarising

- Do the simple thing first
- Small teams with fast feedback loop (showcase often)
- Identify risk early, shift left, and spike
- Continuously measure performance, and stress test
- Isolate context boundaries (e.g. lambdas)
- Solution must prove itself correct

