Project Data Model Definition - **Sprint project**

The SA who normally works with LogisticsZero won the lottery and promptly quit. We've taken over the account and have been immediately tasked with sizing a MongoDB cluster for their new project codenamed *Sprint*.

LogisticsZero is a shipping company that differentiates themselves from other logistics companies by using more environmentally responsible modes of transport.

This is a project to replace their package tracking system. Here are the notes the other SA provided us with before leaving. Use this information along with these mgeneratejs scripts in order to determine what size cluster LogisticsZero will need. Use this template for your sizing calculations. Remember that this is for Enterprise Advanced, so you are not constrained by the standard instance sizes in Atlas.

The customer needs to know:

1. What should their production cluster topology look like? (Replica set, sharded cluster, number of shards, etc.)
2. How many 128GB ram pools will they need to buy for **production only**?
3. How fast does their storage solution need to be? (Measured in IOPS)
4. How many CPU cores should they assign to each node?

In order to figure this out, you'll need to follow these steps:

1. Determine an appropriate schema and generate a sample dataset with mgeneratejs and load it into MongoDB (not Atlas, just use your locally installed mongod)
2. Determine which indexes should be created based on access patterns described below. Create those indexes for your sample dataset.
3. Use data size calculations based on your sample data as well as the given information to determine the MongoDB cluster size.

Notes:

- Size the system for this moment in time, don't worry about including additional resources for future data growth

To submit this project:

Create a folder in Google drive. Put your mgeneratejs templates (or any other code you might have used to generate data) as well as your sizing spreadsheet in this folder.

Also include a document explaining why you believe your schema to be the optimal design for this customer's use case.

As usual, reach out to our dedicated slack channel if you have any questions.

Here is the information the previous SA provided:

Volume

- We are moving 3 billion packages every 4 months

Delivery data

For tracking each of the deliveries we need to know the following details:

- **Tracking Number:** 1Z Number (*) - *Automatically generated*
- **Package RFID**: RFID code of the package, internal - *Automatically generated*
- **Status:** Package Status, visible to customers. Defines the state of the package on the delivery.
    - Can be : Received, AtOrigin, InCustoms, InTransit, AtDestinationWarehouse, OutOnDelivery, Delivered, Exception, Returned
- **Scheduled Delivery:** ISO Date

- **Shipped From:** Address Information (**)
- **Optional - Shipped by email**: Email of sender
- **Optional - Shipment reference**: The identifier specified by the sender during the shipping process. This reference could be a purchase order (PO) number, a customer number, a company name, a Bill of Lading number, or a phrase that identifies that shipment -- such as "10 crates" or "gift for Mom." Shipment references can be any combination of letters and numbers up to 35 characters.
- **Shipped To:** Address Information (**)
- **Optional - Shipped to email**: Email of the recipient
- **Shipped or Billed on:** Shipping/Billing Date
- **Service Type:** Name of Service (Express plus, express, express saver, standard).
- **Weight:** Appropriate measure in kilograms.
- **Dimensions**
    - Length, width, height all in Centimeters
    - Oversized: Yes/No
- ***Optional***
    - **Declared value**: Value estimated in Dollars
    - **Multiple Packages:** Number of Packages
    - **Additional features:**
        - Carbon-neutral : Yes/No
        - Saturday delivery: Yes/No
        - No third Party delivery: Yes/No
        - Additional insurance : Yes/No
    - **VAT Number**

(*) LogisticsZero tracking numbers appear almost always with 14 digits.

(**) Addresses need the following details:

- Country
- State / Province
- City / Town
- Postcode
- Street
- Apartment & suite
- Optional - Department

RFID & Tracking number

When the package is received, a pseudo-random RFID is generated for it that makes it uniquely identifiable. This RFID is assigned to the package and will be used to track it down and scanned all the way from source to destination.

Now, for customers and 3rd party companies to track the shipment down a LogisticsZero tracking number will be generated too. Customers will use this number to get the status of their delivery.

Tracking history

Tracking details displayed include:

- **Date:** Date of the activity shown in that line item (yyyy-mm-dd)
- **Time:** Time of the activity shown in that line item (hh:mm:ss)
- **Location:** Location of the activity shown in that line item. Can be a warehouse name like “Kent, UK” or “Heathrhow Airport” etc.
- **GPS Location**: Approximated Location (lat, long) worldwide of the package.
- **Activity:** An individual event (e.g., origin scan, departure scan, export scan, customs entry, customs exit, import scan, destination scan, out on delivery, delivered, (optionally held)) in the shipment process. From the list of scanned

activities, you can build a picture of the shipping progress of each package from origin to delivery destination.

- **Internal Activity**: Yes / No.
  Not visible to customers. Some actions like loading and unloading from a truck/airplane/ship should not be visible but still need to be tracked.

For each package we need to track the history. When customers enter the LogisticsZero website they can track a package and list all the details from the package such as origin and destination as well as the activity:

This tracking can also be done via the mobile app, no need to be logged in, just know the package tracking id.

To get detailed information about the payment they will need to login and get to a different window.

Lifetime of a delivery

When a package is received, the new package information will be inserted on the system. All shipping and delivery data filled by the LogisticsZero employee.

The system will create an entry in the tracking history from where it's being sent and the activity will be tagged as "origin scan".

All throughout the journey there will be scans when loading the package on LogisticsZero vehicles, unloading, arriving at the warehouse etc. Packages going overseas will have the events for customs entry, customs exit etc.

The final steps for all deliveries are destination scan, out on delivery and finally delivered.

If a package can't be delivered it will go back to the LogisticsZero facility and delivery will be retried. If finally unsuccessful after a few attempts, the package will hold the status of help and will be stored at the latest LogisticsZero facility.

Tracking

Customers can track their packages by using the LogisticsZero tracking number.

We can also track by combination of origin and destination data plus the reference number. The data required is the origin country, date shipped (range), destination country and the shipping reference.

Payments

Payments are also tracked on the old system, and for our Sprint project we also want that information to be registered on the system.

- Tracking number
- Payment method, can be one of the following: Paypal or Payment card.
- If Payment card:
    - Card Type
    - Card number
    - Expiration month and year
    - CVV
    - Billing address (**)
- If Paypal:
    - Paypal operation hash
- If cash:
    - LogisticsZero Access point details
        - ID - 7 digit number
        - Address
        - Postcode
        - Town
        - City
        - Country
- VAT ID
- Optional Customs broker ID
- Optional Promo code

Each package delivery will need to have payment information registered on the system.

Payment data must be stored for 4 months.

Summary of how data is accessed

- LogisticsZero Tracking number will be used by customers via the site and app as well as 3rd parties to track the delivery.
- Customers can also track packages by using the origin country, date shipped (range), destination country and the shipping reference, each of these are required.

- Whenever the package is moved, loaded, unloaded etc. the item will be scanned and the RFID will be used for that purpose.
- Payment data is read by shipment tracking number. Payments are inserted once and read 2 times before the final shipment (one to verify payment info is correct and the second time to push the data to an archival system), payment information never gets updated.
- The data that needs to be responding really quickly is data for the last 10 days.
- Ideally data should be indexed for the entire 4 month period.

Workload details

- 3000 million new packages every 4 months.
    - Packages older than 4 months are moved to archive.
- Desired latency for reads on data for the last week:
    - 100 ms, 99 percentile.
- Desired latency for writes :
    - 100 ms, 99 percentile.
- Expected 'Writes' per second :
    - 12.000 writes per second
    - Most of these writes are updates to the package while on transit.
- Expected 'Reads' per second :
    - 29.000 reads per second
    - Many of these reads are done by RFID readers while on transit to know how to route it.