Case Study #2

Memi Lavi
www.memilavi.com

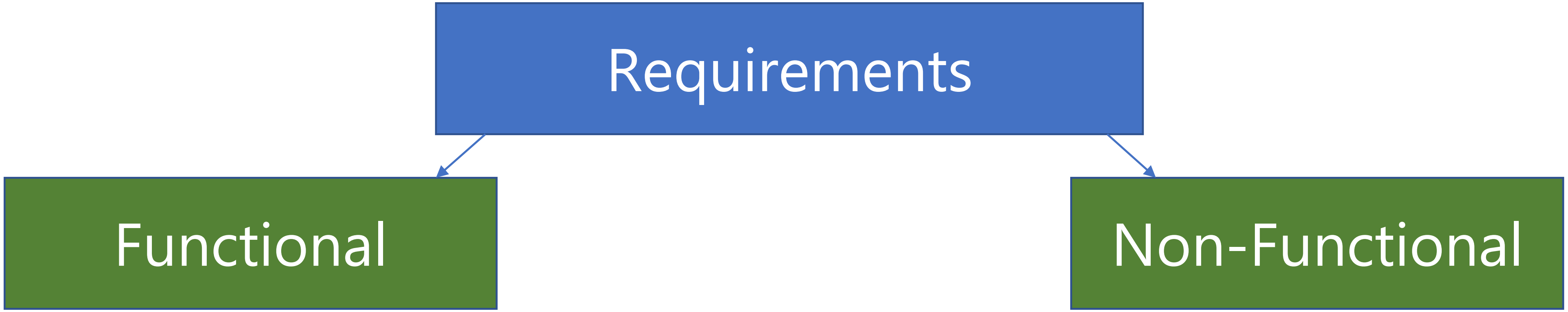
Muscar.auto

The Real Autonomous Car

Muscar.auto

- Manufactures autonomous systems for vehicles
- Has >10,000 vehicles on the roads right now
- Expects more than 200,000 vehicles by end of year
- Needs to reliably receive telemetry from cars and display data about them

Muscar.auto



Requirements

Functional

What the system should do

1. Web Based
2. Receive telemetry from cars (location, speed, breakdowns, etc)
3. Store telemetry in a persistent store
4. Display dashboards summarizing the data
5. Perform analysis on the data

Non-Functional

What the system should deal with

NFR - What We Know

1. Data intensive system
2. Not a lot of users
3. A lot of data
4. Performance is important

NFR - What We Ask

1. *“How many expected concurrent users?”*	10
2. *“How many telemetry messages received per second?”*	7,000
3." *What is the average size of message?”*	1KB
4. *“Is the message schema-less?”*	Yes

Muscar.auto

NFR - What We Ask

5. *"Can we tolerate some message loss?"*	Sort of…
6. *"What is the desired SLA?"*	Highest Possible

Data Volume

- 1 Message = 1KB
- 7,000 messages / sec = 7MB / sec

=> ~25GB / hr

=> ~605GB / day

=> ~221TB / year ← That’s a lot!

Retention Period

Defines for how long records are kept in the database

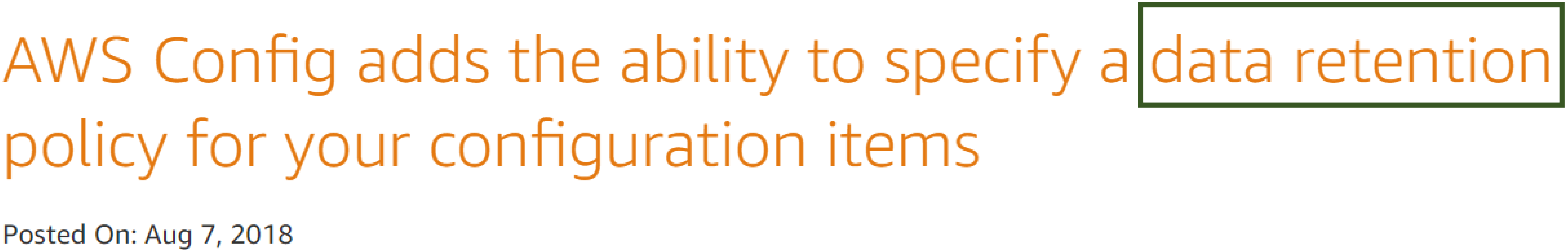
What happens to them after the retention period?

- Deleted
- Moved to archive data store

Retention Period

Motivation:

- Keep database from exploding
- Improve query performance



AWS Config adds the ability to specify a data retention policy for your configuration items

Posted On: Aug 7, 2018

Muscar.auto

Retention Period

Muscar needs two types of data:

- Operational, near-real-time (location, speed, etc.)
- Aggregated and ready for analysis (BI – Business Intelligence)

Retention Period

Data Type	Used for...	Retention Period
Operational	Monitor real time data from cars. Performance is critical	
Aggregated	Reports, BI. Not real time, can be slower.	

Retention Period

Data Type	Used for...	Retention Period
Operational	Monitor real time data from cars. Performance is critical	1 week
Aggregated	Reports, BI. Not real time, can be slower.	Forever

Muscar.auto

Data Volume

- 1 Message = 1KB
- 7,000 messages / sec = 7MB / sec

  => ~25GB / hr

  => ~605GB / day

  => ~221TB / year

Data Volume

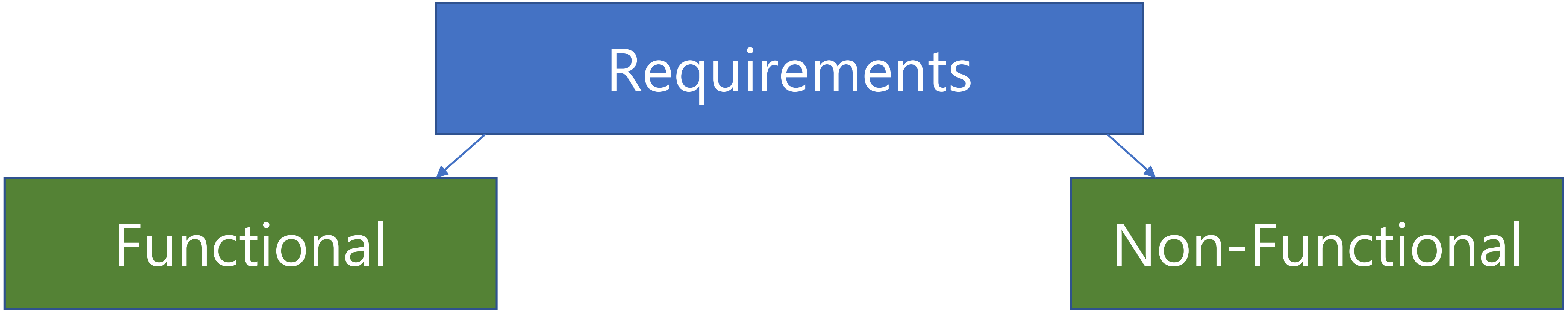
- 1 Message = 1KB
- 7,000 messages / sec = 7MB / sec

=> ~25GB / hr

=> ~605GB / day

=> ~4TB / week

Muscar.auto



Requirements

Functional

What the system should do

1. Web Based
2. Receive telemetry from cars (location, speed, breakdowns, etc)
3. Store telemetry in a persistent store
4. Display dashboards summarizing the data
5. Perform analysis on the data

Non-Functional

What the system should deal with

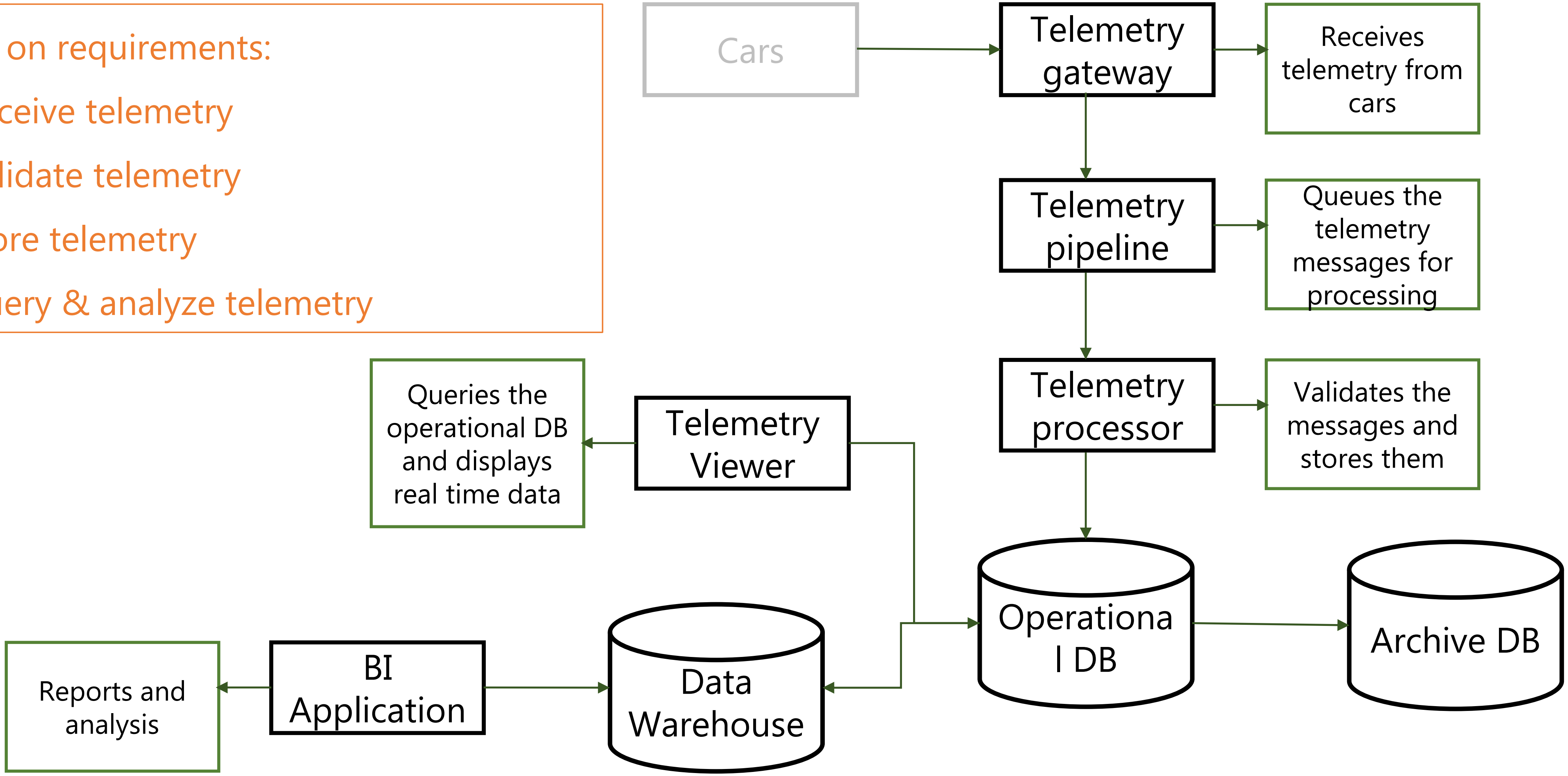
1. 10 Concurrent users
2. 7,000 msgs/sec
3. Max data in the operational DB: 4TB
4. Mission critical
5. Performance is critical

Muscar.auto

Components

Based on requirements:

1. Receive telemetry
2. Validate telemetry
3. Store telemetry
4. Query & analyze telemetry

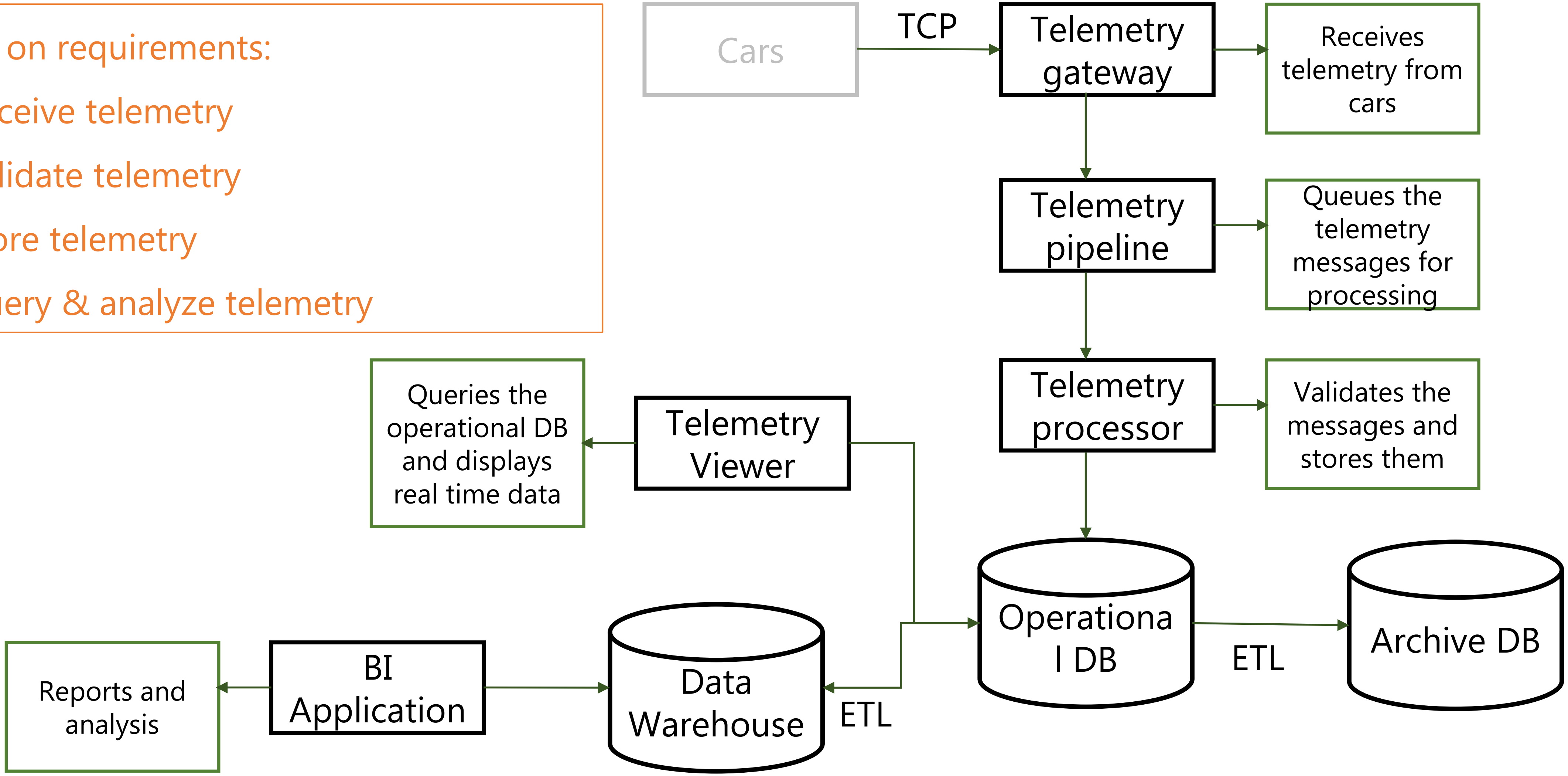

Cars
Telemetry gateway
Receives telemetry from cars
Telemetry pipeline
Queues the telemetry messages for processing
Telemetry processor
Validates the messages and stores them
Queries the operational DB and displays real time data
Telemetry Viewer
Operational DB
Archive DB
Reports and analysis
BI Application
Data Warehouse

Muscar.auto

Messaging

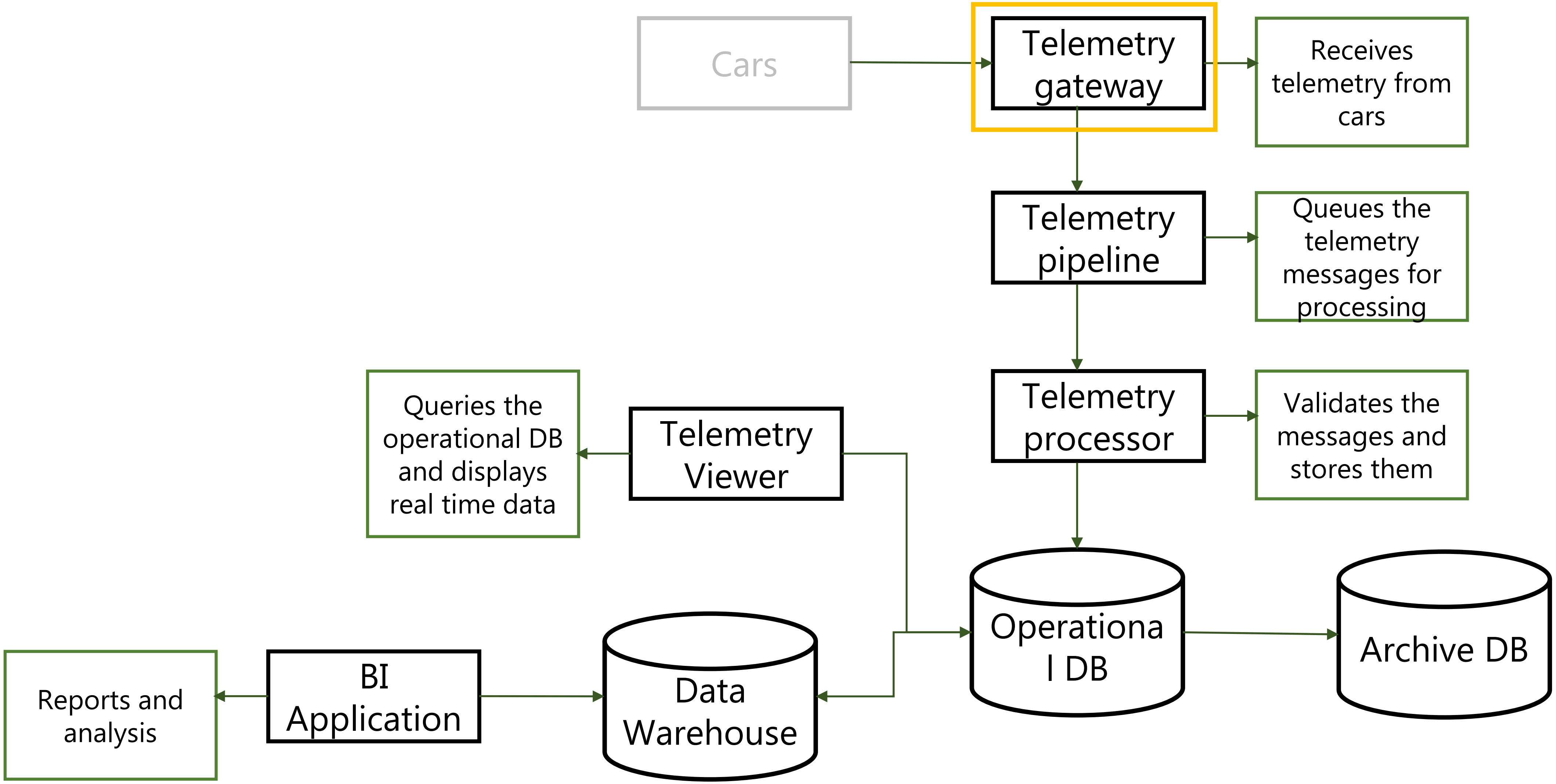
Based on requirements:

1. Receive telemetry
2. Validate telemetry
3. Store telemetry
4. Query & analyze telemetry


Cars
TCP
Telemetry gateway
Receives telemetry from cars
Telemetry pipeline
Queues the telemetry messages for processing
Telemetry processor
Validates the messages and stores them
Operationa
l DB
ETL
Archive DB
Telemetry Viewer
Queries the operational DB and displays real time data
ETL
Data Warehouse
BI Application
Reports and analysis

Muscar.auto

Components


Cars
Telemetry gateway
Receives telemetry from cars
Telemetry pipeline
Queues the telemetry messages for processing
Telemetry processor
Validates the messages and stores them
Telemetry Viewer
Queries the operational DB and displays real time data
Operationa l DB
Archive DB
Data Warehouse
BI Application
Reports and analysis

Telemetry Gateway

What it does:

- Receives telemetry data from cars using TCP
- Pushes the telemetry data to the pipeline

Application Type

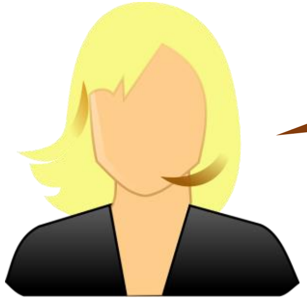
- Web App & Web API ✗
- Mobile App ✗
- Console ✓
- Service ✓
- Desktop App ✗

Technology Stack

Considerations:

- Load (7,000 msgs/sec)
- Performance
- Team’s current knowledge
- Environment (OS, etc)

Technology Stack



Our developers are familiar with Python, and are experts in JavaScript.
In addition, we use only Linux servers.

Python can't be used for the gateway

Too slow

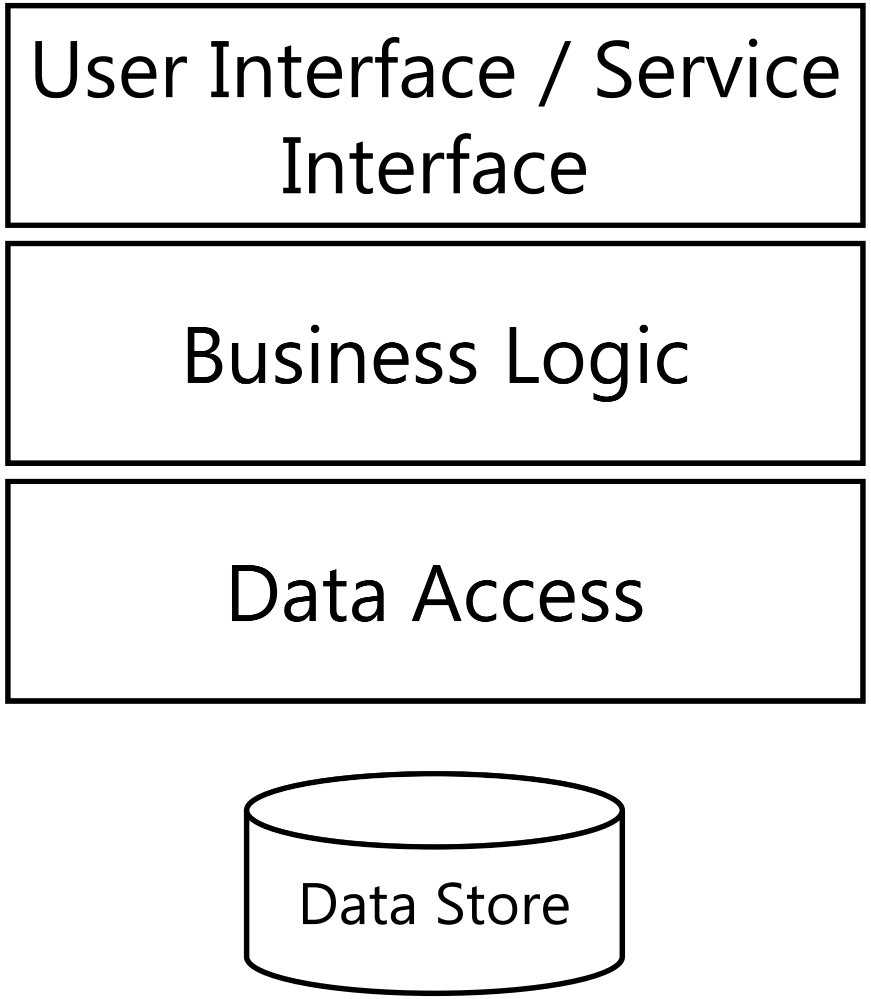
We look for a language with great performance, runs on linux, and leverages current skills (Python & JavaScript)

Technology Stack

- ✓ Great performance
- ✓ Runs on Linux
- ✓ Leverages JS skills

Architecture

Traditional:



User Interface / Service Interface

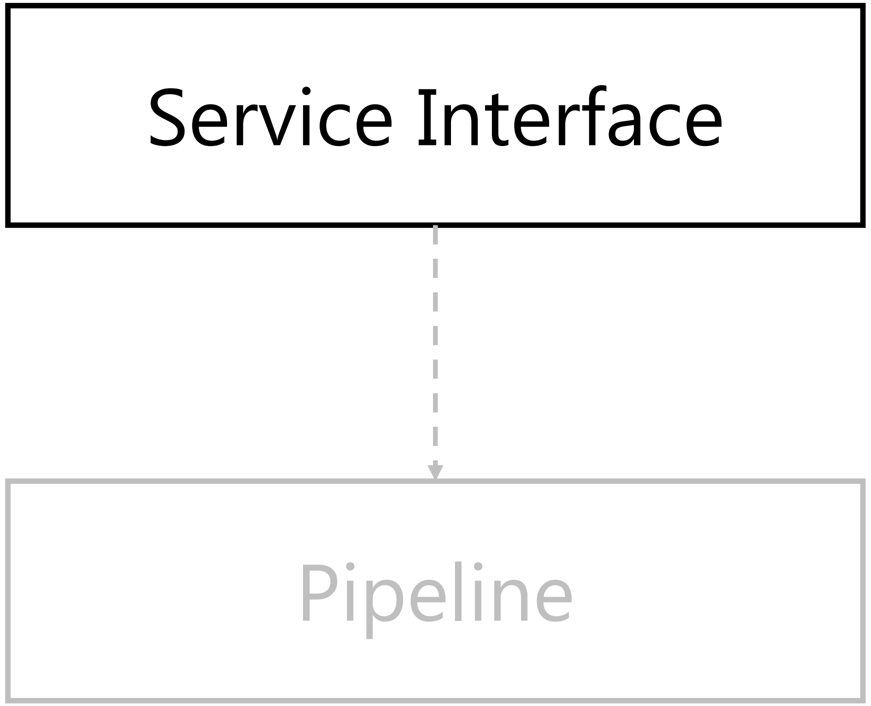
Business Logic

Data Access

Data Store

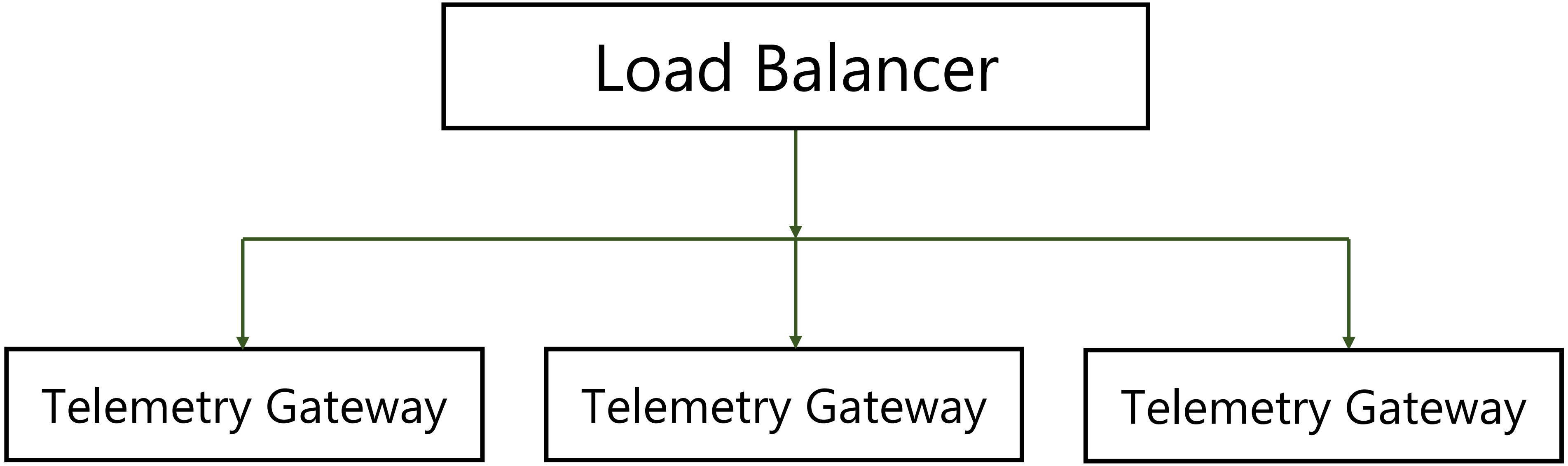
Architecture

In our case:



Muscar.auto

Telemetry Gateway Redundancy



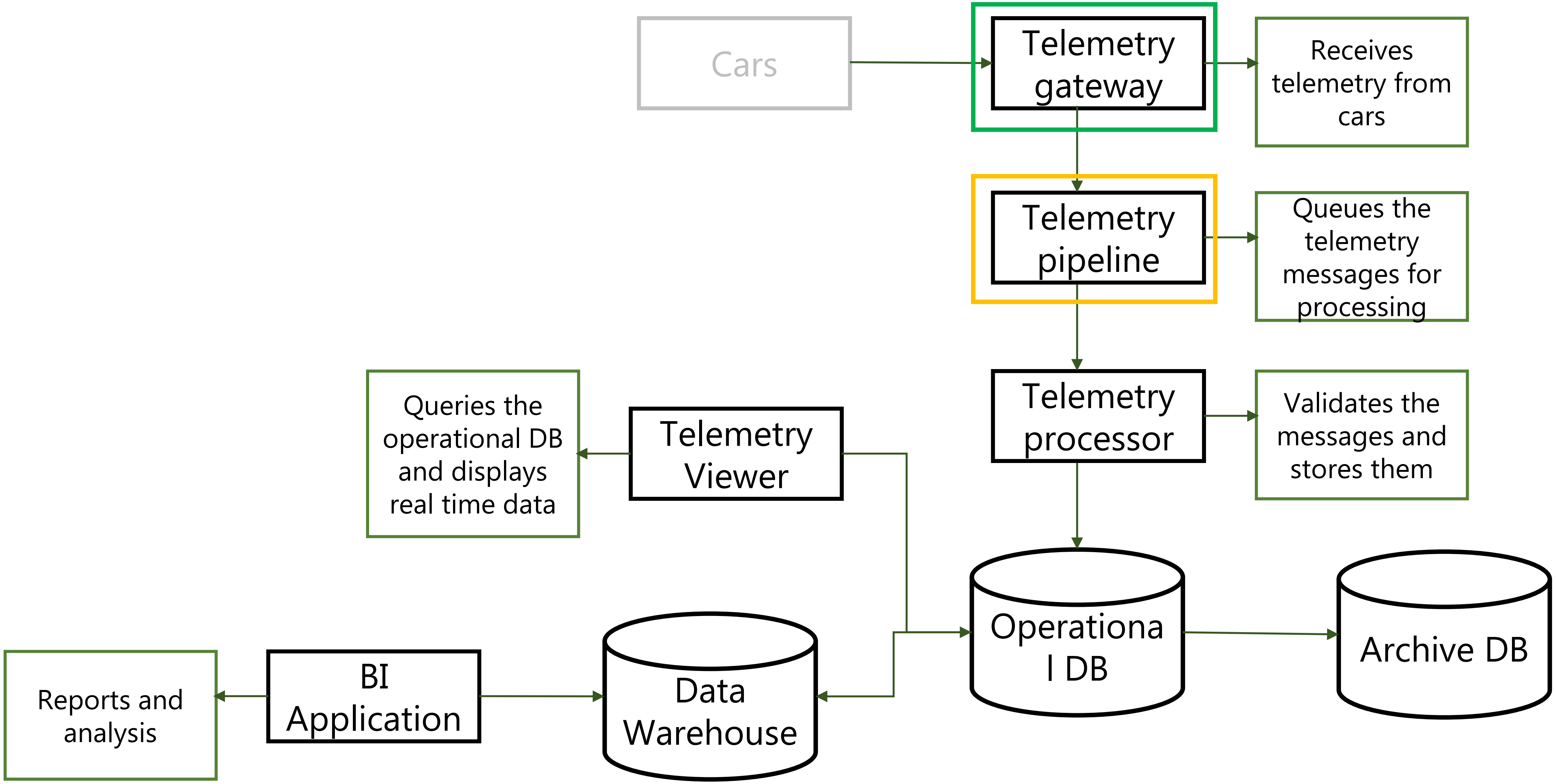
Load Balancer

Telemetry Gateway

Telemetry Gateway

Telemetry Gateway

Components



Cars
Telemetry gateway
Receives telemetry from cars
Telemetry pipeline
Queues the telemetry messages for processing
Telemetry processor
Validates the messages and stores them
Queries the operational DB and displays real time data
Telemetry Viewer
Operationa l DB
Archive DB
Reports and analysis
BI Application
Data Warehouse

Telemetry Pipeline

What it does:

- Gets the telemetry messages from the gateway
- Queues the telemetry for further processing
- Basically – a queue for streaming high volume data

Telemetry Pipeline - Questions

1. Is there an existing queue mechanism in the company?

No

2. Develop our own or use 3rd party?

Telemetry Pipeline - Questions

Let’s look around…

A distributed streaming platform

Telemetry Pipeline - Kafka

Pros:

- Very popular
- Can handle massive amount of data
- High availability support

Cons:

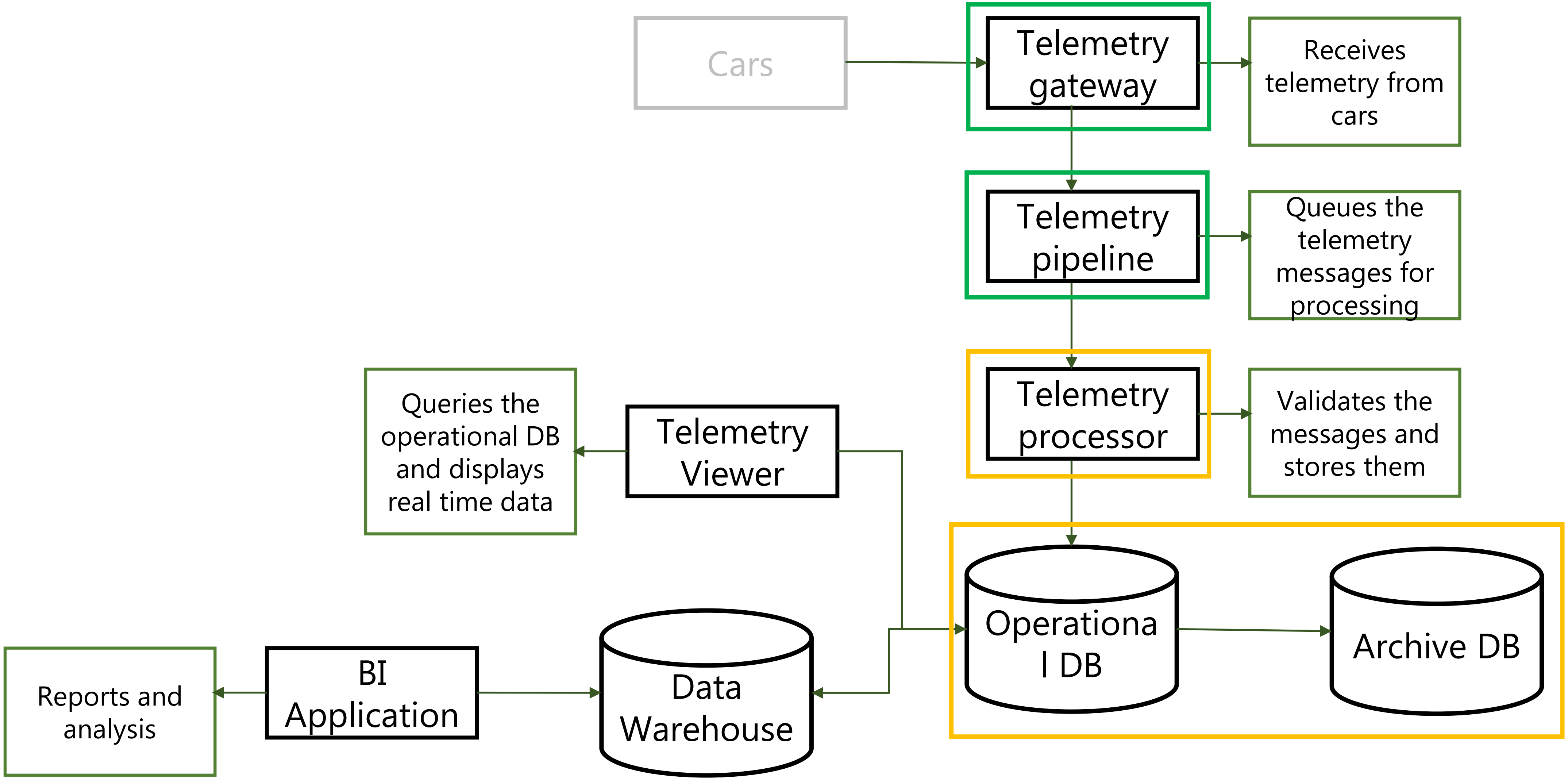
- Complex set-up
- Complex configuration

Telemetry Pipeline - Decision

APACHE
kafka®
A distributed streaming platform

Muscar.auto

Components


Cars
Telemetry gateway
Receives telemetry from cars
Telemetry pipeline
Queues the telemetry messages for processing
Telemetry processor
Validates the messages and stores them
Telemetry Viewer
Queries the operational DB and displays real time data
Operationa l DB
Archive DB
Data Warehouse
BI Application
Reports and analysis

Telemetry Processor

What it does:

- Receives the messages from the pipeline
- Processes the messages (mainly validation)
- Stores the messages in a data store

Application Type

- Web App & Web API ✗
- Mobile App ✗
- Console ✓
- Service ✓
- Desktop App ✗

Technology Stack

For:

- The processor
- The datastore

Technology Stack

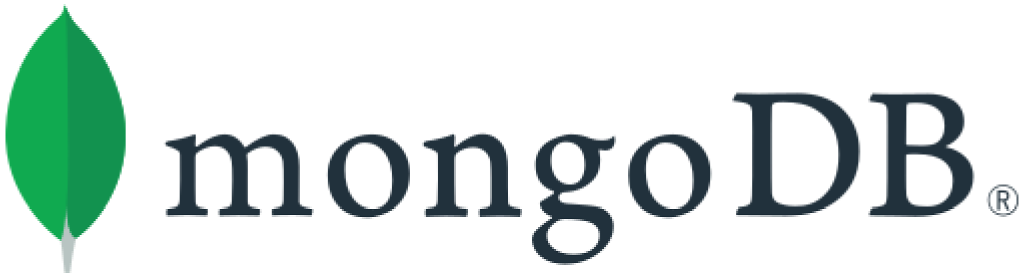
The Processor:

- Already used in the system
- Fast
- Great Kafka support

Technology Stack

The Datastore – what we’re looking for:

- Schema-less message support
- Quick retrieval
- No complex queries



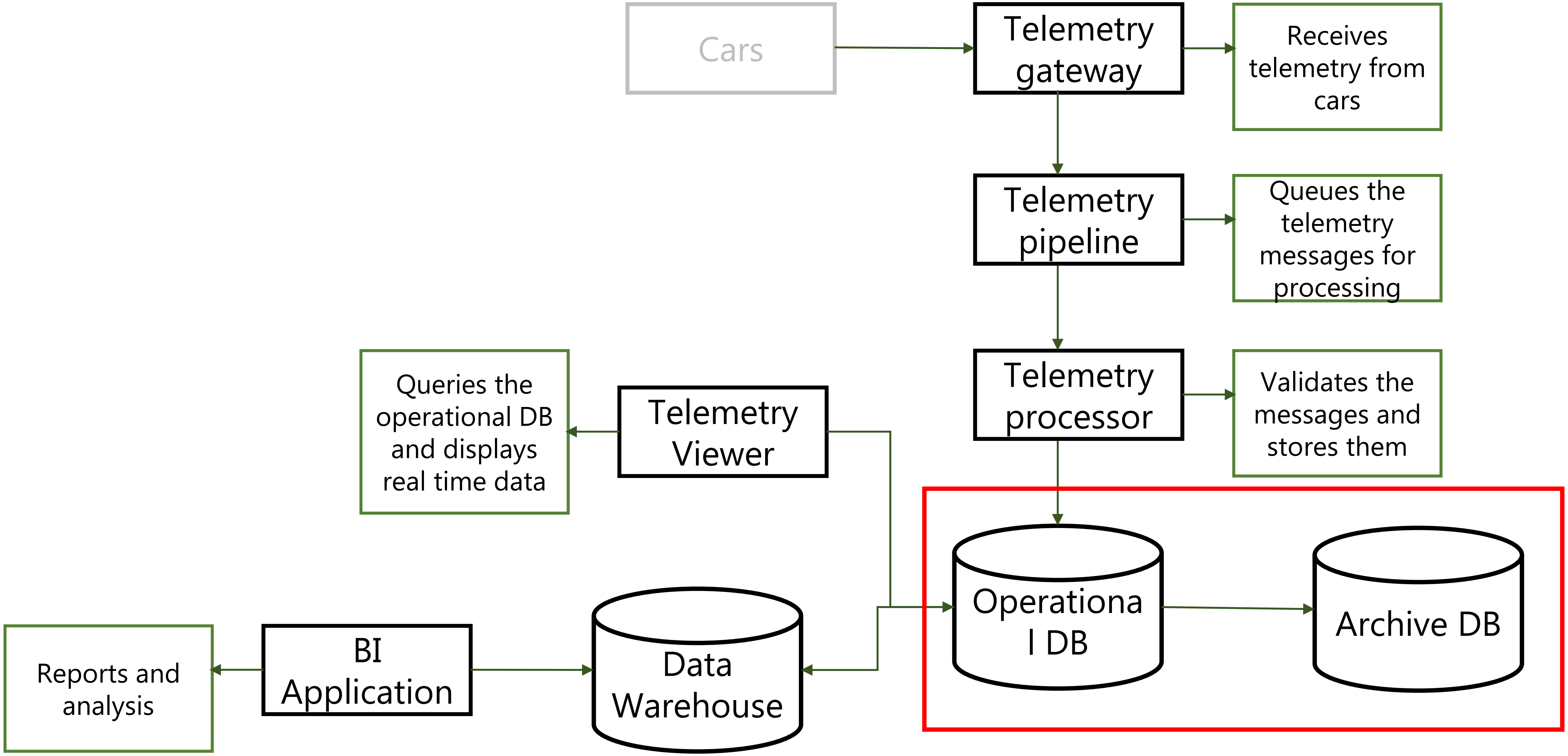
Technology Stack

mongoDB®

- Schema-less message support ✓
- Quick retrieval ✓
- No complex queries ✓

Muscar.auto

Components


Cars
Telemetry gateway
Receives telemetry from cars
Telemetry pipeline
Queues the telemetry messages for processing
Telemetry processor
Validates the messages and stores them
Queries the operational DB and displays real time data
Telemetry Viewer
Operationa l DB
Archive DB
Reports and analysis
BI Application
Data Warehouse

Muscar.auto

Technology Stack

Archive– what we’re looking for:

- Support for a huge amount of data (221TB / Year)
- Not accessed frequently
- No need for fast retrieval
- Save costs

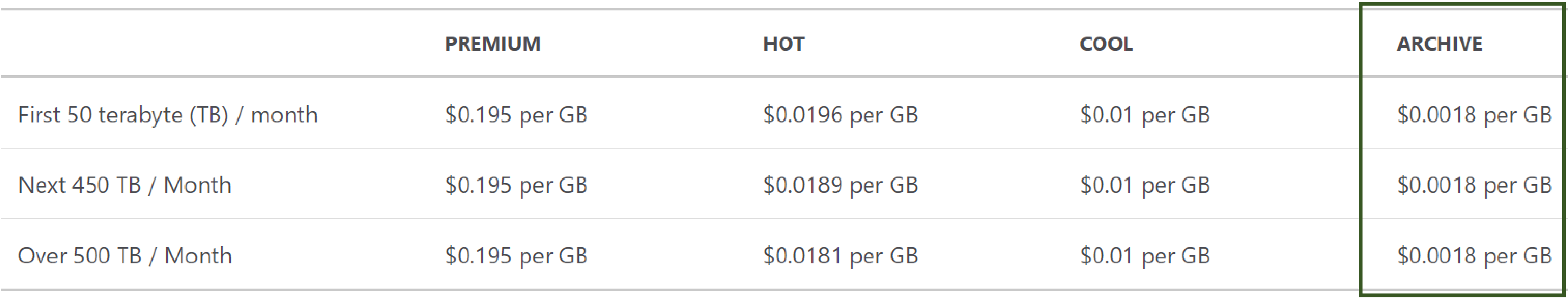
Going To The Cloud

Cloud Storage:

- Huge amounts of data (221TB / Year) ✓
- Not accessed frequently ✓
- No need for fast retrieval ✓
- Save costs ✓

Cloud Storage

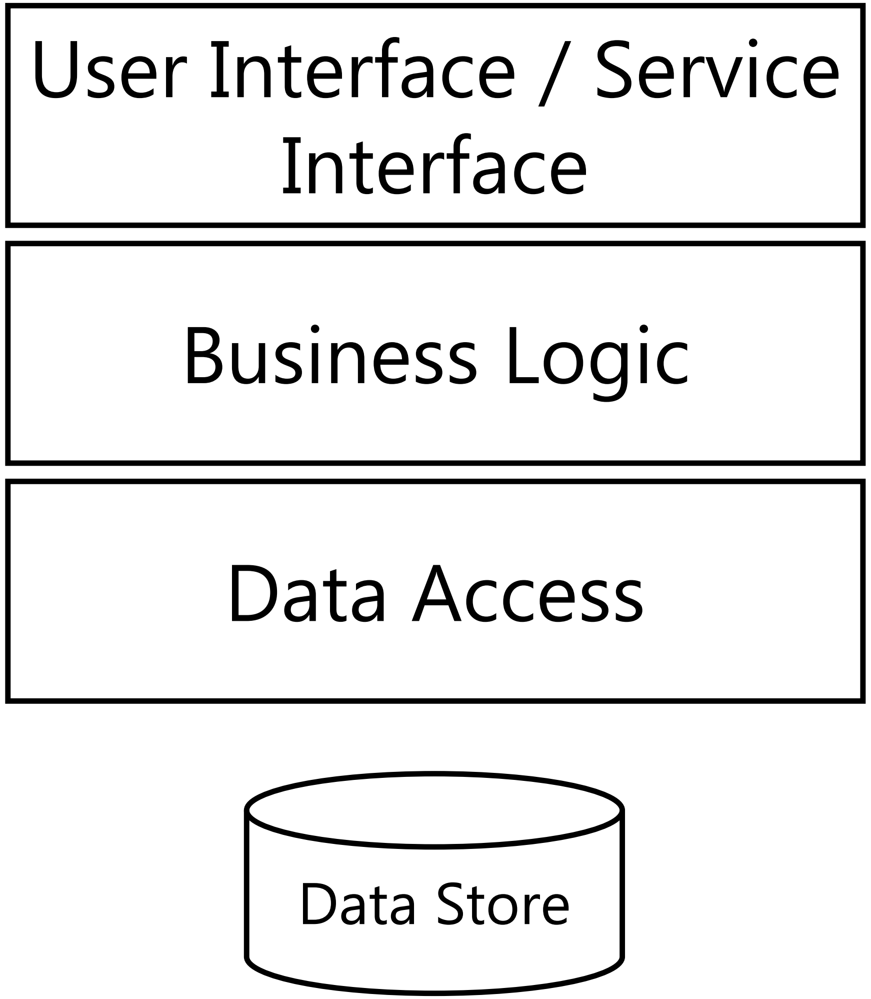
Microsoft Azure



	PREMIUM	HOT	COOL	ARCHIVE
First 50 terabyte (TB) / month	$0.195 per GB	$0.0196 per GB	$0.01 per GB	$0.0018 per GB
Next 450 TB / Month	$0.195 per GB	$0.0189 per GB	$0.01 per GB	$0.0018 per GB
Over 500 TB / Month	$0.195 per GB	$0.0181 per GB	$0.01 per GB	$0.0018 per GB

After 1 year: 221TB =>
398$

Architecture



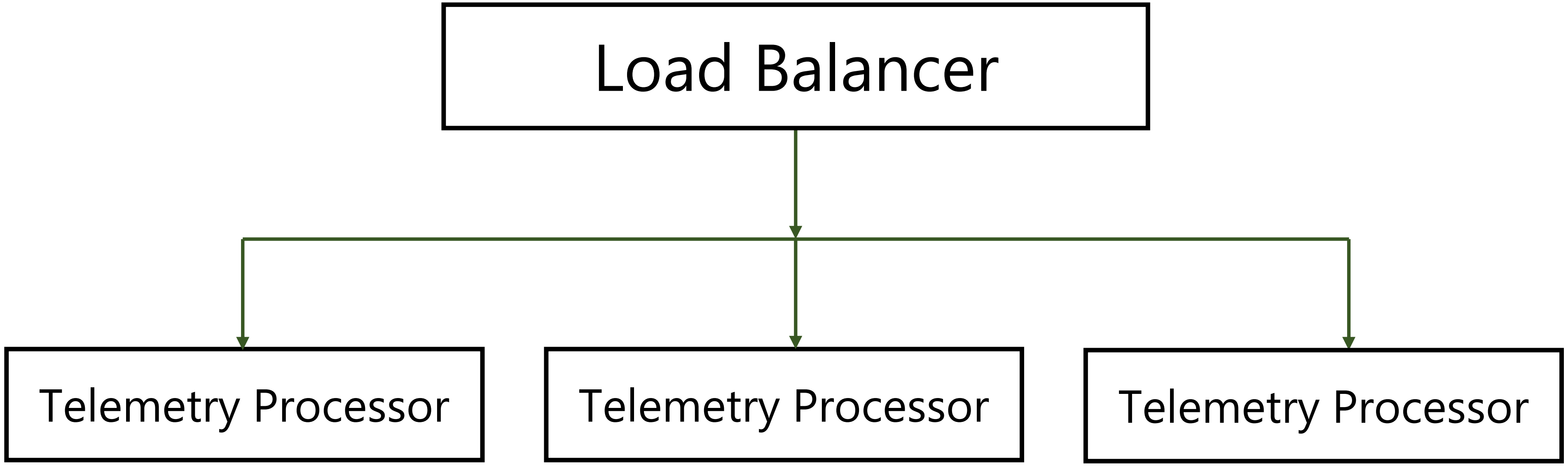
User Interface / Service Interface

Business Logic

Data Access

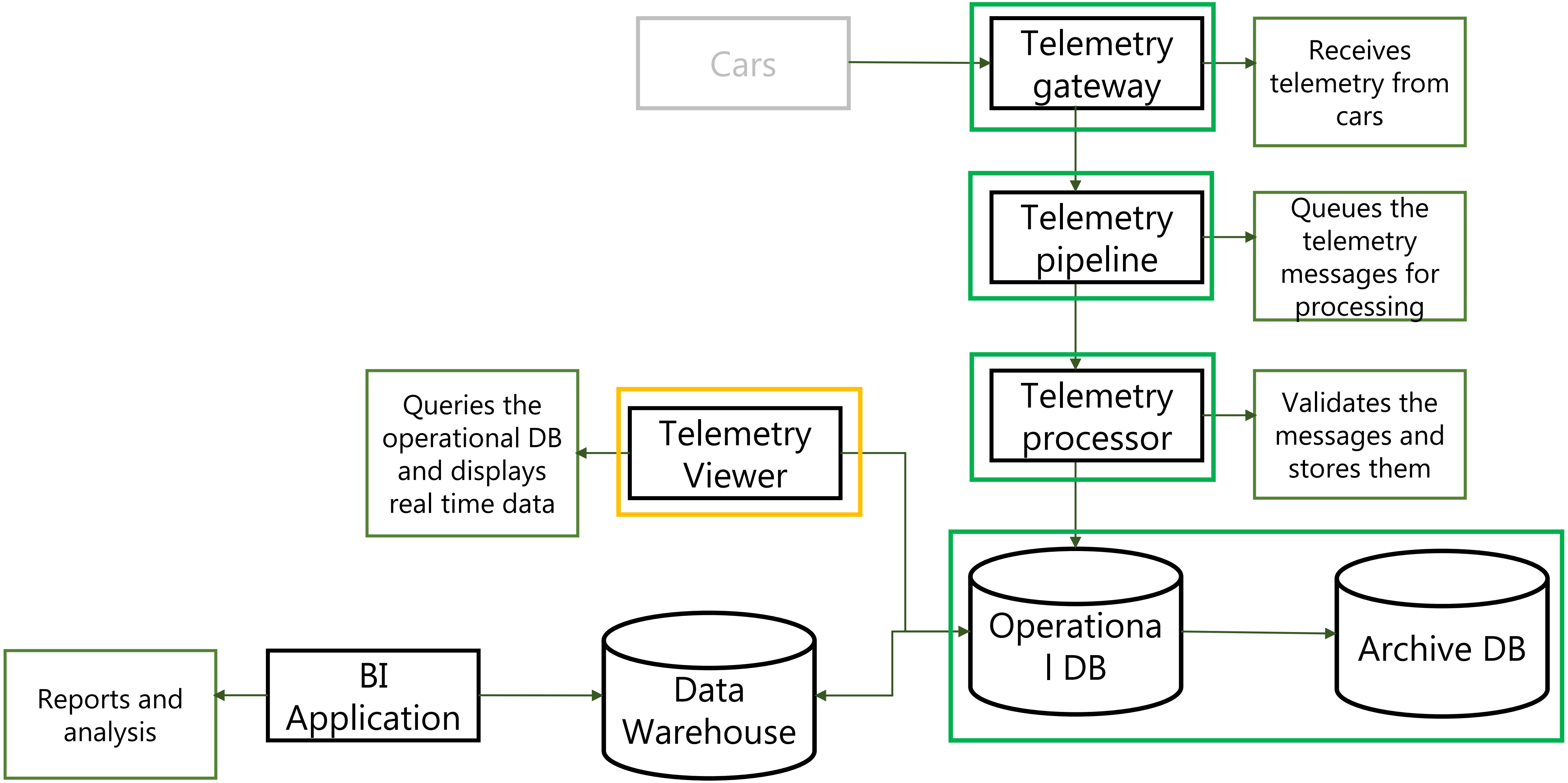
Data Store

Telemetry Processor Redundancy



Muscar.auto

Components


Cars
Telemetry gateway
Receives telemetry from cars
Telemetry pipeline
Queues the telemetry messages for processing
Telemetry processor
Validates the messages and stores them
Queries the operational DB and displays real time data
Telemetry Viewer
Operationa l DB
Archive DB
Reports and analysis
BI Application
Data Warehouse

Muscar.auto

Telemetry Viewer

What it does:

- Allows end users to query telemetry data

- Displays real time data

What it doesn’t:

- Analyzes the data

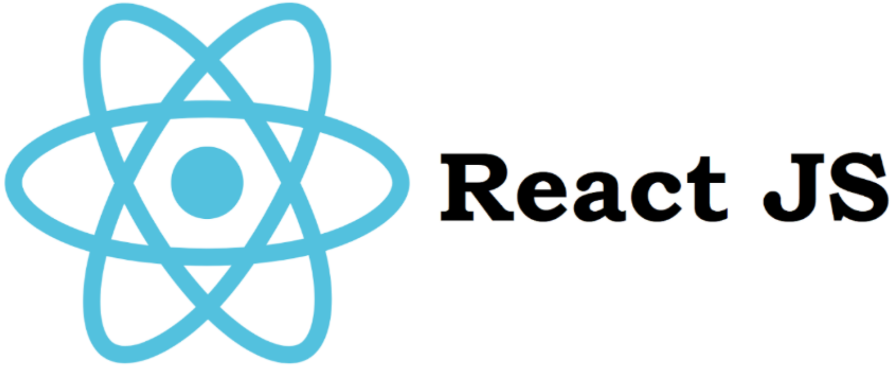
Application Type

- Web App & Web API ✓
- Mobile App ✗
- Console ✗
- Service ✗
- Desktop App ✗

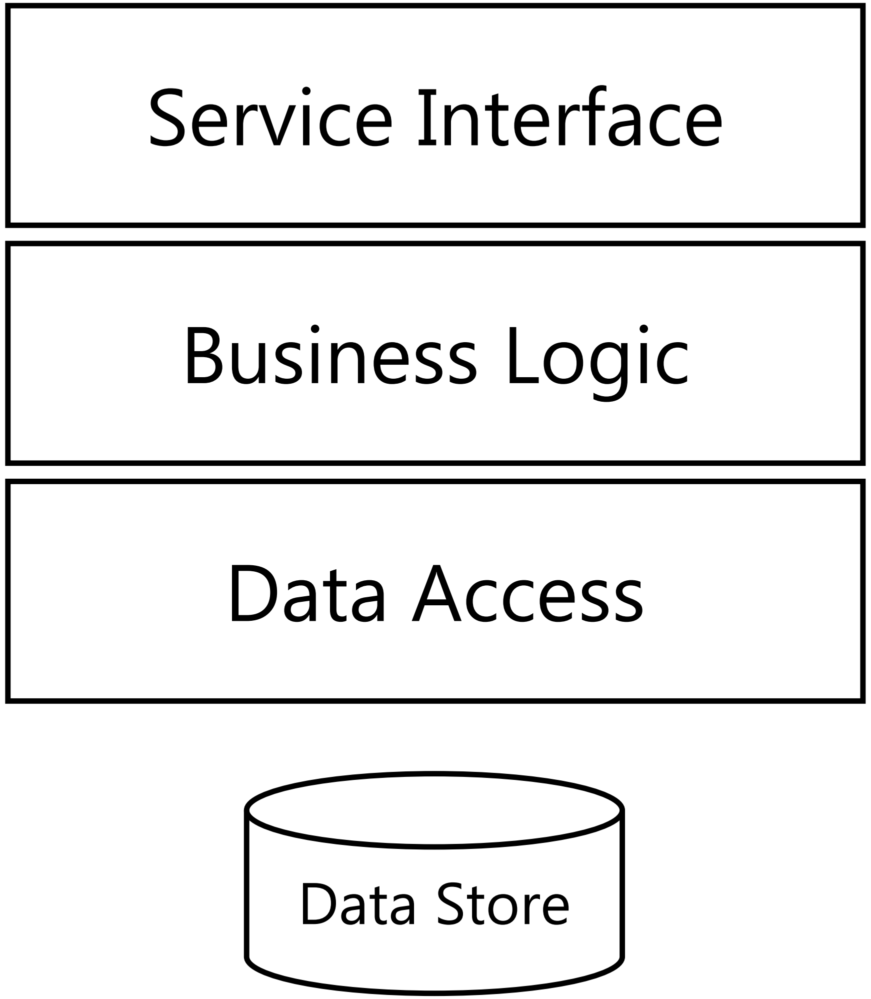
Technology Stack

Back End

Front End



Architecture



Service Interface

Business Logic

Data Access

Data Store

API

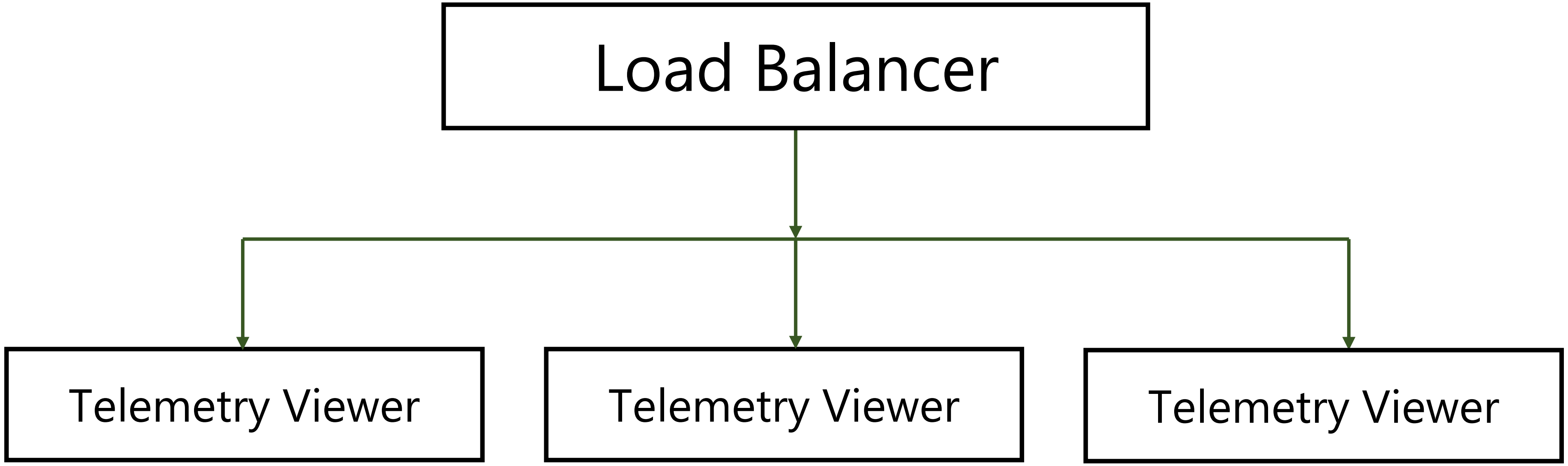
- Get latest errors for all cars
- Get latest telemetry for specific car
- Get latest errors for specific car

API

Functionality	Path	Return Codes
Get latest errors for all cars	GET /api/v1/telemetry/errors	200 OK
Get latest telemetry for specific car	GET /api/v1/telemetry/*{carId}*	200 OK 404 Not Found
Get latest errors for specific car	GET /api/v1/telemetry/errors/*{carId}*	200 Ok 404 Not Found

Muscar.auto

Telemetry Viewer Redundancy


Load Balancer

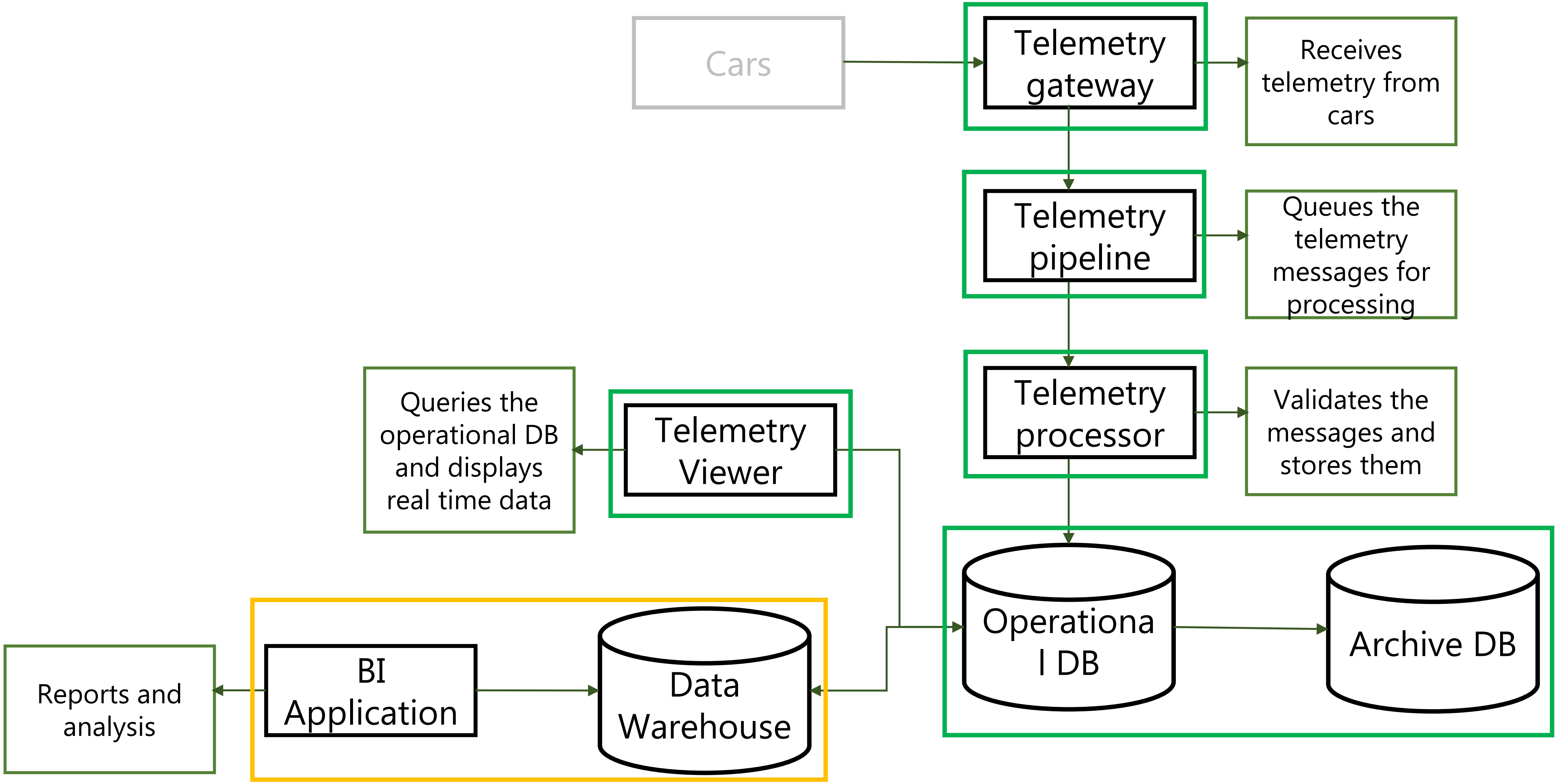
Telemetry Viewer

Telemetry Viewer

Telemetry Viewer

Muscar.auto

Components


Cars
Telemetry gateway
Receives telemetry from cars
Telemetry pipeline
Queues the telemetry messages for processing
Telemetry processor
Validates the messages and stores them
Queries the operational DB and displays real time data
Telemetry Viewer
Operationa l DB
Archive DB
Reports and analysis
BI Application
Data Warehouse

BI Application

Business Intelligence

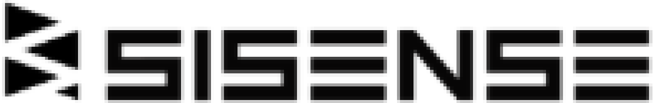
What it does:

- Analyzes telemetry data
- Displays custom reports about the data, trends, forecasts etc.
  - How many cars did break during the last month?
  - What is the total distance the cars drove?

Application Type

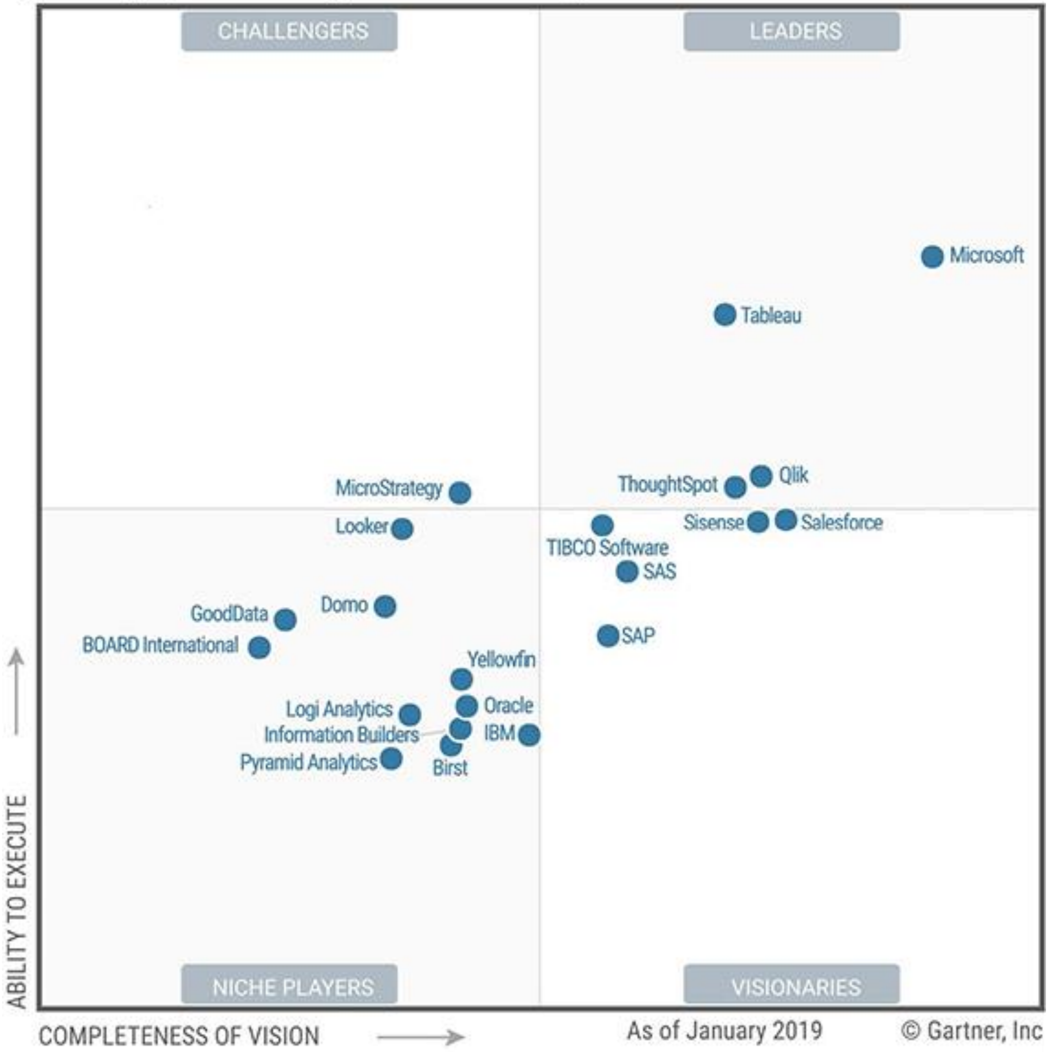
- Doesn’t matter
- BI Application is ALWAYS based on an existing tool

BI Tools



BI Tools

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



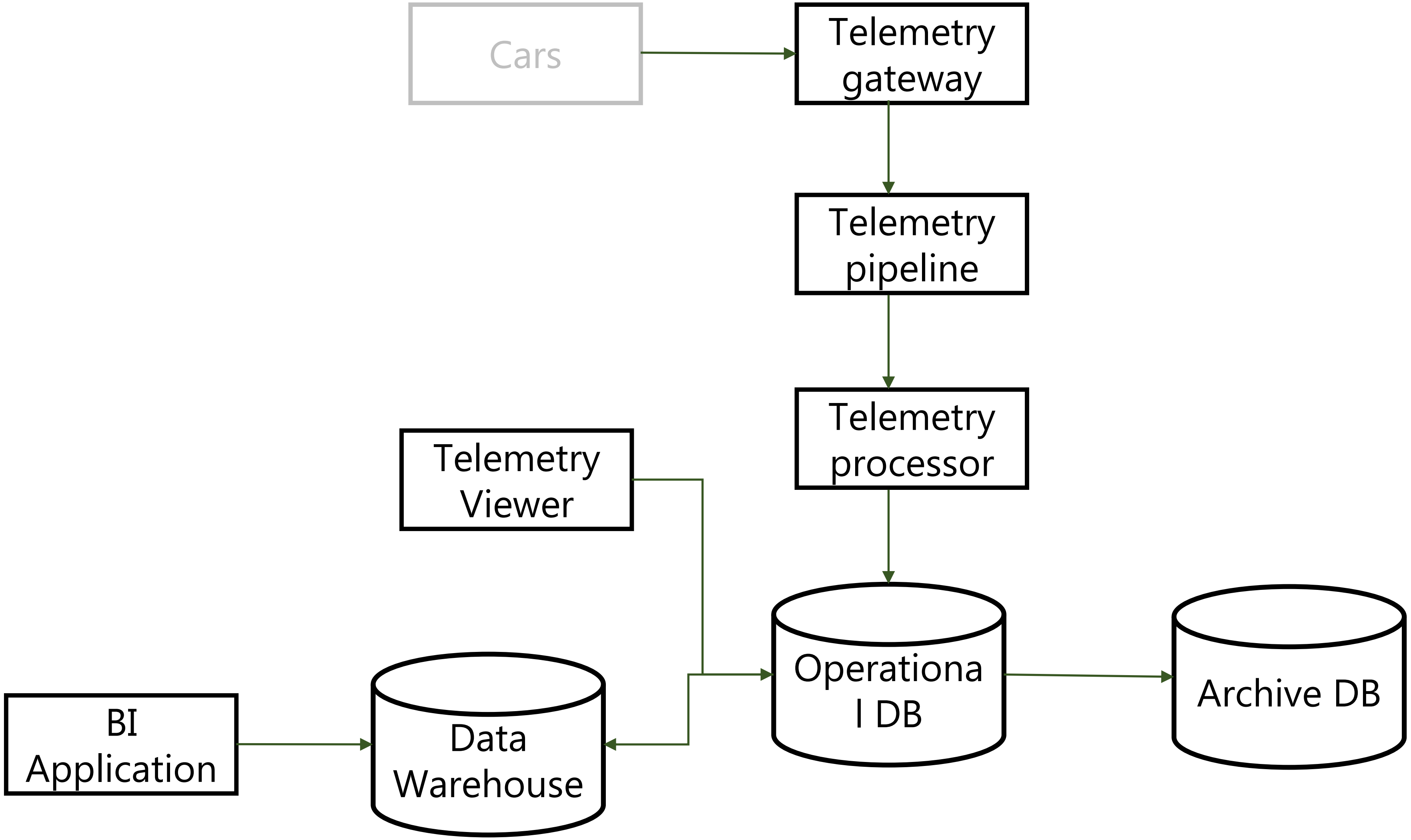
Source: Gartner (February 2019)

BI Tools

- An important lesson:
  - Designing BI solution is NOT part of the architect’s job
  - ALWAYS use BI expert for this task

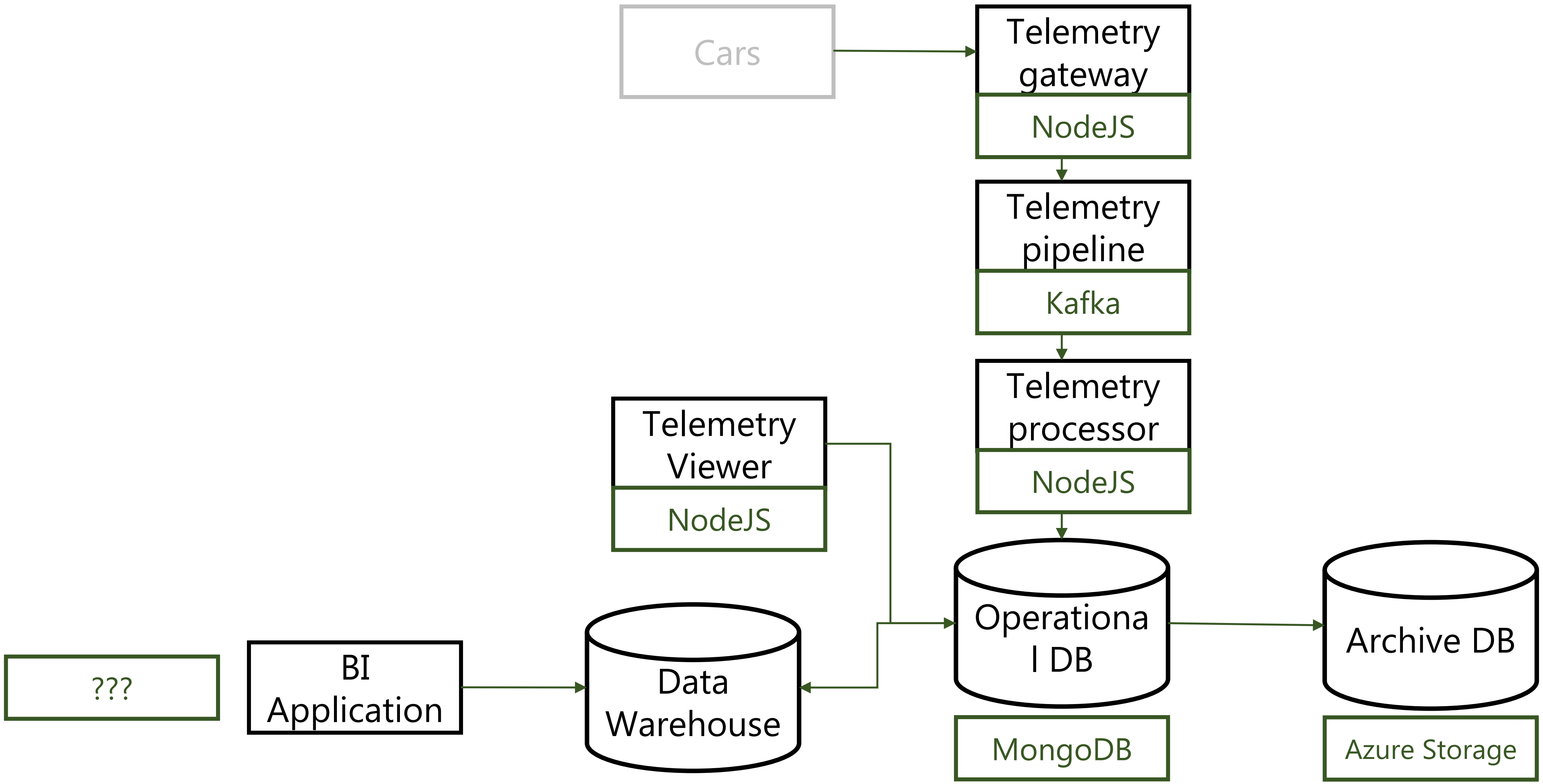
Muscar.auto

Logic Diagram



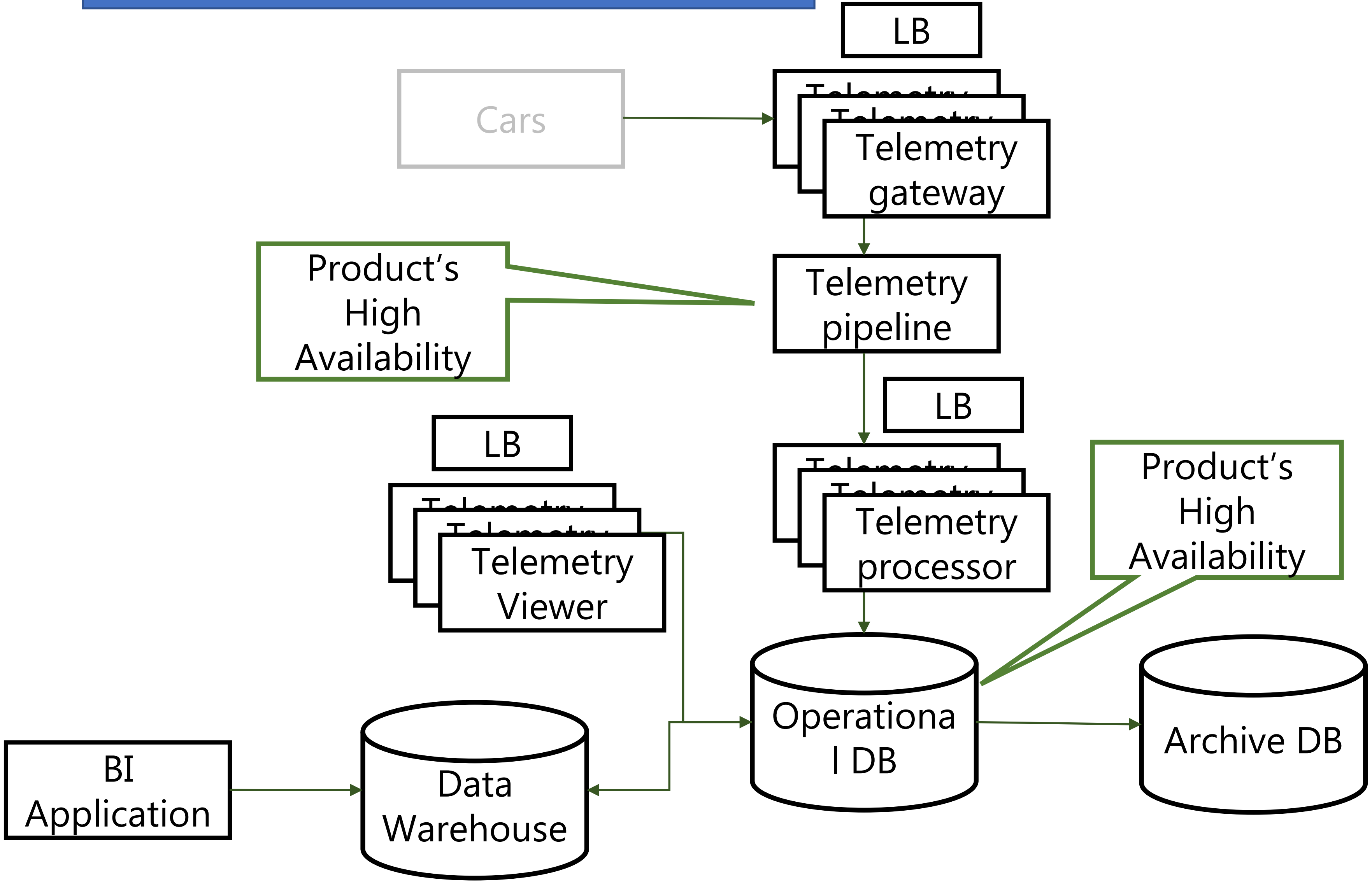
Muscar.auto

Technical Diagram



Muscar.auto

Physical Diagram


LB
Cars
Telemetry
Telemetry
Telemetry gateway
Product's High Availability
Telemetry pipeline
LB
LB
Telemetry
Telemetry
Telemetry Viewer
Telemetry
Telemetry
Telemetry processor
Product's High Availability
Operationa l DB
Archive DB
BI Application
Data Warehouse