Data Storage and Database Requirements

Andrew Cope

Project Overview

Energy Management System
Outlet Modules read power, voltage, current consumption from connected device
Data is sent to main unit
Main unit stores data and presents it through web interface

Data Storage

Data needs to be stored in order to present it to the user.

Data Storage Options:
- One single file
- Database
- Memory

Why a Database?

- Queries!
- Large data sets handled easily
- ACID (Atomicity, Consistency, Isolation, Durability) Compliance (usually)
- Easier to manage data (DBMS)

Risk Specification

Marketing Requirements

1. The system shall accurately monitor power consumption.
2. The system shall allow for control of individual outlets.
3. The system shall provide intuitive visual representations of data

Risk Specification

Marketing Requirements	Engineering Requirements	Justification
1	The voltage, power, current, and scale should be stored for each outlet reading.	Storing current and voltage, along with power, will allow for accurate readings and possible corrections.
1,2,3	The outlet names, addresses, and readings will be stored in a database.	A database allows for ease of access to all of the outlet readings. The database will be accessed through the web interface.
2	A table in the database will store the name and address of each outlet module.	The user will be able to determine more easily what outlet modules are what, and the address allows for communication between the outlet modules.
3.	After 30 days, the readings are condensed/averaged into larger time segments (minutes or hours) for that day.	The size of the main unit's storage is limited, and the user will not need to see the precise outlet readings from 30 days ago. Compressing the data into broader readings will maintain a visual representation, but take up smaller space on the main unit.
3	The outlet readings will store the current time at insertion to the database.	The time data will allow the user to view specific time ranges for an outlet module, if need be.

Risk Investigation - DBMS

DBMS = Database Management System
- Creating, Updating, Querying, Administrating
- Popular DBMSs:
  - Microsoft SQL Server
  - MySQL
  - PostgreSQL
  - SQLite
- ACID = Atomicity, Consistency, Isolation, Durability

Microsoft SQL Server

- Developed by Microsoft
- Available on Windows Server OS only
- Commercial license
- ACID Compliant
- Supported in smaller number of programming languages
- Multiple locking techniques, based on transaction taking place

MySQL

Developed by Oracle
Available on all operating systems
GPL open-source commercial license
ACID Compliant (Storage Engine dependent)
Supported by many programming languages
Table or Row-level locking (Storage Engine)

PostgreSQL

Developed by PostgreSQL
Available on all operating systems
BSD open-source license
ACID Compliant
Fewer languages supported than MySQL, but still a number of major languages
Supports both locking mechanisms

SQLite

Developed by Richard Hipp (originally)
Available on all operating systems
Open source through public domain license
ACID Compliant
Available in MANY languages
Supports variation of table-level locking

Criteria	Solutions			
	Microsoft SQL Server	MySQL	PostgreSQL	SQLite
Server OS	**-1**	+1	+1	+1
License	**-1**	+1	+1	+1
Java	+1	+1	+1	+1
Perl	0	+1	+1	+1
PHP	+1	+1	+1	+1
Python	+1	+1	+1	+1
ACID Compliant	+1	+1	+1	+1
Concurrency	+1	+1	+1	+1
Data Types	+1	+1	+1	**-1**
Total Rating	4	9	9	7
Rank	4	1	1	3
Consider?	No	Yes	Yes	Yes

MySQL - Storage Engine

Memory	MyISAM	InnoDB

• Data Stored in memory
• Database flushed to disk once enough changes have been made
• Quicker access times
	• NOT ACID Compliant
	• Table-level Locking
	  ○ Read-Heavy
	• Three separate files
		• ACID Compliant
		• Row-level Locking
		  ○ Write Heavy
		• Typically 1-2 files

MySQL Storage Engine Analysis

Criteria	Memory	MyISAM	InnoDB
Reliability	-1	+1	+1
Locking	+1	-1	+1
Performance	+1	0	+1
Transactions	+1	-1	+1
Storage Medium	-1	+1	+1
Total Rating	1	0	5
Rank	2	3	1
Consider?	No	No	Yes

Risk Mitigation - Database Tables

Outlet Module Table

Column	Data Type	Max Data Size
Outlet ID	Tiny/Small Integer	1 - 2 Bytes
Name	Varchar (100)	101 Bytes
Address	Binary (64)	8 Bytes
Frequency	Tiny Integer	1 Byte
Real_Time	Bit (1)	1 Byte
	Max Size Per Row:	112 - 113 Bytes

Risk Mitigation - Database Tables

Column	Data Type	Max Data Size
Reading ID	Integer	4 Bytes
Outlet Module ID	Tiny / Small Integer	1 - 2 Bytes
Scale	Tiny Integer	1 Byte
Voltage	Tiny Integer	1 Byte
Current	Small Integer	2 Bytes
Power Factor	Small Integer	2 Bytes
Time	Timestamp	4 Bytes
Real Time	Bit (1)	1 Byte
	Max Size Per Row:	16 - 17 Bytes

Database Size - Outlet Module Table

# of Outlet Modules	Size Per Table Row	Total Estimated Size
1	112 B	112 B
10	112 B	1.09 KB
50	112 B	5.47 KB
100	112 B	10.94 KB
300	113 B	33.11 KB
1,000	113 B	110.35 KB
10,000	113 B	1.08 MB
65,535	113 B	7.06 MB

Database Size - Outlet Reading Table

# of Outlet Modules	Update Interval	Table Size Per Day	Table Size Per Year
10	1 second	13.18 MB	4.7 GB
50	1 second	65.92 MB	23.5 GB
100	1 second	131.84 MB	47 GB
500	1 second	700.38 MB	250 GB
10	15 seconds	0.88 MB	0.3 GB
50	15 seconds	4.39 MB	1.6 GB
100	15 seconds	8.79 MB	3.13 GB
500	15 seconds	46.69 MB	16.64 GB

Risk Mitigation - Data Compression

Every day, take all readings from 30-31 Days ago and merge them into larger/broader readings.

Saves space!

Data Compression - Outlet Reading Table

# Outlet Modules	Original Data Size		4 -> 1 Compression (1 Minute)		240 -> 1 Compression (1 Hour)	
	Size/Day	Size/Year	Size/Day	Size/Year	Size/Day	Size/Year
1	96 KB	34 MB	30 KB	8 MB	384 B	137 KB
10	1 MB	300 MB	230 KB	80 MB	3.75 KB	1.34 MB
50	4 MB	1.56 GB	1.1 MB	401 MB	18.75 KB	6.7 MB
100	9 MB	3.13 GB	2.2 MB	802 MB	37.5 KB	13.4 MB
500	47 MB	16.64 GB	11.7 MB	4.16 GB	199 KB	71 MB

Testing

1. Outlet Module Reading Simulation
2. Real-Time Outlet Module Reading
3. Time Range Outlet Module Reading
4. Data Compression
5. Deletion of Real Time Data After Session Closes

Questions?