SensorCloud
Sensor Data Management
SensorCloud Platform

- Sensor Agnostic
- Time series scalar data
- Cloud hosted
- Scalable
- Developer API
- Meta-data model for sensor networks
- Federating
- Real-time ingestion and processing
- Web streaming
SensorCloud Information Model

- Meta-data distinct from observations
- StarFL OGC discussion document. Developed in Hobart by CSIRO
- Observations model based upon OGC Observations and Measurements
Sensor Platform

- An instance of a specific asset/structure which sensors are mounted. (eg. A specific vehicle, a weather station tripod, an individual person)
Sensor

• A specific sensor. (Eg. Greenspan MP65 with serial number 12345678)
Deployments

• Platform Deployment
  • A relationship that ties a Platform to a location within a time period
  • Eg. *Platform A was deployed at Frogmore creek between 1/2/2015 and 4/4/2015*

• Sensor Deployment
  • A relationship that ties a Sensor to a Platform
  • Eg. *Sensor MP47 with serial no 12345 was install on platform C between 1/1/2015 and 1/3/2015 at a depth of 5m*
Locations

• GeoLocation
  • Static or Dynamic
  • Dynamic locations are linked to a GeoLocation data stream

• RelativeLocation
  • Static; \((x, y, z)\) metres (eg. The anemometer is mounted 10m above the Platform)
Sensor Characteristic

• Define capabilities of a sensor model.

• Datasheet

• Eg. The Greenspan MP65 can measure electrical conductivity with +/- 1% accuracy
Calibration

- Conformance test or field sensor calibration
- Calibration event recorded
Stream

• A collection of observations
• Called ‘phenomenon’ in V1
• Optionally associated with Sensor in V2
Observations

• The actual data (>99% of volume)
• Version 2 types:
  • GeoLocation (eg. GPS sensors)
  • Scalar value (numerical value, eg. Temperature)
  • Vector value (array of values, eg. Spectrometer, depth profile)
  • Sequence value (high resolution, eg. Audio)
  • Image value
SensorCloud Data Storage

• Why MongoDB?
  • Document storage, flexibility in data types
  • GridFS, distributed storage for files (images)
  • Aggregation framework (distributed computation of aggregations)
  • GeoSpatial indexes and queries (for mobile sensors)
  • Easy to horizontally scale
  • Experience within development team
MongoDB Structure

• V1 learning (1.1 billion observations)
  • Scalar data is small in volume. Index size large vs data.
  • Nodes use a lot of RAM, performance degrades as data/ram ratio increases

• V2
  • Aggregated storage, multiple observations in a single ‘document’
  • Indexes smaller for high frequency (faster than hourly) data streams
  • MongoDB aggregation pipeline unwinds results
  • Improved sharding to balance data streams across nodes in cluster.
Queries

• Most common query is to get temporal range of data stream. Eg. Get all data between 1st Jan and 3rd March for the data stream called ‘mycooldatastream’.

• Temporal Aggregation
  • Summary data gets too large (esp. On mobile devices)
  • Aggregation performed by distributed MongoDB cluster.

• Geospatial:
  • “get data collected in the paddock with boundary x”
  • And in temporal period y
  • Also get all sensor attached to the same platform?

• Various encodings
  • CSV, JSON, GeoJSON

• Multivariate queries
  • Get all data from sensor a,b,c with time range x

• Examples
Data Ingestion Framework

• **Sensor Messaging Gateway (SMG)**

```
Sensor Message Gateway (SMG-1)

Source(s)  | Processor(s)  | Sink  | JMS
          |              | JMSSink |

SMG-2

ActiveMQ Cluster

SMG-2

Sink SMG (JMS Source and Mongo Sink)

MongoDB Sharded/Replicated Cluster

Real-time Client (e.g., WebSockets)

FTP Client, FTP Server, HTTP Client, HTTP Server, SOAP, MQTT, TCP, etc. CSV, JSON, XML,
```
Data Sources

• Configurable Generic polled file import (CSV, TSV, fixed width, FTP, HTTP)
• Existing library of data sources
  • Campbell Scientific
  • Libelium
  • PACP (DPF/CSIRO)
  • AgIsp (DPF/CSIRO)
  • ROS (Robotic platforms)
  • BoM/SILO
• Custom Data Sources (Java, Python)
• +more
Data Processing

- Heart rate processing example
- Subscribes to Message queue for heart signal
- Perform processing
- Publish back heart rate BPM

![Diagram of data processing flow](image-url)
Web Streaming

• JavaScript streaming
• Uses STOMP interface in RabbitMQ
**Python API**

- Uses pandas
  - Python Data Analysis Library (http://pandas.pydata.org/)
- Python example

```python
api = SensorCloudV1()
# Different ways to access data streams
aphenomenon = api.get('demo','smg_oyster','oystag_347','2') # or
aphenomenon = aNetwork.get('smg_oyster','oystag_347','2') # or
aphenomenon = aPlatform.get('oystag_347','2') # or
aphenomenon = aSensor.get('2') # or
aphenomenon = aSensor.get_phenomenon('2') # or

anotherPhenomenon = aSensor.get_phenomenon('9')

start = datetime(2014,12,25)
end = datetime(2015,1,1)

print "single variable output"
d = api.get_observations(aphenomenon, start, end)
print d

print "multivariate output"
d = api.get_observations_multi([aphenomenon, anotherPhenomenon], start, end)
print d
```
Authentication and Authorization

- HTTP ‘Basic’ authentication
- All SensorCloud metadata and streams belong to one or more groups
- User ‘roles’ defined permissions for a group.
- Eg. John works on a project called Aquaculture. That project has its own private sensors, and all needs access to some private sensor data. The role ‘aqua_researcher’ allows read access to all data in the ‘public’ group and read access to data in the ‘private_aqua’ group.
Model services

- SensorCloud provides raw sensor data by reference
- Model Services provides a framework to:
  - ‘Wrap’ existing models as web services
  - Model inputs and outputs described in REST API
  - Currently support for Kepler, R, Python, Java
Gridded Data Services

• THREDDS
  • Data management for netCDF data sets
  • Provide catalogue and the following services
    – Web Map Service (WMS)
    – Web Coverage Service (WCS)
    – OpenDAP
    – NetCDFSubsetService
    – HTTP
SensorCloud Tools

- Generic visualisation and data extraction
- Sensor failure alerting (sms, email)
- Infrastructure monitoring/alerting (email)
SensorCloud Projects

• Users
  • Sense-T Research Projects
    – Viticulture (AWSs)
    – Dairy and Beef (AWSs)
    – Aquaculture (Water Quality)
    – Water management (AWSs, Streamflow)
  • Aquaculture (Oysters and AgIsp)
  • Soil Carbon Monitoring
  • TERN / ACEF (Water Quality)
  • Bees (AWSs)
Applications using SensorCloud

- TSQAP Decision Support Application
- Viticulture decision support web application
- Water Management dashboard
- Pasture Growth dashboard
- Water Quality Dashboard
Thanks