

[소프트웨어정책연구소 세미나]

IoT 사물데이터 활용 기술 동향

2018. 9. 17

이기용

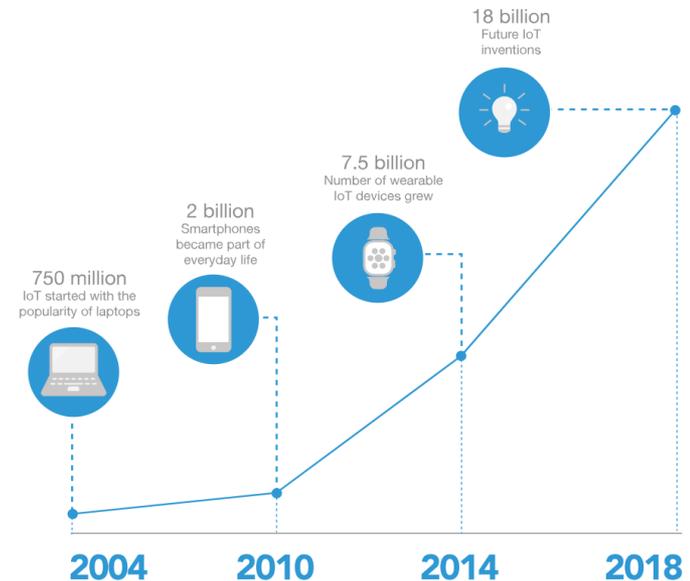
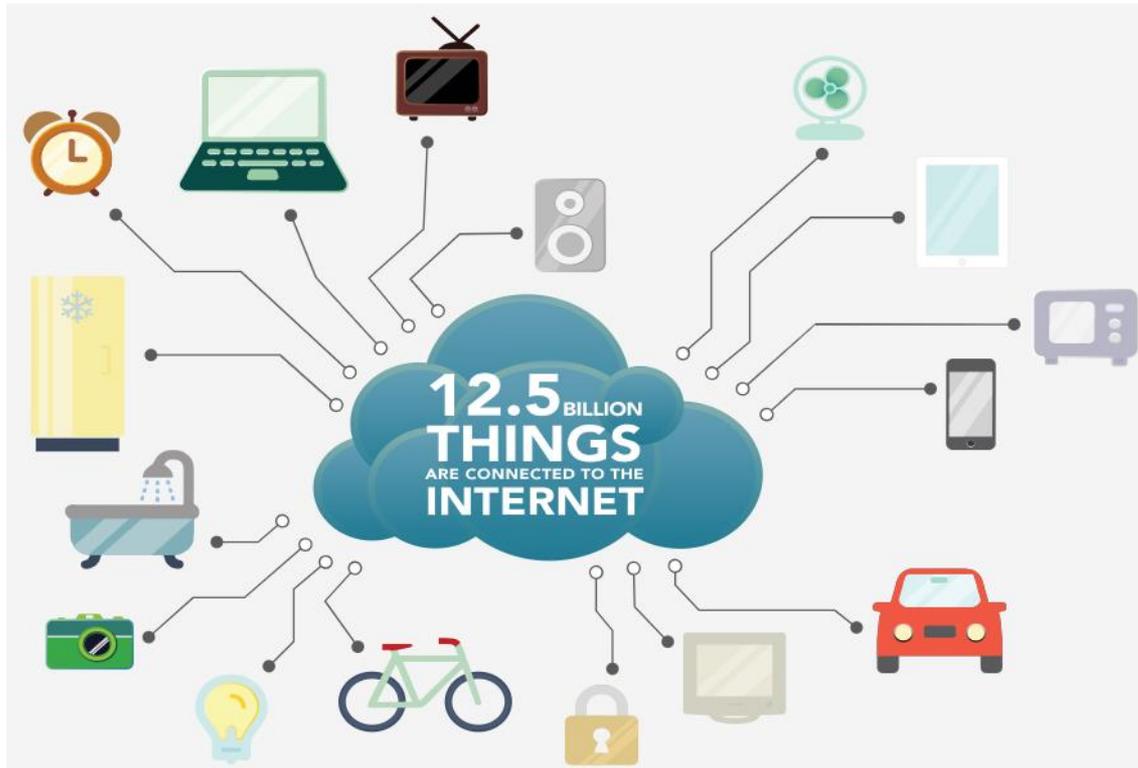
Data & Knowledge Engineering Lab.
Division of Computer Science
Sookmyung Women's University



- 사물데이터 활용의 필요성
- 사물데이터 활용 서비스 동향
- 사물데이터 활용을 위한 검색 기술 연구 소개
- 결론

사물인터넷(IoT) 기술 확산 및 보편화

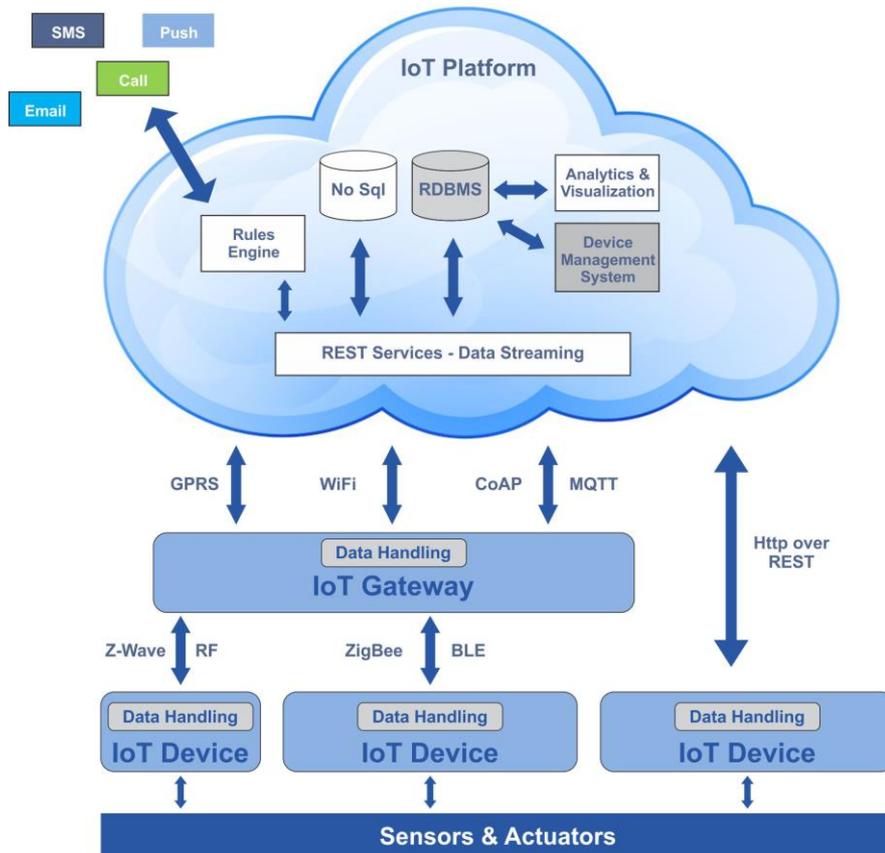
- 사물인터넷(Internet of Things, IoT) 기술 보급과 함께 점차 더 다양한 단말들이 사물인터넷에 연결되고 있음
 - 그에 따라 단말들이 생성하는 사물데이터도 급증하는 추세



전 세계 사물인터넷 단말 수 증가 추이
(www.tado.com, 2018)

IoT 핵심 기술

- 지금까지 IoT 확산에 기여한 핵심 기술은 단말과 인터넷 연결 및 단말 간 데이터 전송 기술에 집중됨



– 사물 연결 기술

- 컴퓨터가 아닌 일반 사물을 인터넷에 연결시키는 기술

– 사물 제어 기술

- 일반 사물을 데이터 전송을 통해 제어하는 기술

– 사물 간 데이터 전송 기술

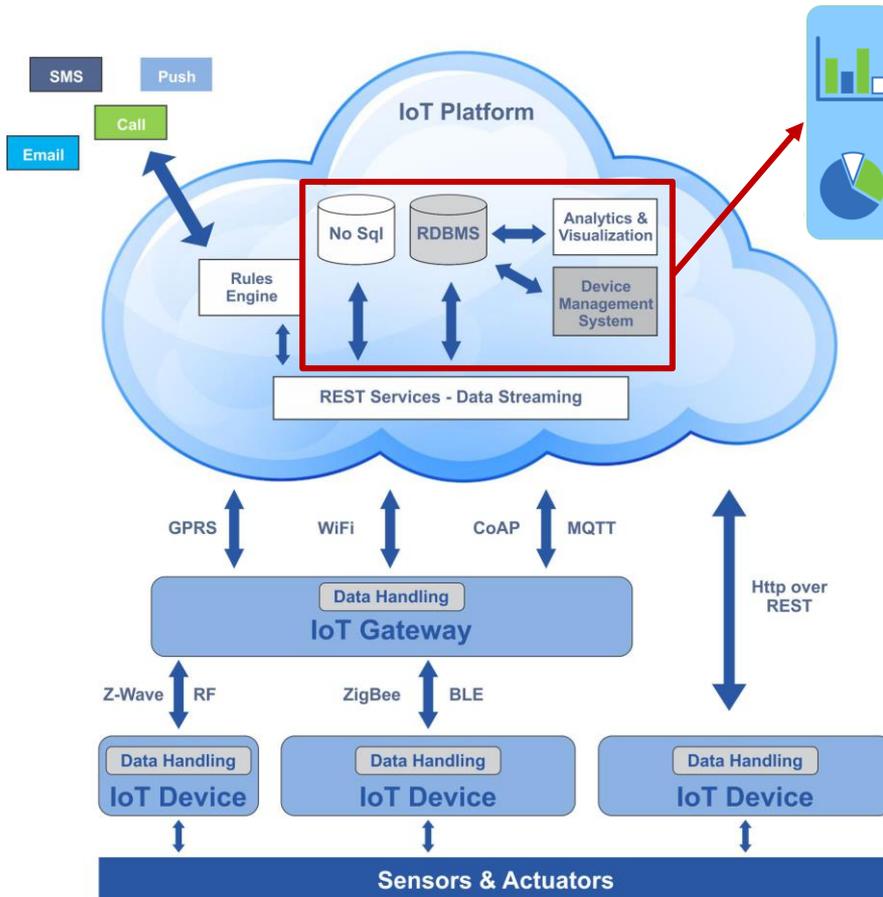
- 대규모 사물 간 서로 데이터를 보내고 받는 기술

– 기술 표준화

- 여러 사물 간 상호 운용을 위한 표준 (AllJoyn, oneM2M, Thread, ...)

‘데이터 분석 및 활용’ 기반 IoT

- 하지만 단순히 사물 연결에 그치지 않고 사물들이 발생시킨 데이터를 분석하여 활용하려는 시도가 점차 늘고 있음



- **사물데이터 저장 기술**
 - 대용량 사물데이터를 확장성 있게 저장하는 기술
- **사물데이터 검색 기술**
 - 대규모 이종 사물데이터 중 원하는 데이터를 효율적으로 검색하는 기술
- **사물데이터 분석 기술**
 - 사물 빅데이터 및 데이터 스트림에 대한 사후 또는 real-time 분석 기술
 - 사물데이터에 대한 데이터마이닝, 기계학습 기술 (예측, 이상치 탐지, 빈발 패턴 탐색 등)

‘데이터 분석 및 활용’ 기반 IoT

■ (예) 코카콜라 ‘Freestyle Soda Machine’

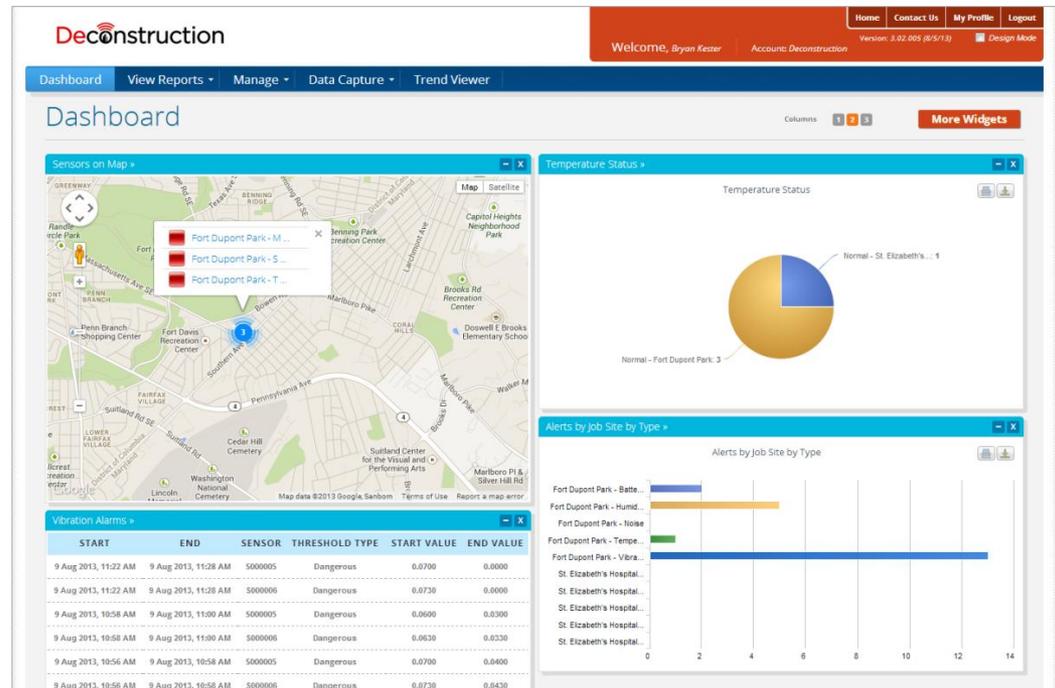
- 자판기에서 147가지 맛의 음료를 소비자 선택대로 혼합하여 제공
- 자판기마다 무선인터넷 칩으로 음료 소비량, 시간대 별 판매량 통계, 인기 음료 등의 **정보를 실시간으로 본사에 전송**
- 코카콜라 측은 이러한 데이터를 토대로 특정 지역의 인기 음료, 재고 등을 파악하는 한편, 지역적 취향을 고려한 **마케팅 전략 수립 가능**



코카콜라 프리스타일
음료 자판기 앱과
실제 앱 이용 모습
(www.coca-colafreestyle.com)

‘데이터 분석 및 활용’ 기반 IoT

- (예) 디컨스트럭션(Deconstruction) ‘mBuilder’
 - 공사 현장의 온도, 습도, 소음, 진동 등을 실시간으로 모니터링
 - 수집된 데이터를 분석하여 주변 지역의 체감 소음 및 진동 **예측**
 - 일정 수준 이상의 피해가 예측되면 이를 즉시 통지



mBuilder 센서와 분석 결과를 보여주는 대시보드 화면

‘데이터 분석 및 활용’ 기반 IoT

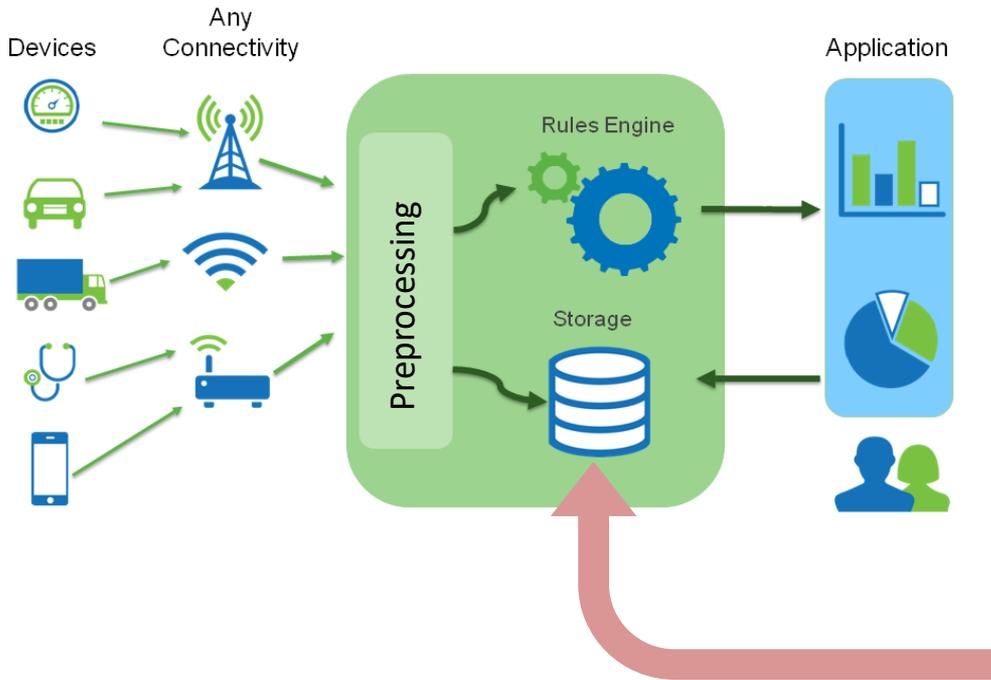
■ (예) IoT 생활가전에서 사물데이터를 활용하는 시나리오

생활가전	IoT 기반 서비스	인공지능 탑재로 진화된 가전 서비스
에어컨	<ul style="list-style-type: none"> - 인체감지 센서 연동을 통한 자동 on/off 기능 - 원격 에어컨 제어 	<ul style="list-style-type: none"> - 인체감지 센서와 공간학습 기능을 통해 실내환경을 감지하고, 사람의 위치와 수를 파악해서 냉방 공간, 냉방 모드, 공기청정 가동 등을 스스로 결정 - 사람이 머무르는 공간에만 바람을 내보내 에너지절약
냉장고	<ul style="list-style-type: none"> - 스마트폰 앱을 통한 식품의 종류, 유통기한 등의 정보 관리 및 식품목록 확인 - 온라인 쇼핑 	<ul style="list-style-type: none"> - 온도, 습도, 동작감지, 거리측정, 노크, 문 여닫기 등 각종 센서를 부착하여 사용자의 행동을 인지 - 도어가 열리는 횟수와 시간을 분석해 절전 운전 - 온도와 습도가 높은 한여름에는 음식물이 쉽게 상하지 않도록 제균 기능 강화
로봇청소기	<ul style="list-style-type: none"> - 원격 청소 제어 	<ul style="list-style-type: none"> - 가구, 전선, 사람, 동물 등의 사물을 인지하고 장애물을 스스로 판단하여 보다 꼼꼼하게 청소
세탁기	<ul style="list-style-type: none"> - 스마트폰 앱을 통한 자가진단 및 세탁 종료 알림 - 원격 세탁 제어 	<ul style="list-style-type: none"> - 집안 환경분석 및 세탁물 오염도 확인을 통해 최적 세탁옵션(세탁코스와 시간, 물의 양과 온도 등) 결정 - 날씨를 파악하여 습한 날씨에는 보다 강력한 탈수를 제공하고, 미세먼지가 많은 날은 행굼 시간을 추가 - 고객이 자주 적용하는 세탁옵션을 학습하고 상황에 맞는 세탁옵션 추천

<자료> 이병주, “OCF 표준적용 및 추진전략,” OCF Korea Forum 창립기념세미나, 2017. 3. 28, 재구성

사물데이터의 특징 (1/2)

- 최근 들어 많은 IoT 플랫폼에서 **사물데이터 저장기능** 제공
 - 센서 데이터, 이벤트 데이터, 환경 데이터, 행동 데이터 등
 - 지금까지 대부분은 휘발성 데이터로 단기 저장 후 삭제되었음

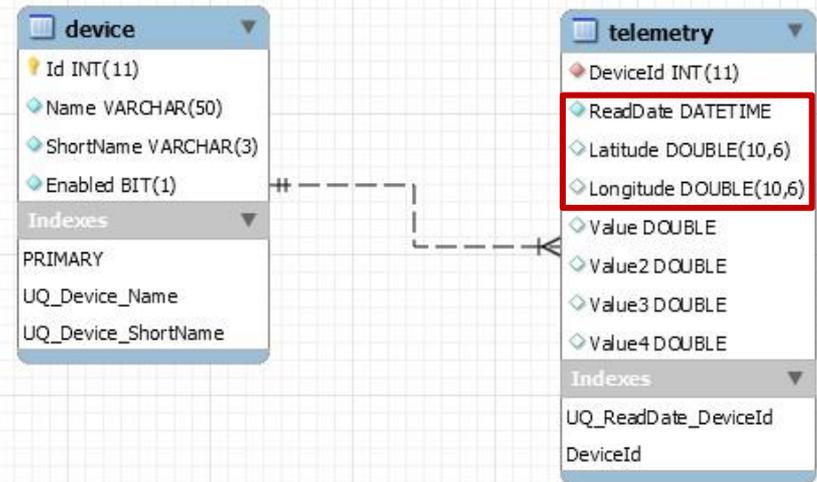
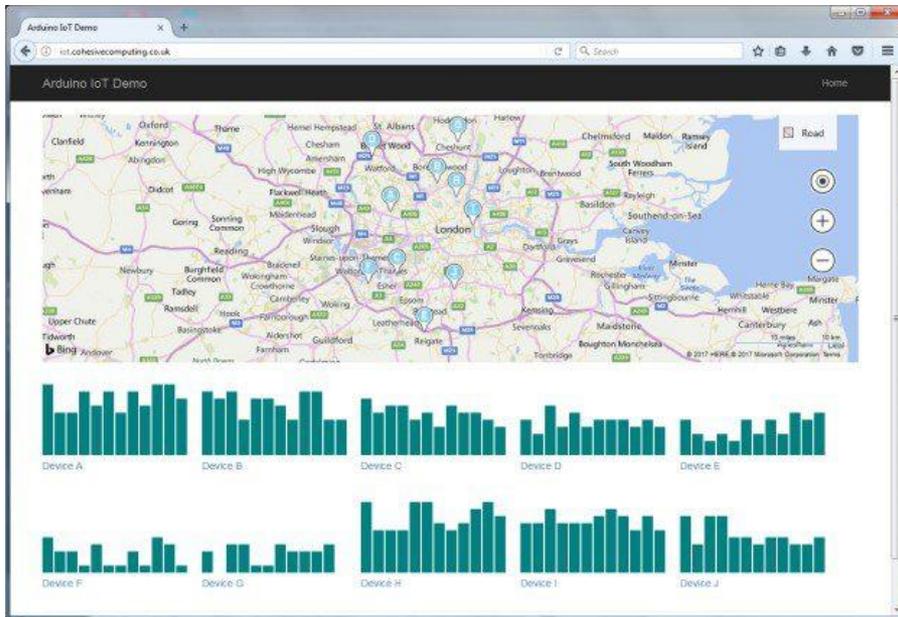


사물데이터 예

사물ID	시간	위도	경도	온도	습도
1	17:12	34	123	33	...
2	17:12	35	128	32	...
1	17:14	34	123	34	...
3	17:15	36	122	36	...

사물데이터의 특징 (2/2)

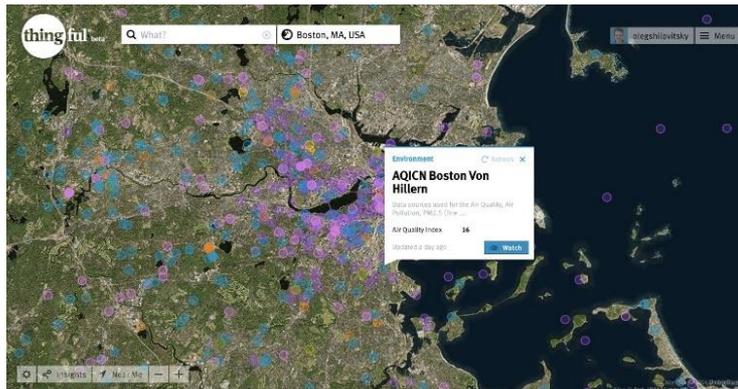
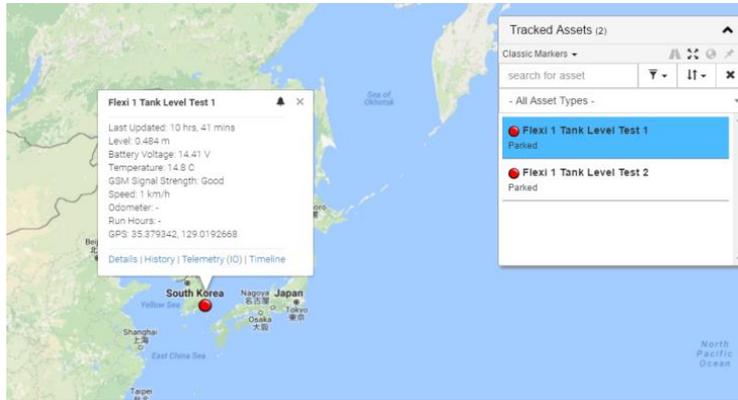
- 사물데이터는 크게 다음 4가지 특징을 가짐
 - **시공간**: 일반적으로 생성된 위치 및 시간 정보를 포함
 - **이종**: 서로 다른 사물의 데이터는 서로 다른 형태를 가짐
 - **대용량**: 다수의 사물이 오랜 시간 동안 데이터를 생성
 - **연속성**: 각 사물이 지속적으로 데이터를 생성



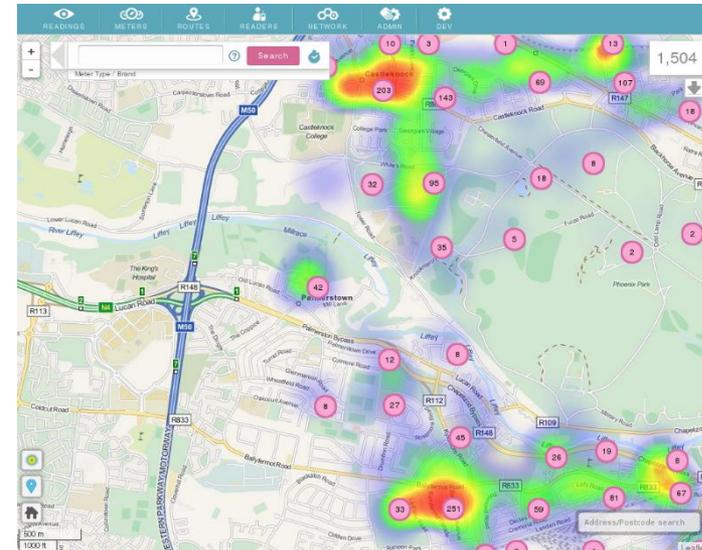
사물데이터 검색 기술 현황 (1/2)

- 하지만 지금까지 사물데이터에 **특화된** 검색 기술은 거의 존재하지 않음
 - 단순 브라우징: 사물 선택 후, 해당 사물의 현재 또는 과거 데이터 탐색
 - 단순 검색: 지정한 공간 또는 시간 범위 내에 수집된 사물 데이터 검색

Flexi



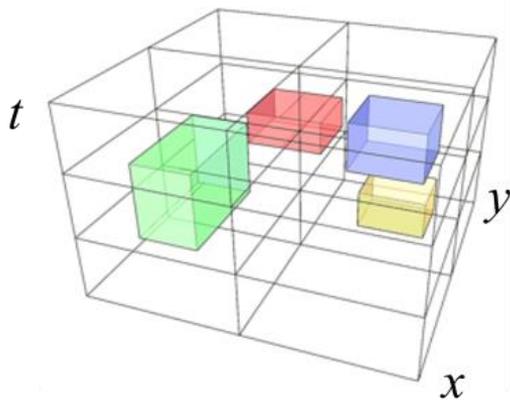
Thingful



Tematra

사물데이터 검색 기술 현황 (2/2)

- 따라서 사물데이터의 특성을 반영한, 사물데이터에 특화된 차세대 검색 기술 개발이 필요함
 - 시공간: 시간 또는 공간 정보에 대한 새로운 검색 기술
 - 이종: 다양한 사물이 생성한 이종 데이터에 대한 통합 검색 기술
 - 대용량: Scale-out 방식의 분산 저장 및 검색 기술
 - 연속성: 최소한 주기적으로 새로운 내용을 검색 결과에 반영



사물ID	시간	위도	경도	온도	습도	PM2.5	...
1	17:12	34	123	33	-	-	...
2	17:12	35	128	-	-	26	...
1	17:14	34	123	34	-	-	...
3	17:15	36	122	-	63	-	...
...

시공간

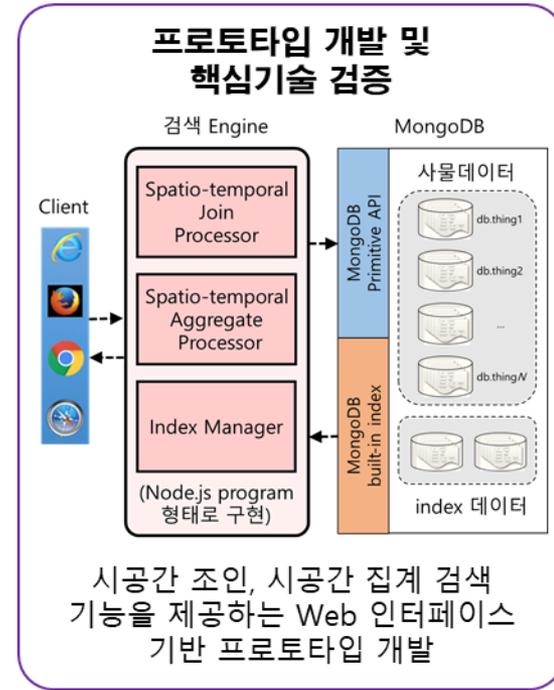
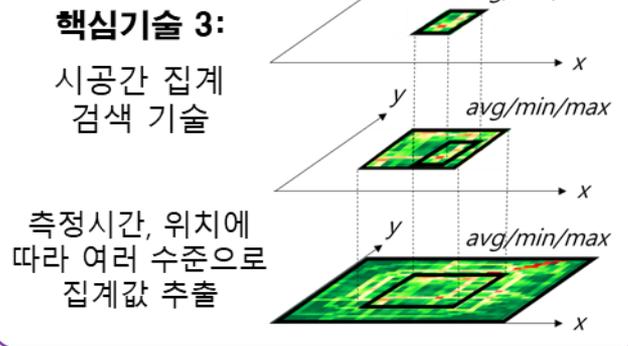
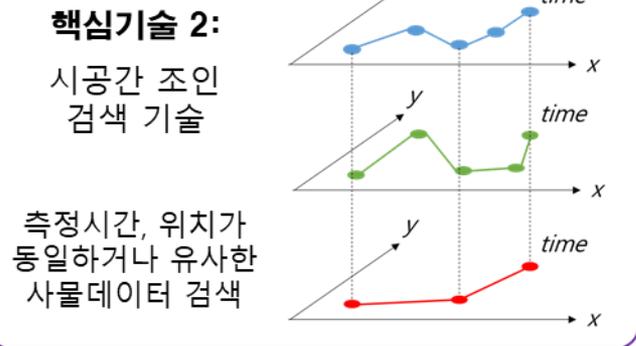
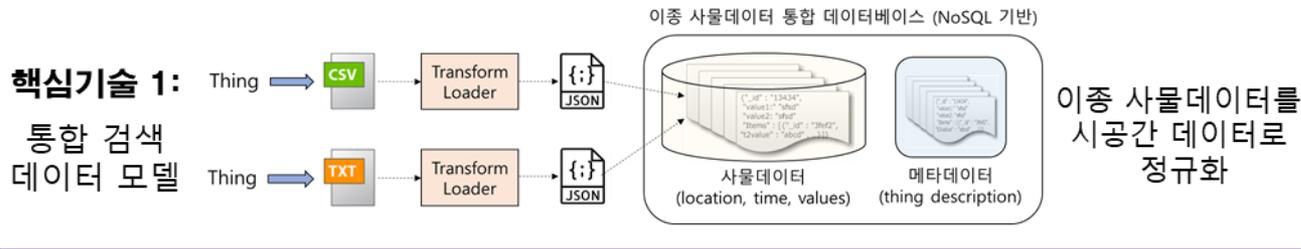
이종

대용량 연속성

현재 진행 중인 과제

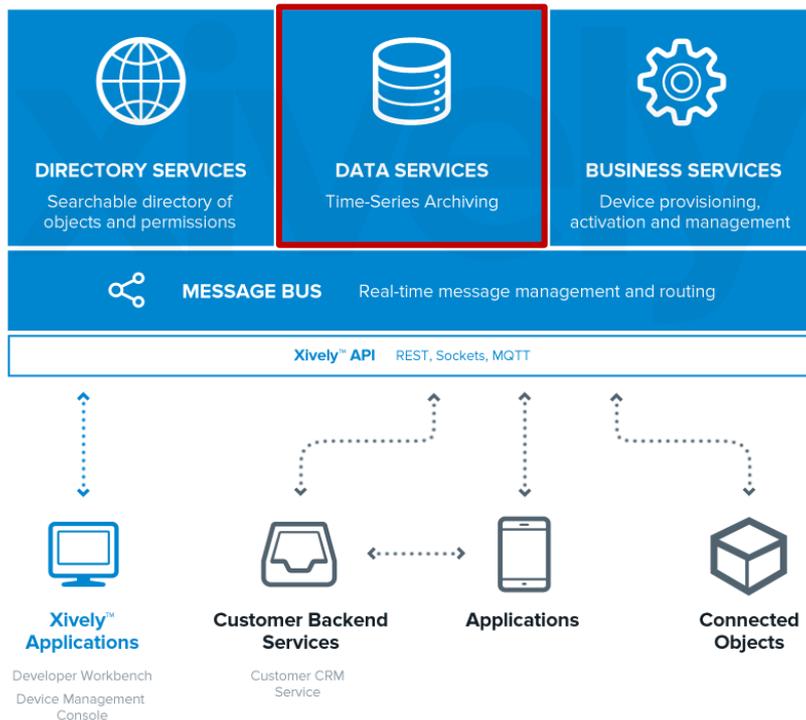
과제명: 사물데이터 검색엔진 기술연구

- 위탁기관: 한국과학기술정보연구원 (2018.5.31 - 10.31)
- 연구책임자: 이기용 (숙명여자대학교 소프트웨어학부 부교수)



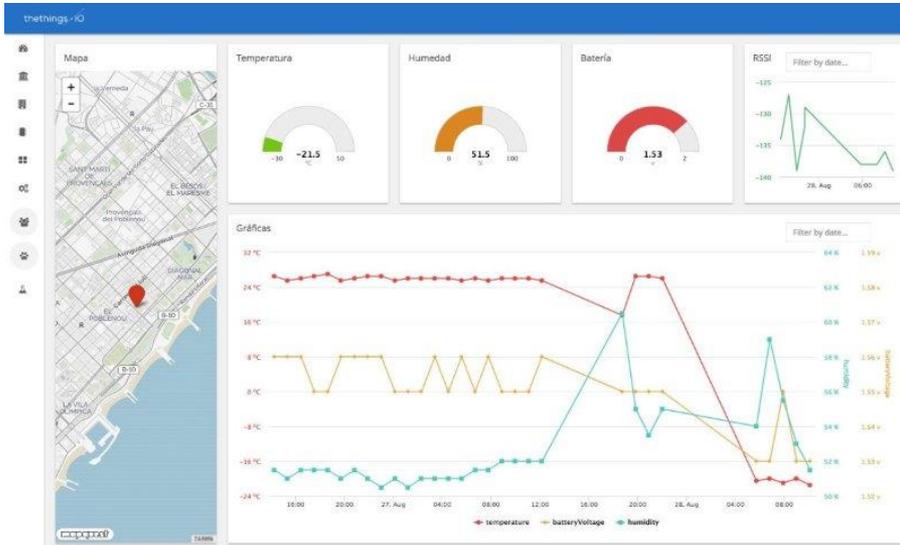
사물데이터 활용 서비스 동향 (1/12)

- 최근 들어 사물들이 발생시킨 데이터를 활용하는 기능을 제공하는 IoT 플랫폼들이 속속 개발되고 있음
 - 단순 사물간 데이터 전송 기능 → 사물데이터 저장, 검색, 분석 기능
- 사물데이터 저장, 검색, 분석 서비스 제공 플랫폼 예

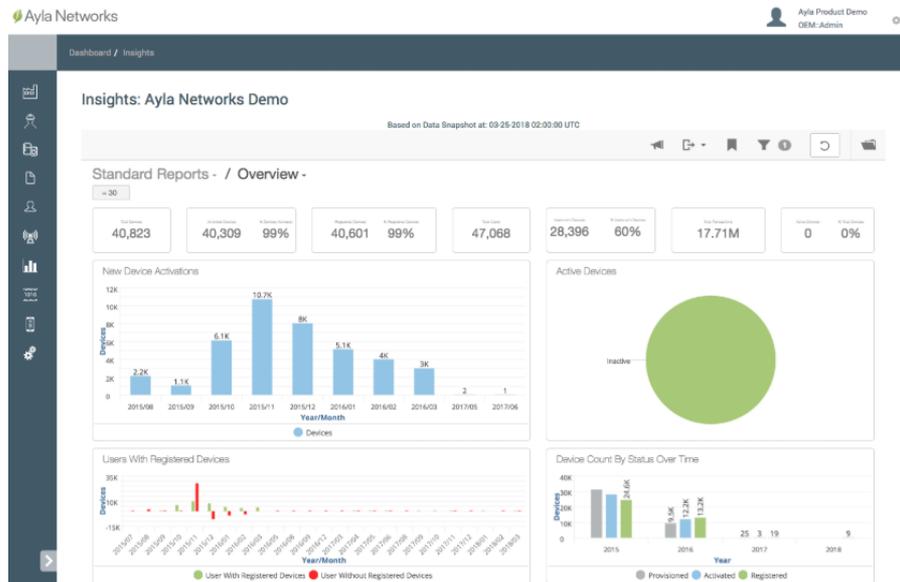


- Xively (<https://xively.com/>)
 - 최근 Google이 인수한 IoT 플랫폼
 - 사물 데이터를 **time series** 형태로 저장 (Timestamp, Category, Value, String)
 - **Timestamp**에 따른 데이터 검색 가능 (최신 데이터, 지정 범위 데이터)
 - Xively API로 time series data 접근 가능

사물데이터 활용 서비스 동향 (2/12)

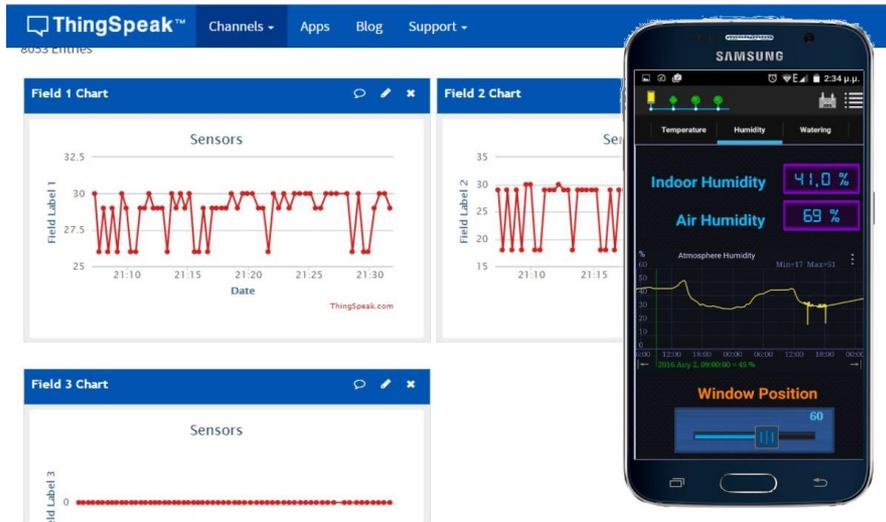


- TheThings.io (<https://thetings.io>)
 - 플랫폼에 연결된 device의 실시간 데이터 제공 (위치, 온도 등)
 - 개발자는 API를 통해 각 device의 실시간 정보를 얻을 수 있음
 - AWS 상에서 구현됨

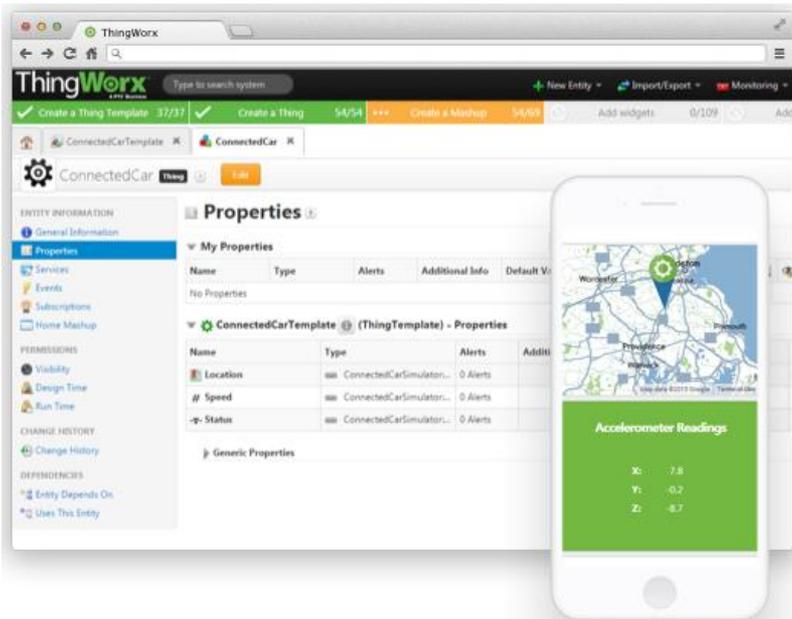


- Ayla Networks (<https://www.aylanetworks.com>)
 - 대량의 장치를 관리할 수 있도록 해주는 기업용 IoT 플랫폼
 - 각 장치 데이터를 표준 포맷으로 저장
 - 데이터 분석을 위해 데이터를 외부 cloud로 내보내는 기능 제공
 - 각 장치의 성능 및 상태 데이터 제공

사물데이터 활용 서비스 동향 (3/12)

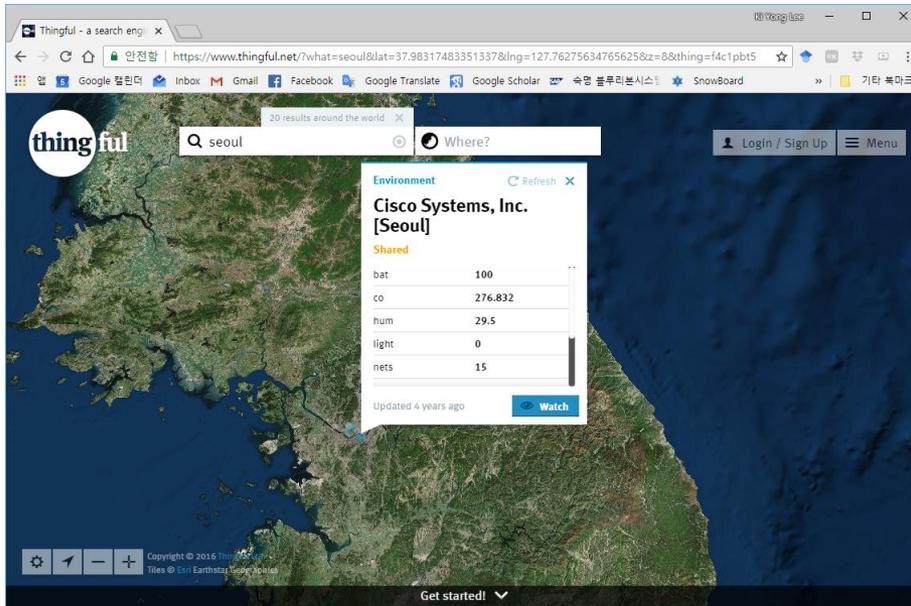


- ThingSpeak (<https://thingspeak.com/>)
 - Sensor data를 cloud에 저장 (온도, 습도, 압력 데이터 등)
 - 데이터 분석, 시각화 기능은 **MATLAB**과 연동하여 제공
 - 데이터 변환, 통계, 결합 기능 제공
 - 특정 event 발생 시 trigger 기능 제공



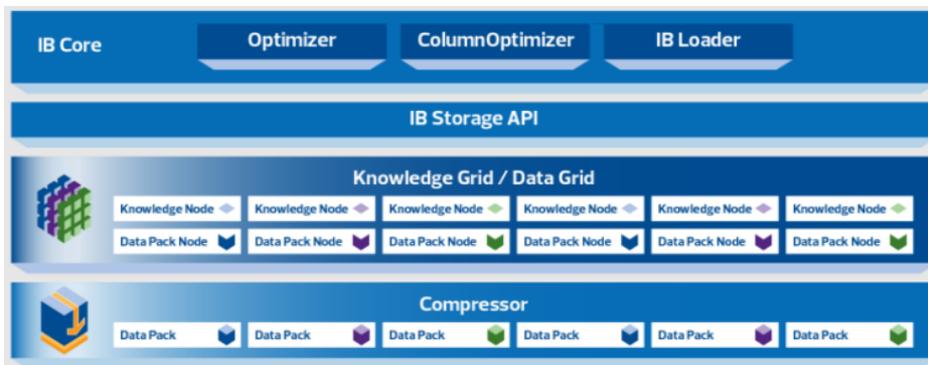
- ThingWorx (<https://thingworx.com>)
 - 데이터 기반 의사 결정 지원 IoT 플랫폼
 - 각 thing을 모델로 표현할 수 있음
 - **SQUEAL (Search, Query, Analysis):** 모델 속성 기반 검색 기능 제공 (name, tag, type, shape 등)
 - 데이터 분석 기능 제공 (예측 모델 생성, 이상치 탐지, 실시간 모니터링 등)

사물데이터 활용 서비스 동향 (4/12)

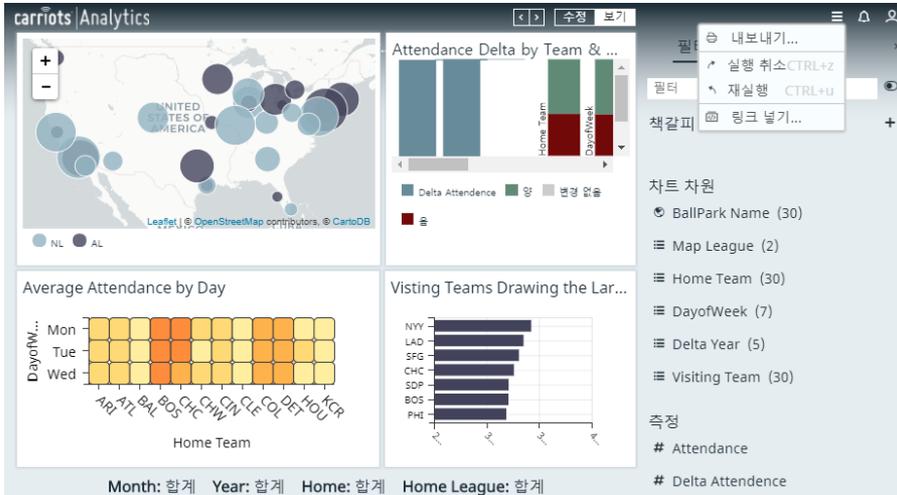


- thingful (<https://www.thingful.net/>)
 - 전세계 대상 IoT 데이터 검색 엔진
 - 등록된 thing의 데이터를 실시간으로 접근할 수 있는 API 제공
 - Datapipes: 사물데이터 전송 제공
 - 위치에 따른 단순 검색만 가능

- InfoBright (<http://www.infobright.com/>)
 - IoT 데이터에 특화된 **column-oriented relational database**
 - SQL 질의 및 그의 분산 처리 기능 제공
 - 분산 IoT 데이터 접근을 위해 **Knowledge Grid**라는 색인과, Knowledge Grid를 활용하여 질의를 처리하는 Optimizer layer를 가짐

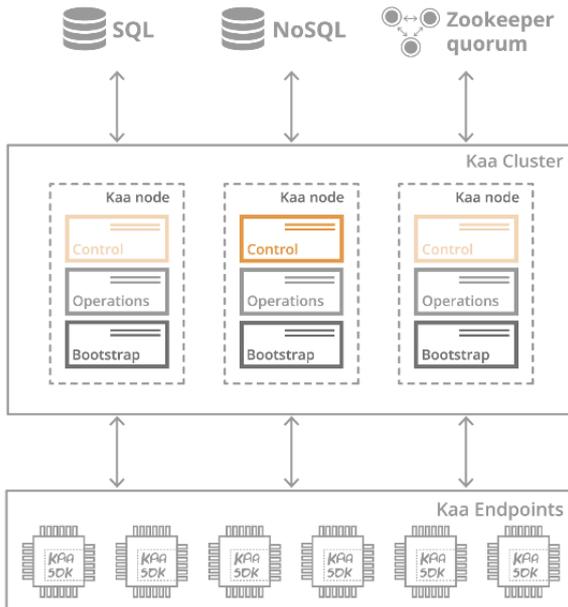


사물데이터 활용 서비스 동향 (5/12)



■ Carriots (<https://carriots.com/>)

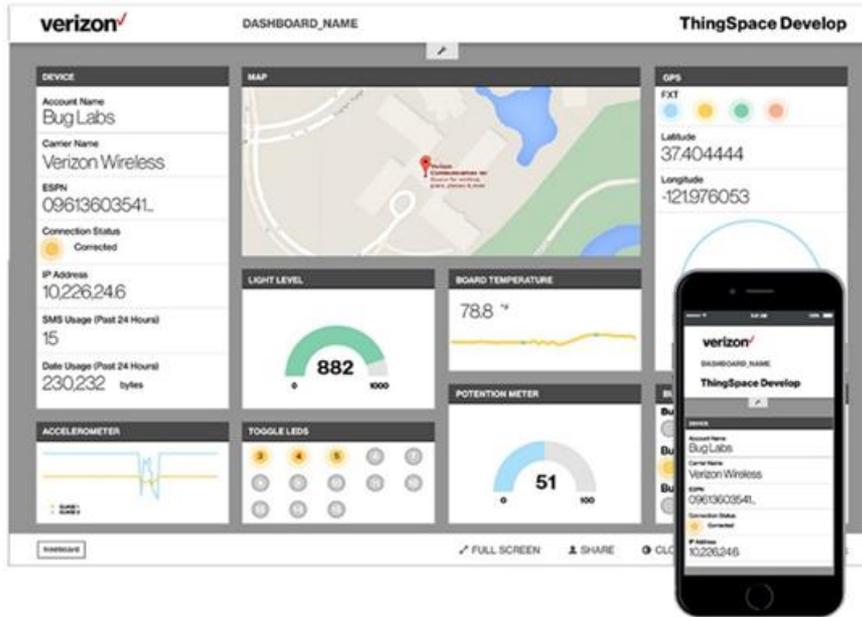
- **Carriots Platform:** IoT 데이터 수집, 저장
- **Carriots Analytics:** 데이터 시각화, 공유, 내보내기 제공 (특히 시각화에 집중)
- 데이터 접근을 위한 API 및 SDK 제공
- 데이터는 자체 개발 NoSQL에 저장



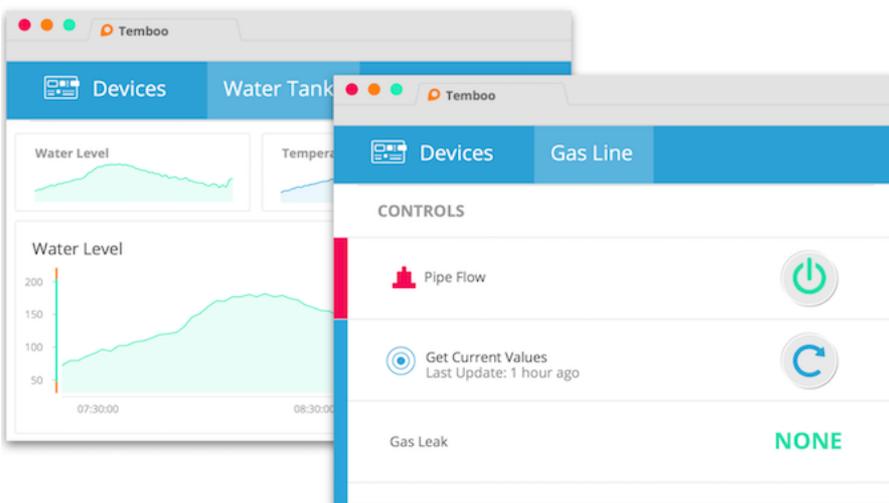
■ KAA (<http://www.kaaproject.org/>)

- Cloud 기반의 open source IoT 플랫폼
- Structured/unstructured data 저장 가능
- **데이터 분석:** time series에 대한 filtering 및 aggregate 기능 지원
- **데이터 시각화:** 차트, 지도, 테이블, 게이지 등 여러 widget 지원
- Storage로 SQL과 NoSQL 사용 가능

사물데이터 활용 서비스 동향 (6/12)

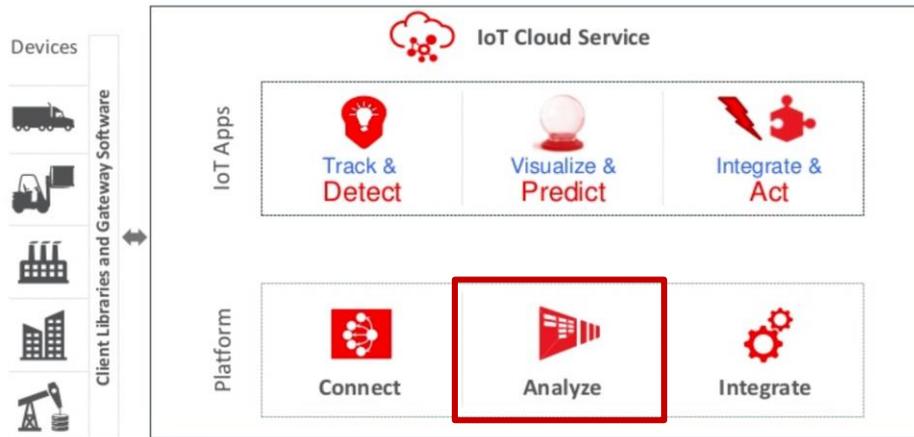


- Freeboard (<https://freeboard.io/>)
 - 각 thing들의 현재 상태를 시각화해서 보여주는 것이 주요 목적인 IoT 플랫폼
 - Open source며 쉬운 GUI 구현 기능 제공
 - 표, 그래프, 차트 등 다양한 widget 제공
 - 복잡한 질의, 분석 기능은 제공하지 않음

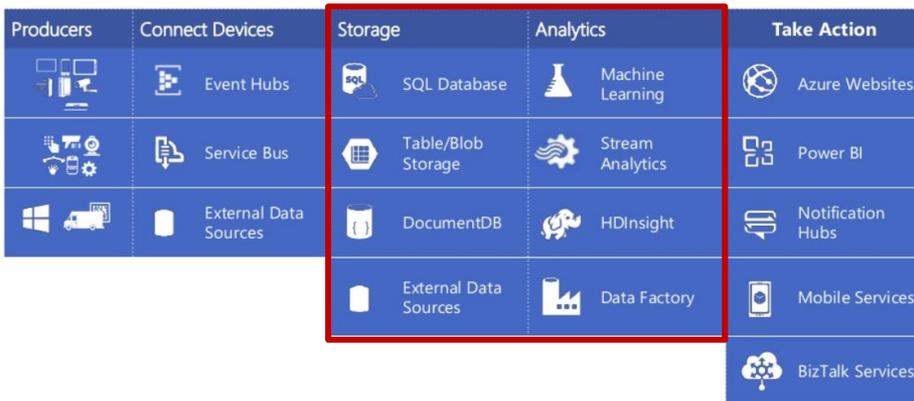


- temboo (<https://temboo.com/>)
 - 각 장치들이 발생시킨 데이터를 저장, 접근, 시각화, 공유하는 기능 제공
 - 저장: XLS, CVS, XML, JSON 형태로 저장
 - 시각화: 시간에 따른 그래프 출력 기능
 - 공유: 이메일, 클라우드로 전송 기능
 - 고수준 검색, 질의 기능 미제공

사물데이터 활용 서비스 동향 (7/12)

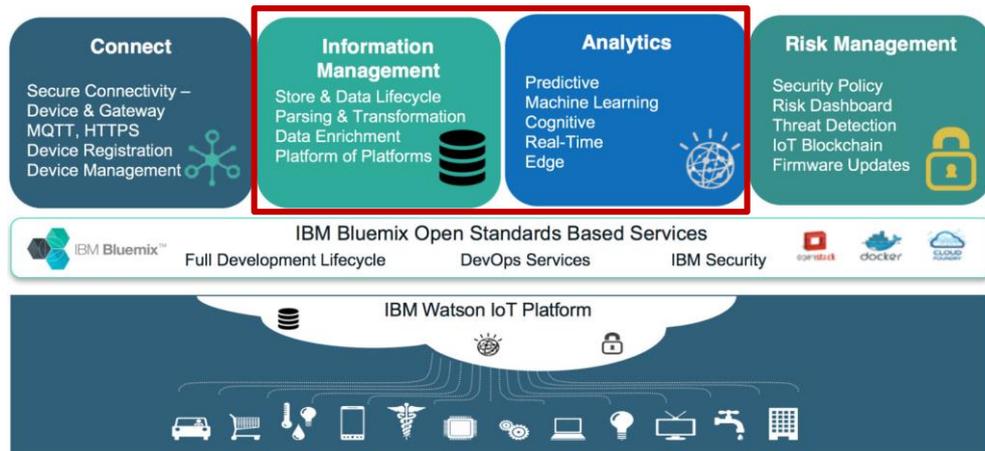


- Oracle IoT (<https://cloud.oracle.com/iot>)
 - 데이터 저장, 분석, 기계학습 기능 제공
 - **저장**: Oracle database 사용
 - **분석**: 실시간 스트림/이벤트 처리 및 SQL에 기반한 ad-hoc 질의 기능 제공
 - **기계학습**: built-in 기계학습 기능 제공 (이상치 탐지, 예측모델 생성, 추천)

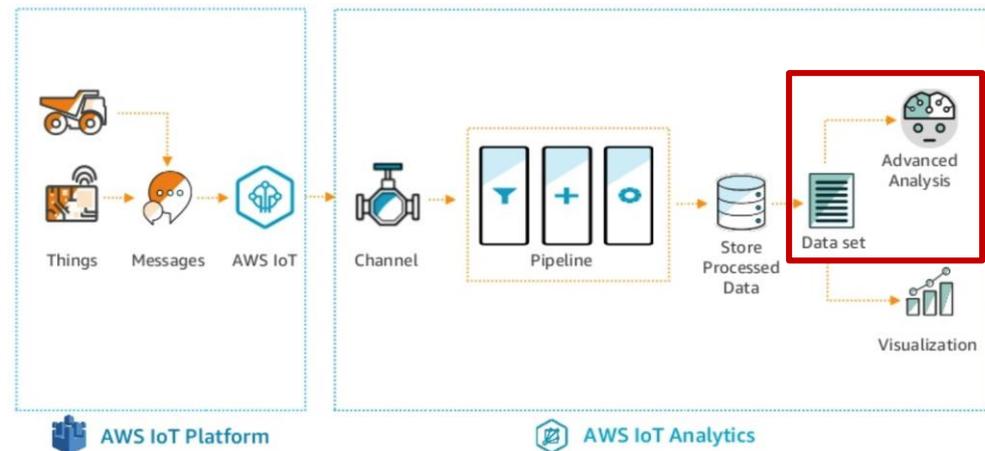


- MS Azure IoT suite (<https://azure.microsoft.com/overview/iot>)
 - IoT 관련 여러 solution들의 집합
 - IoT Hub: 사물 연결, 모니터링, 제어
 - IoT Edge: 클라우드가 아닌 단말에서 데이터 분석 기능 제공 (filtering 등)
 - **저장**: Azure Cosmos DB (SQL + NoSQL)
 - **분석**: Time Series Insights, Machine Learning, Stream Analytics 등 다른 제품들과 연계

사물데이터 활용 서비스 동향 (8/12)

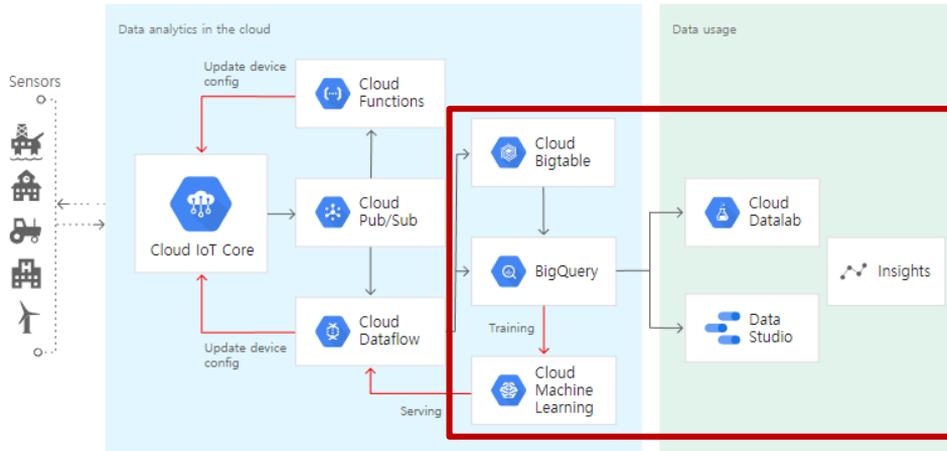


- IBM Watson IoT platform (<https://www.ibm.com/internet-of-things>)
 - 실시간 데이터 접근 및 과거 데이터 저장 기능 제공 (IBM Bluemix cloud)
 - 장치 데이터 및 성능 데이터에 대한 정제, 필터링, 변환, 집계 기능 제공
 - **Watson API**로 자연어처리, 기계학습, 이미지 및 텍스트 분석 기능 제공



- AWS IoT (<https://aws.amazon.com/iot>)
 - 여러 AWS service로 구성됨
 - AWS IoT Core: Cloud 기반 IoT 플랫폼 (기기 연결, 관리, 통신, 데이터 전송)
 - **AWS IoT Analytics**: 장치 데이터 filtering 및 정제, time series 변환 및 저장, SQL 질의 기능, machine learning 기능 제공 (예측 모델 등)

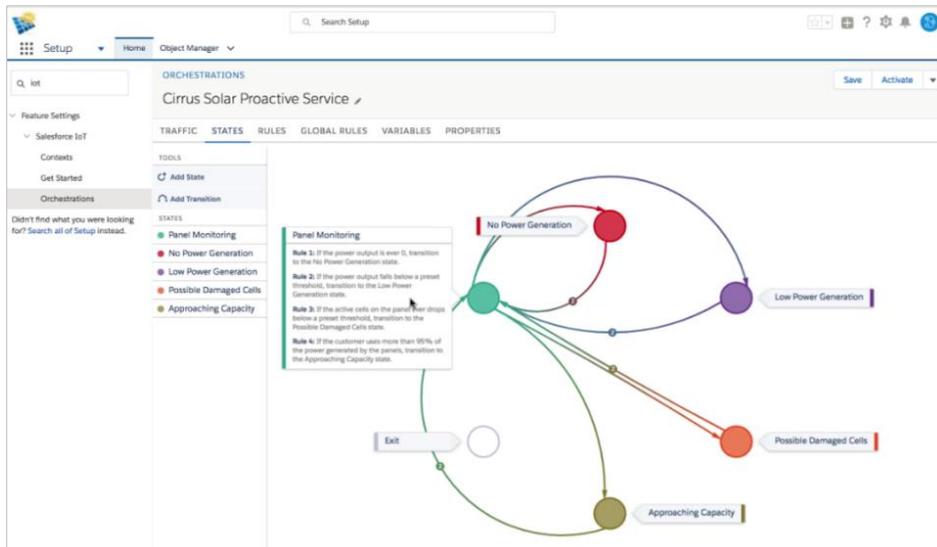
사물데이터 활용 서비스 동향 (9/12)



■ Google Cloud IoT

(<https://cloud.google.com/solutions/iot/>)

- IoT 데이터에 대한 분석 및 시각화 기능
- **Cloud Bigtable**: 데이터 저장소 (NoSQL)
- **BigQuery**: 데이터 분석 (SQL 사용)
- **Cloud Machine Learning**: 기계학습 기능
- **Data Studio**: 보고서, 시각화
- **Cloud Datalab**: BQ, ML 대화형 interface

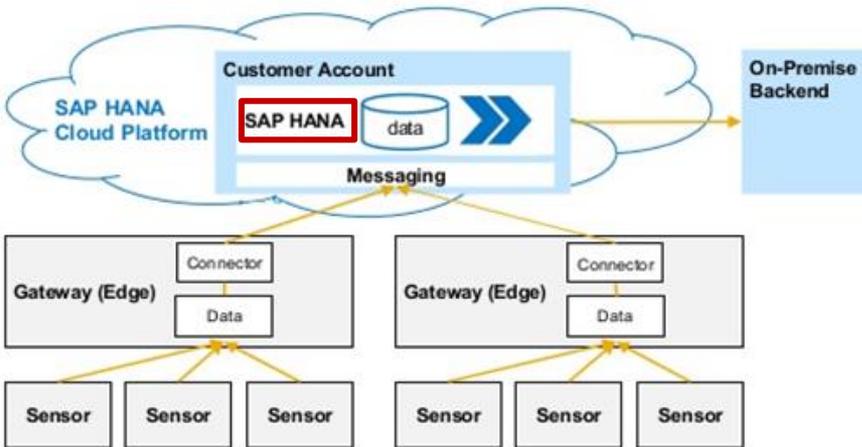


■ Salesforce IoT Cloud

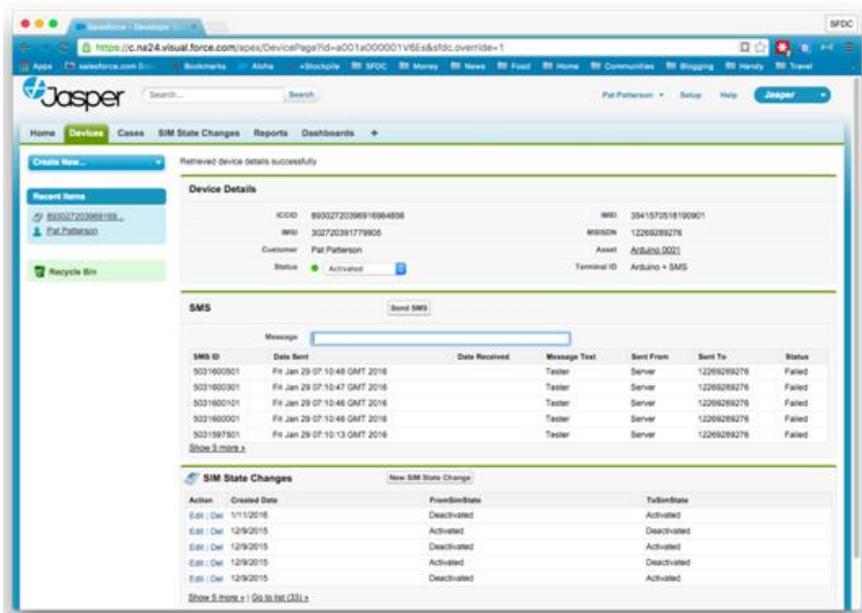
(<https://www.salesforce.com/products/salesforce-iot/>)

- 사물 데이터를 Salesforce cloud로 전송
- 다양한 장치 데이터를 집계, 필터링, 통합 등을 거쳐 JSON, CSV 등 형태로 저장 가능
- 실시간 규칙 검사로 action trigger 가능
- 시각화 외 고수준 질의 미제공

사물데이터 활용 서비스 동향 (10/12)

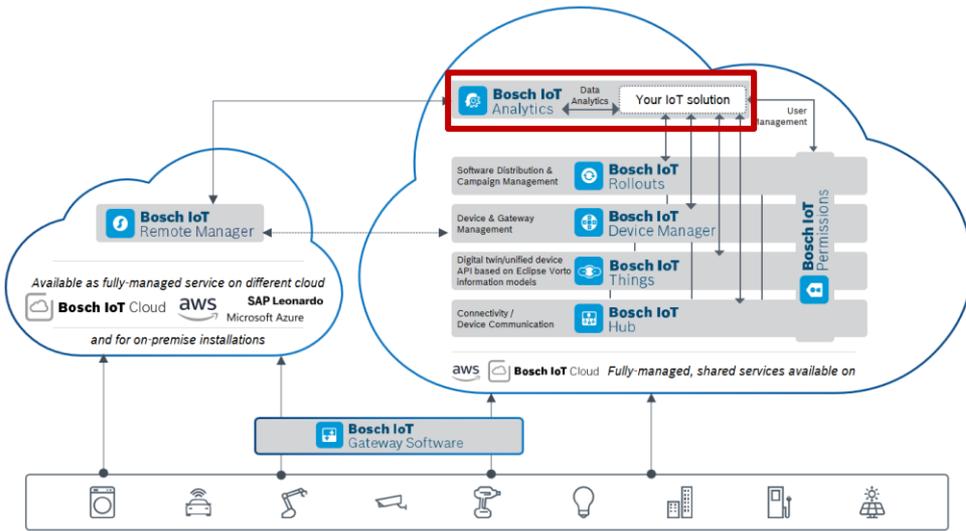


- SAP Cloud Platform for IoT (<https://www.sap.com/products/iot-platform-cloud.html>)
 - 실시간 처리 목표 in-memory IoT 플랫폼
 - **SAP HANA**(in-memory RDBMS)가 제공하는 geospatial processing (위치 기반 탐색), time-series processing (집계, 예측, 비교), 시각화 기능 사용 가능
 - Data stream에 대한 패턴 탐지 기능 제공

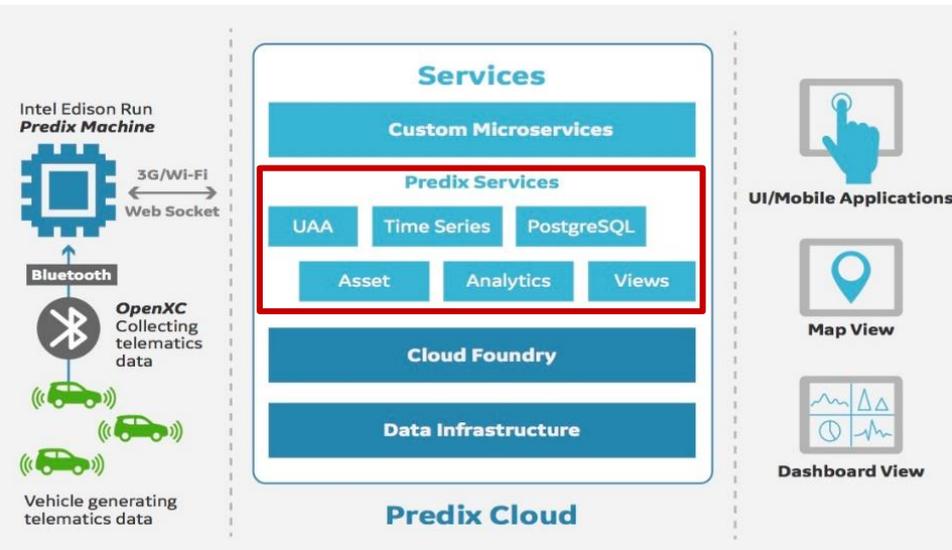


- Cisco Jasper Control Center for IoT (<https://www.jasper.com/control-center-for-iot>)
 - 실시간 device 제어, 관리, 진단 기능 제공
 - 사물데이터에 대한 실시간 집계 뷰 제공
 - **Time-series pattern 분석** 기능 제공
 - 고급 질의 및 분석 기능 미제공

사물데이터 활용 서비스 동향 (11/12)

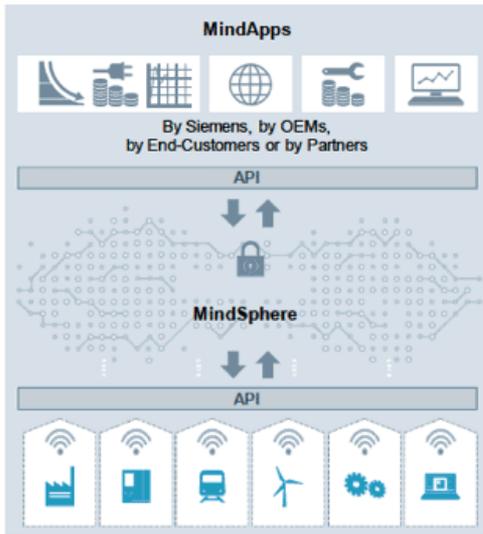


- Bosch IoT Suite (<https://www.bosch-iot-suite.com/>)
 - 클라우드 기반 Platform-as-a-Service IoT 플랫폼으로 여러 service로 구성
 - **Bosch IoT Analytics**: device data 분석 기능 제공 (핵심은 anomaly detection, 그 외 device 이동 패턴 파악, device data monitoring & alarm 기능 제공)
 - Data mining 및 기계학습 기술 사용



- GE Predix (<https://www.predix.io/>)
 - 산업 장비 데이터 수집 및 분석 플랫폼
 - **Data management**: data modeling, time series 저장, 색인, 탐색 기능 제공
 - **Data analytics**: 외부 machine learning library를 사용하여 anomaly detection, text mining, time series 분석 등 제공
 - PostgreSQL 및 Elasticsearch 사용

사물데이터 활용 서비스 동향 (12/12)



- MindApps**
 - Analytical insights
- MindSphere**
 - Various cloud infrastructures
- MindConnect**
 - Plug and play connection

- Siemens MindSphere (<https://siemens.mindsphere.io/>)
 - 클라우드 기반 산업용 IoT 플랫폼
 - **Storage services:** time series service (time series 저장, 색인, 범위 탐색), time series aggregates service (day/hour/min 단위 최대/최소/평균)
 - **Analytics services:** anomaly detection, frequent pattern, trend prediction 등



- Samsung Artik IoT Platform (<https://www.artik.io/>)
 - 다양한 H/W와 S/W들을 cloud로 연결
 - Subscription & notification 기능 제공
 - Device data를 표준화하여 저장
 - 저장된 data에 대한 **aggregate** 및 **시각화** 기능 제공

사물데이터 활용 서비스 동향: 요약 (1/2)

■ 사물데이터 저장 기능

- 현재 대부분의 IoT 플랫폼은 데이터 저장소로 다음을 사용
 - **RDBMS**: Oracle, MySQL, PostgreSQL, ...
 - **NoSQL**: document database (MongoDB), column store (HBase, Cassandra)
 - **File system**: HDFS(Hadoop) 또는 자체 storage

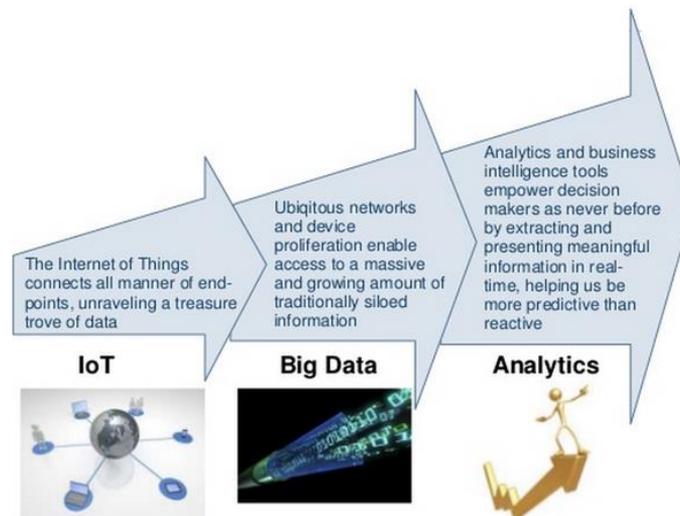


- RDBMS나 NoSQL을 사용하는 경우 해당 시스템의 질의/색인 기능 제공
 - 질의: 대부분 SQL (또는 SQL-like) 질의 기능 제공
 - 조건 검색, 범위 검색, 집계 질의
 - 색인: B-tree, inverted index, hash index, R-tree 등 제공
 - 빠른 조건 검색, 범위 검색 지원

사물데이터 활용 서비스 동향: 요약 (2/2)

■ 사물데이터 활용 기능

- 현재 대부분의 IoT 플랫폼은 다음과 같은 **검색 기능**을 제공
 - 조건을 만족하는 특정 사물의 데이터 검색 (실시간, 과거)
 - Time series 데이터 검색 및 집계 기능
- 최근 들어 많은 IoT 플랫폼이 다음과 같은 **분석 기능**을 추가하는 추세
 - Prediction: 시간 흐름에 따른 특정 값 예측
 - Anomaly detection: 예측이나 trend를 많이 벗어가는 경우 탐지
 - 기타 Data mining 및 machine learning 기반 분석 기능



사물데이터 활용 연구 논문 동향 (1/3)

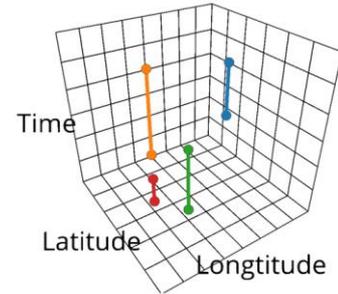
- 사물데이터에 대한 단순 검색 외 다음 검색 기술도 연구됨
 - 유사 thing 검색
 - 유사한 데이터 패턴이나 특징을 가지는 thing 검색
 - 의미 기반 thing 검색
 - Topic 또는 keyword와 관련 있는 thing 검색
- Hadoop 또는 Spark를 사용한 사물데이터 처리 기술 연구
 - Hadoop이나 Spark에서 **사물데이터 분산 저장 및 질의 처리 기법**
 - MapReduce를 활용한 질의 처리 기법
- 사물데이터 저장소 비교 연구
 - **RDBMS**(MySQL)과 **NoSQL**(MongoDB, HBase, ...) 성능 비교
 - 현재는 NoSQL을 사물데이터 저장소로 사용하려는 움직임이 많음
 - Schemaless라는 특징과 확장성 측면에서 유리

사물데이터 활용 연구 논문 동향 (2/3)

■ 많은 IoT 데이터 survey 논문에서 다음 기술 필요성 언급

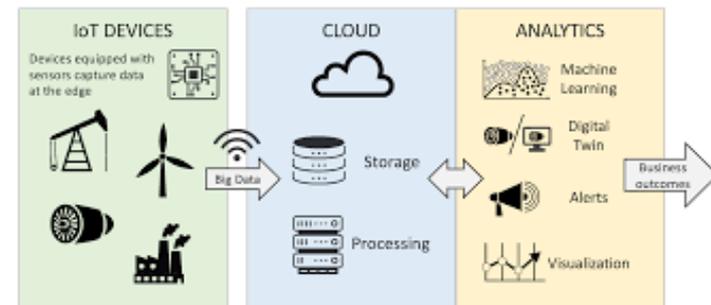
– 시공간 질의 처리 기술

- IoT 데이터는 time series data이자 spatial data임
- 시간 및 공간에 관한 다양한 질의 처리 기술 필요



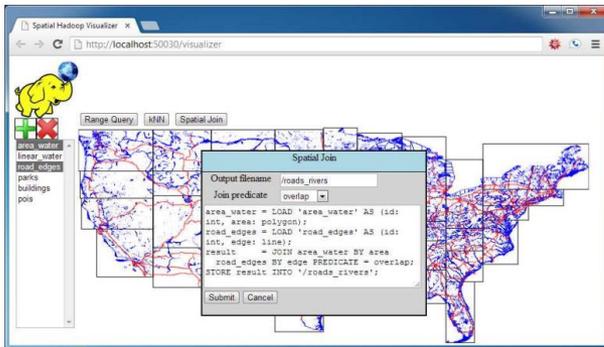
– Data mining 또는 machine learning 기반 데이터 분석 기술

- Prediction: 측정값 예측
- Anomaly detection: 이상치 또는 고장 장치 탐지
- Frequent pattern mining: 자주 발생하는 event 탐지
- Clustering: 유사 패턴을 가진 thing끼리 그룹화
- Classification: 어떤 thing에 발생한 event 판단
- Profiling: 유사 thing들간 공통점 추출

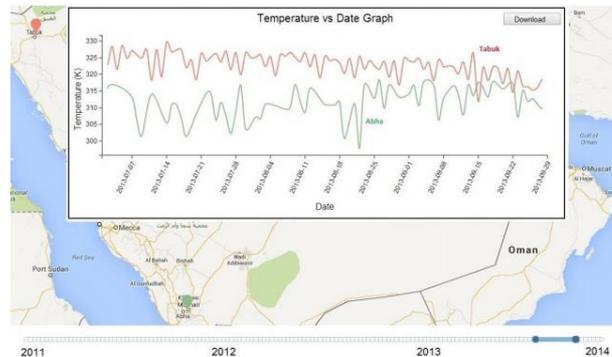


사물데이터 활용 연구 논문 동향 (3/3)

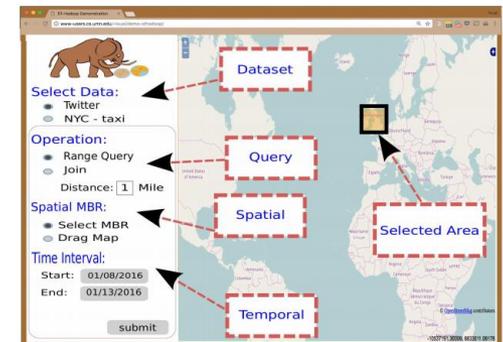
- IoT와는 별도로 최근 대규모 시공간 데이터에 대한 검색 기술이 독립적으로 연구되고 있음
 - IoT 데이터는 시간, 위치가 존재하므로 시공간 데이터로 볼 수 있음
 - 시공간 데이터에 대한 저장, 색인, 질의 지원 시스템
 - SpatialHadoop**: MapReduce를 활용한 공간 데이터 처리 시스템 (2014)
 - SHAHED**: SpatialHadoop 기반 시공간 데이터 처리 시스템 (2015)
 - ST-Hadoop**: MapReduce를 활용한 시공간 데이터 처리 시스템 (2017)



SpatialHadoop



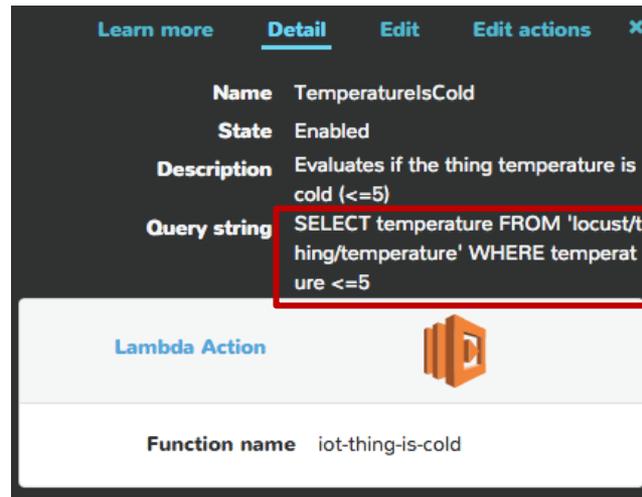
SHAHED



ST-Hadoop

기존 IoT 플랫폼의 한계

- 기존 IoT 플랫폼은 주로 범용 SQL 기반 검색 기능 제공
 - 조건을 만족하는 thing의 데이터 검색 (SELECT)
 - Location, time, state, keyword, type 기반
 - Time series data에 대한 aggregate (GROUP BY)
 - 최대, 최소, 평균, 합계 등 계산

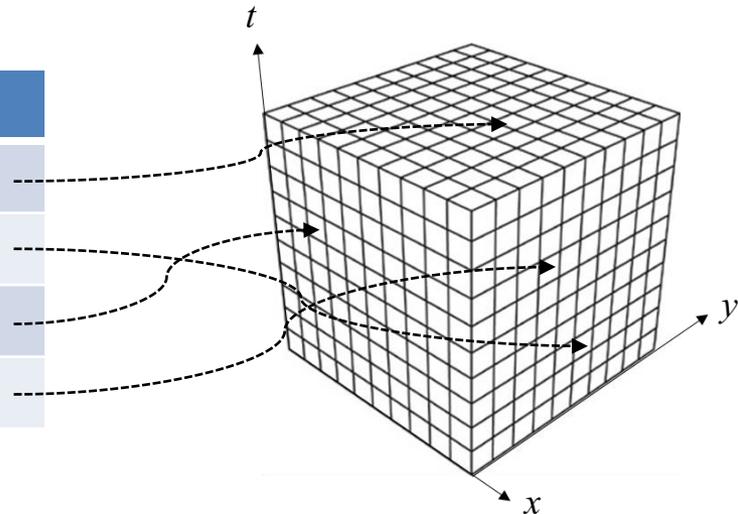


- 아직 IoT 데이터만을 위한 새로운 검색 기술은 없음
 - 대부분은 기존 범용 SQL 처리 기술 활용 (색인, 질의최적화 기술 등)

진행 연구 소개 (1/4)

- 단순 검색 또는 기존 범용 SQL 기술보다 효율적인, **사물데이터에 특화된 검색 기술 개발**
- 사물데이터는 대규모 시공간 데이터로 볼 수 있음
 - 각 데이터는 위치 및 시간 정보를 포함 (측정위치, 측정시간)
 - 각 데이터는 (x, y, t) 3D 공간의 한 점으로 볼 수 있음

시간	위도	경도	온도	...
17:12	34	123	33	...
17:12	35	128	32	...
17:14	34	123	34	...
17:15	36	122	36	...

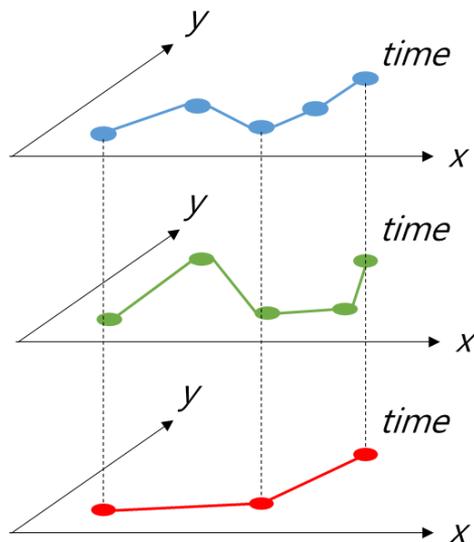


■ 본 연구의 목표

세부목표 1:

시공간 조인 검색 기술

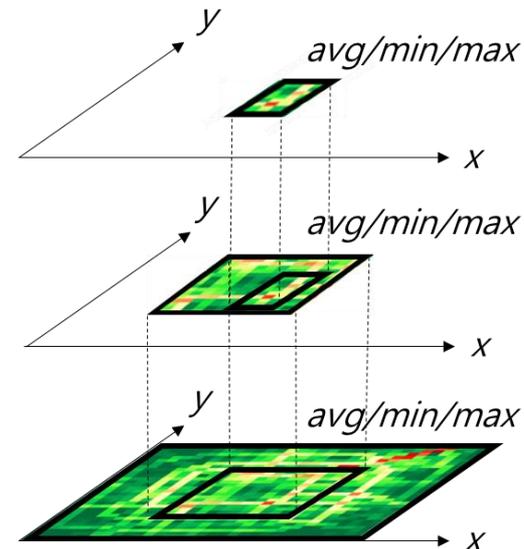
측정시간 및 위치가 동일하거나 유사한 사물데이터 검색



세부목표 2:

시공간 다수준 집계 기술

측정시간 및 위치에 따라 여러 수준으로 집계값 계산

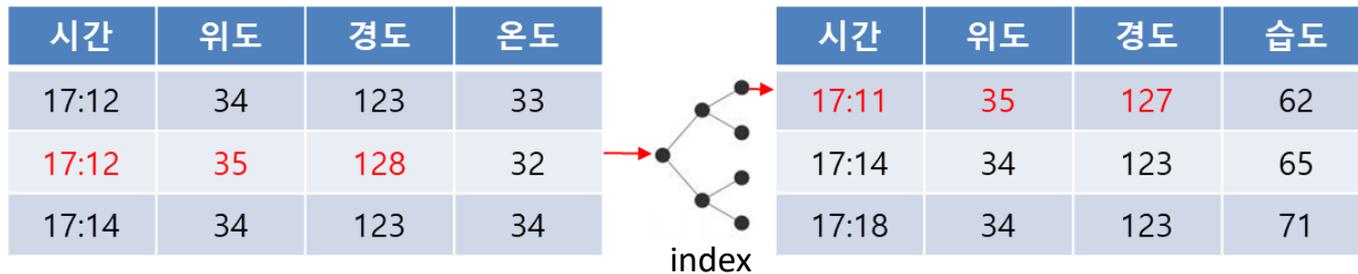


진행 연구 소개 (3/4)

기존 시스템 또는 기술과의 차별점

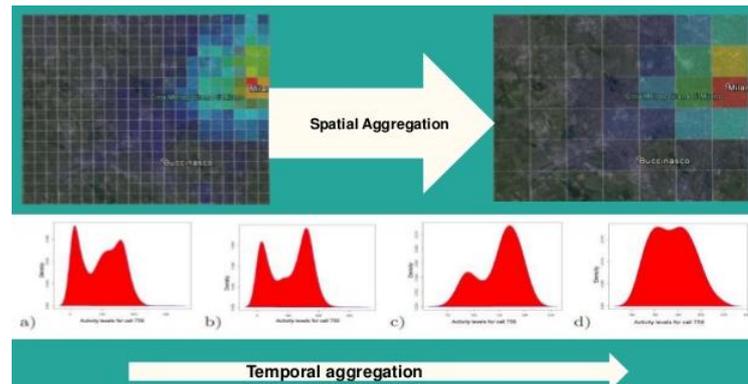
– 시공간 조인

- 기존 RDBMS의 조인 기법을 사용할 수는 있으나 비효율적



– 시공간 다수준 집계

- RDBMS, NoSQL 모두 시공간 다수준 집계를 위한 특별한 기능 미제공

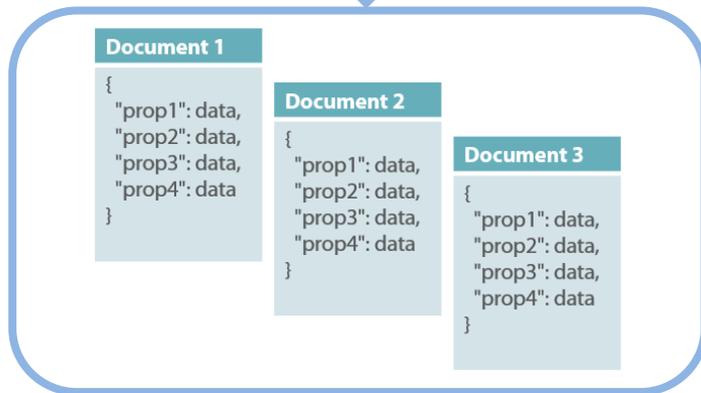


진행 연구 소개 (4/4)

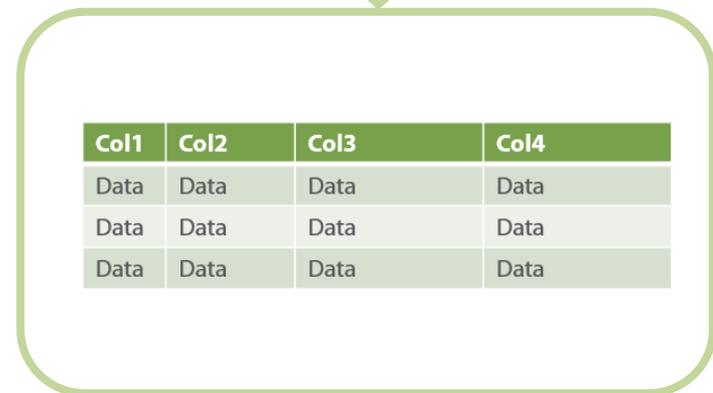
■ 제안 기술의 적용

- 본 연구에서 제안하는 기술은 특정 storage에 종속되지 않음
- Logical scheme 자체는 MongoDB, RDBMS 모두 적용 가능

제안 시공간 기반 검색 기법 & 색인 구조



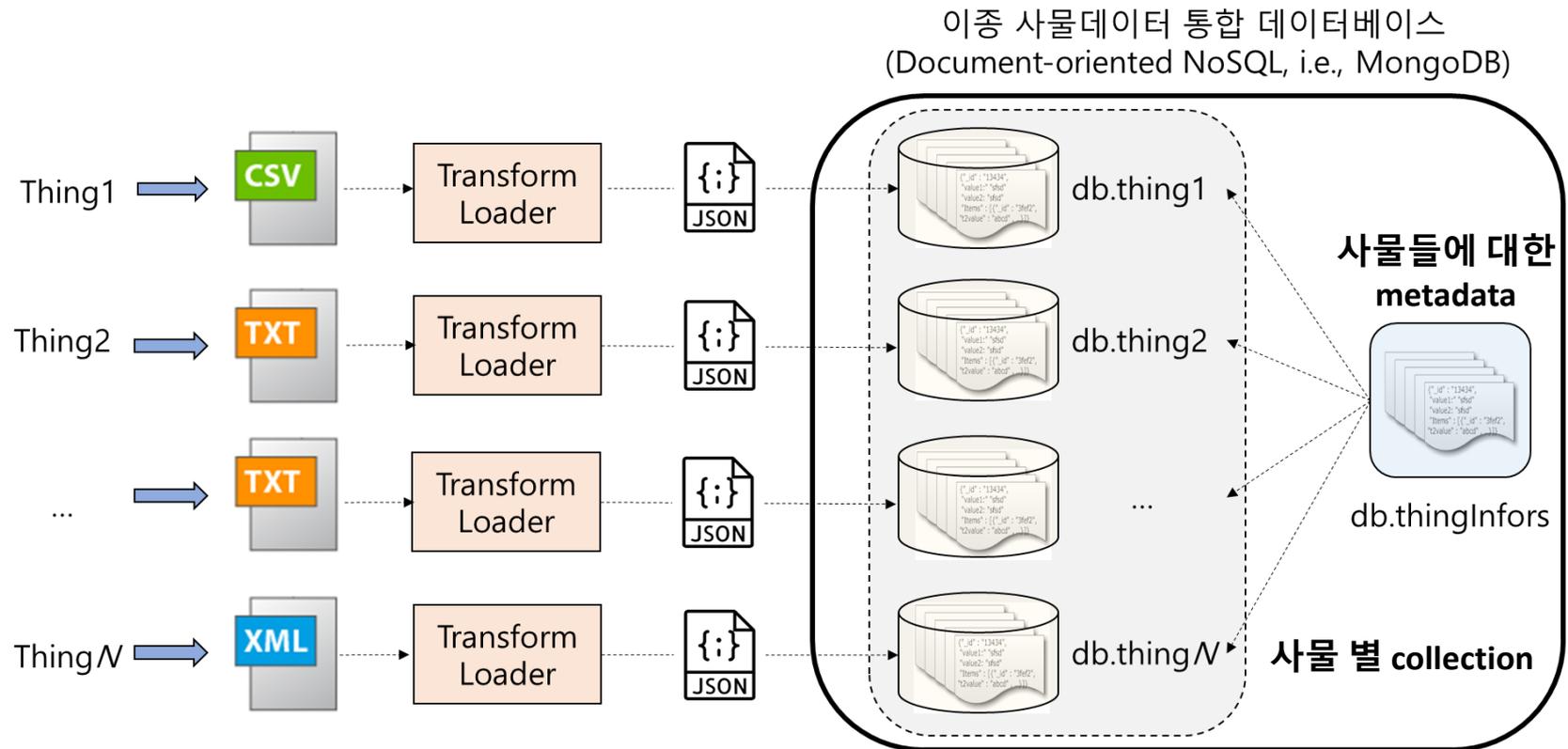
MongoDB



MySQL

사물데이터 검색 시스템 아키텍처 (1/2)

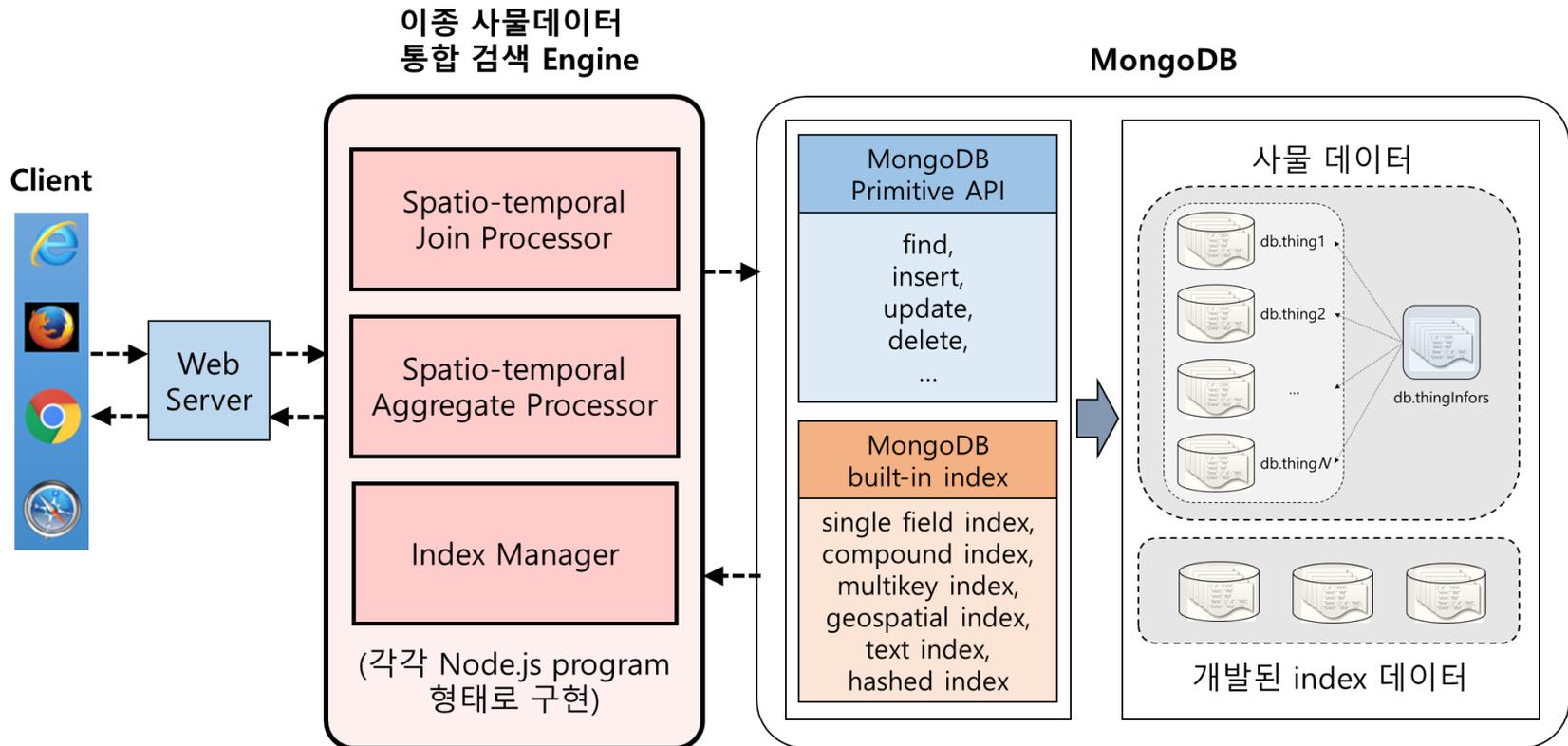
- 사물데이터 저장소로 **MongoDB** 사용
 - 대표적인 document-oriented NoSQL database
 - 서로 다른 사물이 생성한 데이터는 JSON document로 변환 및 표준화되어 MongoDB에 통합 저장됨



사물데이터 검색 시스템 아키텍처 (2/2)

■ 사물데이터 검색 시스템

- MongoDB를 storage로 사용하는 응용 프로그램 형태로 구현
- 웹 기반 인터페이스 제공



시공간 조인 알고리즘 개발

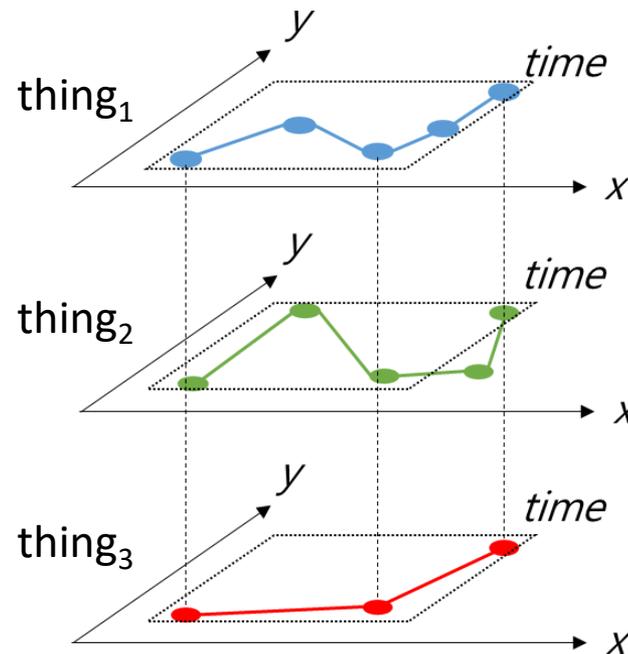
■ 시공간 조인 검색 정의

– Input

- 조인 대상: $thing_1, thing_2, \dots, thing_N$
- 위치 범위: $[x_{min}, y_{min}], [x_{max}, y_{max}]$
- 시간 범위: $[t_{min}, t_{max}]$
- 동일 위치 임계치: θ_S
- 동일 시간 임계치: θ_T

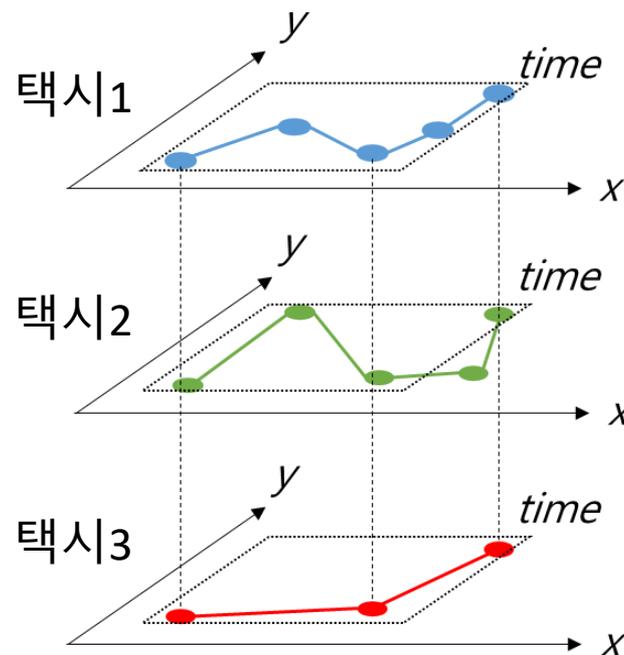
– Output

- 주어진 위치 및 시간 범위 내에서, 동일 위치 및 동일 시간에 측정된 $thing_1, thing_2, \dots, thing_N$ 들의 데이터
- 단, $|(x_1, y_1) - (x_2, y_2)| < \theta_S$ 이면 (x_1, y_1) 과 (x_2, y_2) 를 동일 위치로 간주
- 단, $|t_1 - t_2| < \theta_T$ 이면 t_1 과 t_2 를 동일 시간으로 간주



시공간 조인 검색 사용 시나리오

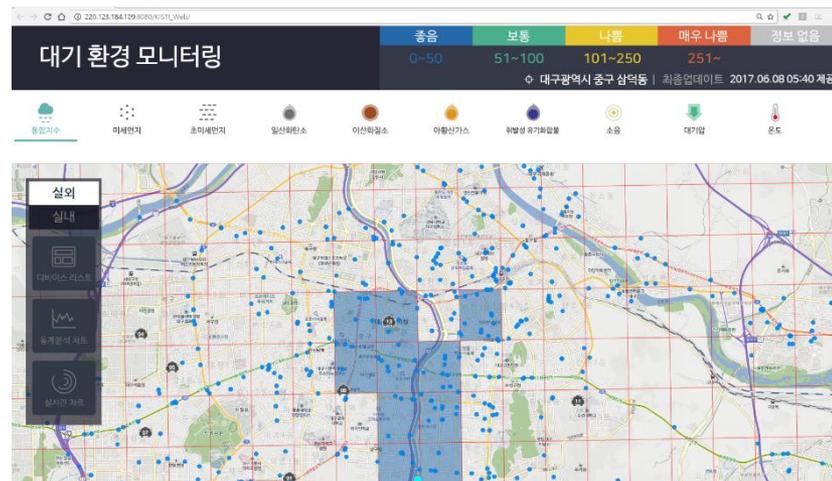
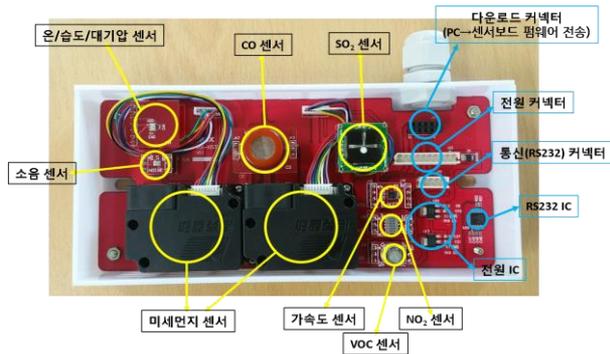
- 서로 각각 독립적으로 생성된 사물데이터 중 **동일한 위치와 동일한 시간에** 측정된 데이터 검색
 - 목적: 데이터 통합 분석, 데이터 결합, 데이터 동기화 등
- 사용 예
 - 여러 택시들이 이동하면서 측정한 데이터 중, 동일한 위치와 동일한 시간에 측정된 데이터 수집
 - 대기질, 날씨, 환경 정보 등
 - 어떤 도로를 따라 독립적으로 측정된 서로 다른 측정 데이터 결합
 - 교통량, 도로상태, 속도, 미세먼지 등



Preliminary Experiment (1/2)

■ 사용 데이터

- 대구시 대기질(air quality) 데이터
 - 택시(사물)들이 운행 중 10초 주기로 수집한 데이터 (2017.6 ~ 현재)
- 수집항목
 - 미세먼지, 온습도, 대기압, NO₂, SO₂, CO, VOC, 진동, 소음 등



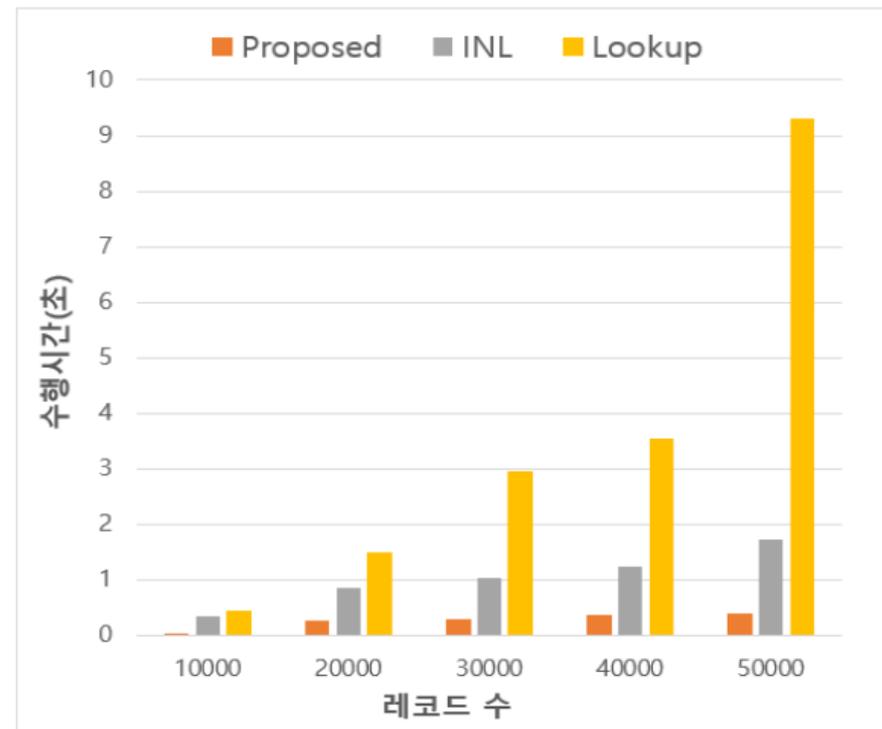
Preliminary Experiment (2/2)

■ 시공간 조인 질의

- 두 택시(택시#1과 택시#8) 데이터 간 시공간 조인 수행
- 매개 변수
 - 동일 위치 임계치 $\theta_s = 10m$
 - 동일 시간 임계치 $\theta_T = 1분$

■ 수행 시간 측정 결과

- INL (Index Nested Loop)
 - 기존 RDBMS 조인 방법
- Lookup
 - MongoDB의 \$lookup 기능 사용
- Proposed
 - 제안 방법

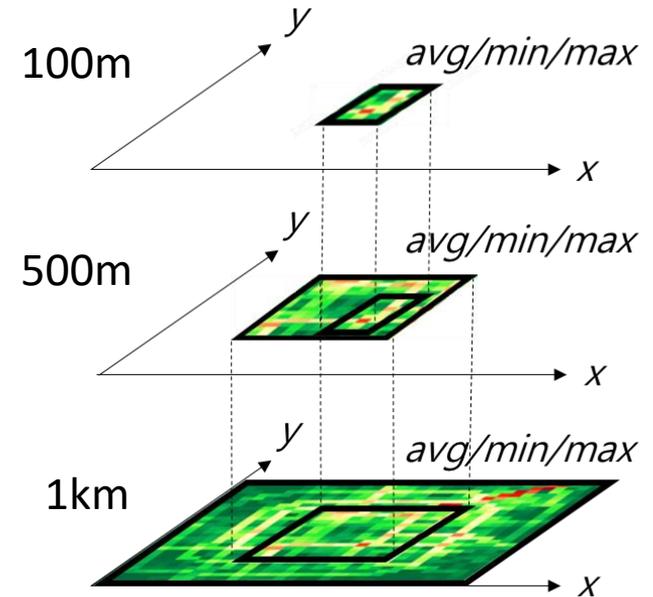


시공간 다수준 집계 알고리즘 개발

■ 시공간 다수준 집계 정의

– Input

- 집계 대상: $thing_1, thing_2, \dots, thing_N$
- 집계 함수: MAX, MIN, SUM, COUNT, ...
- 위치 범위: $[x_{min}, y_{min}], [x_{max}, y_{max}]$
- 시간 범위: $[t_{min}, t_{max}]$
- 위치 집계 수준: 1km, 100m, ...
- 시간 집계 수준: 1개월, 1일, 1시간, ...



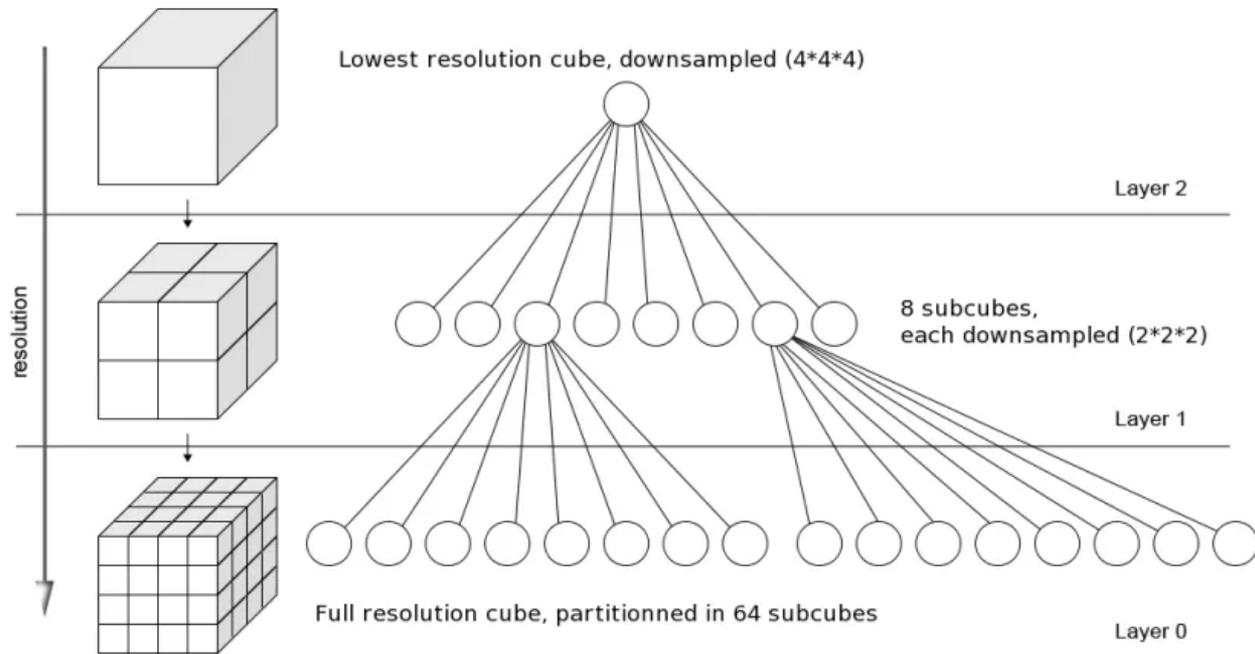
– Output

- 주어진 위치 및 시간 범위 내에서, 주어진 위치 및 시간 집계 수준으로 집계된 $thing_1, thing_2, \dots, thing_N$ 들의 데이터
- 단, 집계 항목은 사전에 지정되어 있다고 가정 (온도, 습도 등)

시공간 다수준 집계 알고리즘 개발

■ 핵심 idea

- (x, y, t) 3D 공간을 분할한 cell들을 계층 구조로 표현 (octree)



- 요청된 시간 및 공간 범위에 해당하는 node를 찾아 집계값 반환



- 과거의 IoT 기술은 주로 사물 간 통신 및 데이터 전송에 집중
 - 수많은 센서, 단말, 소프트웨어 간 원활한 연동 필요
 - 또한 이들의 상호 운용을 위한 기술적 표준 확립이 중요
- 하지만 연결된 단말 증가 및 데이터 생성 규모 증가만으로는 IoT 시장이 더 이상 크게 성장하기 어려움
 - 다음 단계로 필요한 것은 축적된 사물데이터의 구체적 활용 방법
- 그에 따라 지능형 IoT를 위한 **사물데이터 활용 기술**이 활발히 연구되고 있음
 - 사물데이터 활용의 기반이 되는 대규모 시공간 데이터 검색/질의 기술
 - 사물데이터에 대한 **데이터마이닝 및 기계학습** 기술



Thank you!

Any Question?

