Leveraging IoT Best Practices for Automotive

ASAM North America Conference
Big Data in Future Car Development
Novi, Michigan 26 October 2016

Mark Quinsland
mark.quinsland@datastax.com

Association for Standardisation of
Automation and Measuring Systems

About the presenter

DATASTAX

- Background: 20+ years with Emissions, Engine, and Time-Series data.
- Experience: 15+ years with ASAM ODS.
- Creator: ASAM ODS Web Services
- Instigator: ODS / Big Data workgroup
- Solutions Architect: DataStax (the Cassandra NoSQL company)
- Recent projects include: Designing an IoT Database to handle > 1 Trillion ops/day

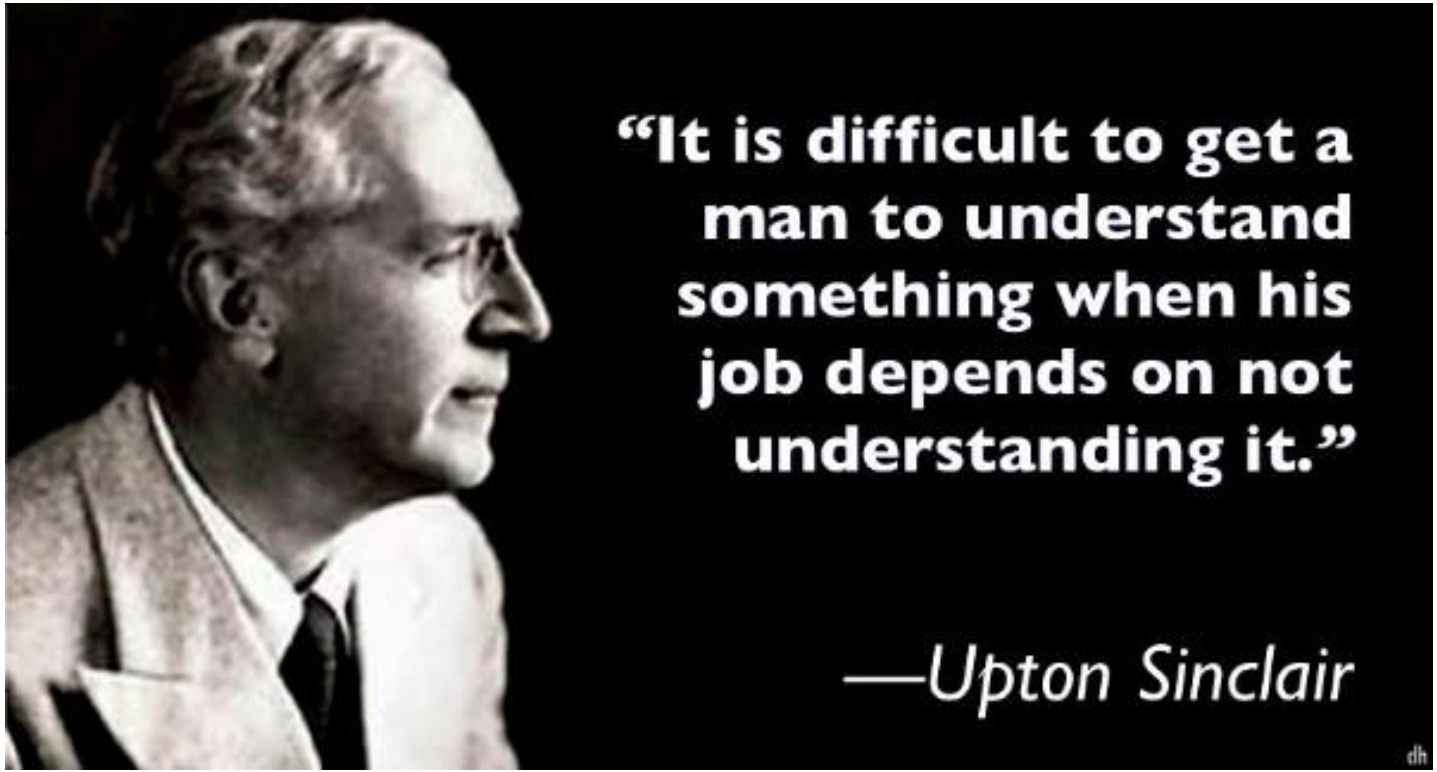
- mark.quinsland@datastax.com

Agenda

- Background
- Relevant examples from the Internet of Things
- Some Design patterns & tools that enable the IoT
- How can these ideas be exploited by ASAM ODS
- Next Steps

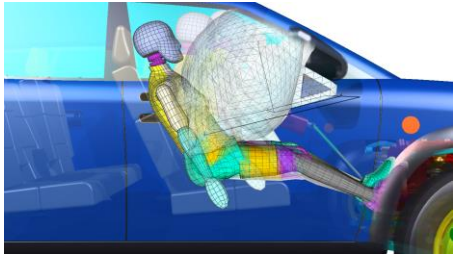
Source: http://www.techrepublic.com/blog/the-enterprise-cloud/the-cloud-is-fundamental-to-the-internet-of-things/

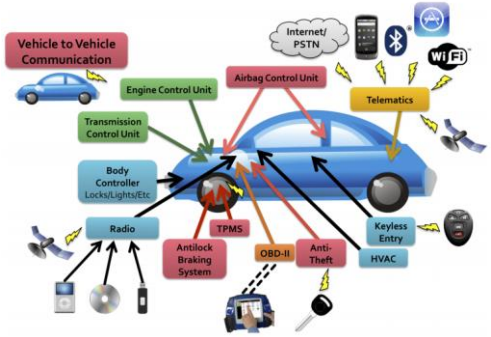
Big Data is a Very Disruptive Idea



http://quotesgram.com/upton-sinclair-quotes/

Evolution of Automobile Testing





- Road Testing
- Lab Testing
- Simulation
- Crowd Testing
- Improved Models

Source: http://www.tested.com/tech/concepts/46154-how-modern-cars-can-be-hacked/
http://www.mscsoftware.com/application/crash-safety
https://corporate.ford.com/content/dam/corporate/en/company/history/1955_Safety_Forum.jpg
http://www.eurocarnews.com/media/pictorials/555/2894.jpg

Where Can Big Data Help Automotive R&D

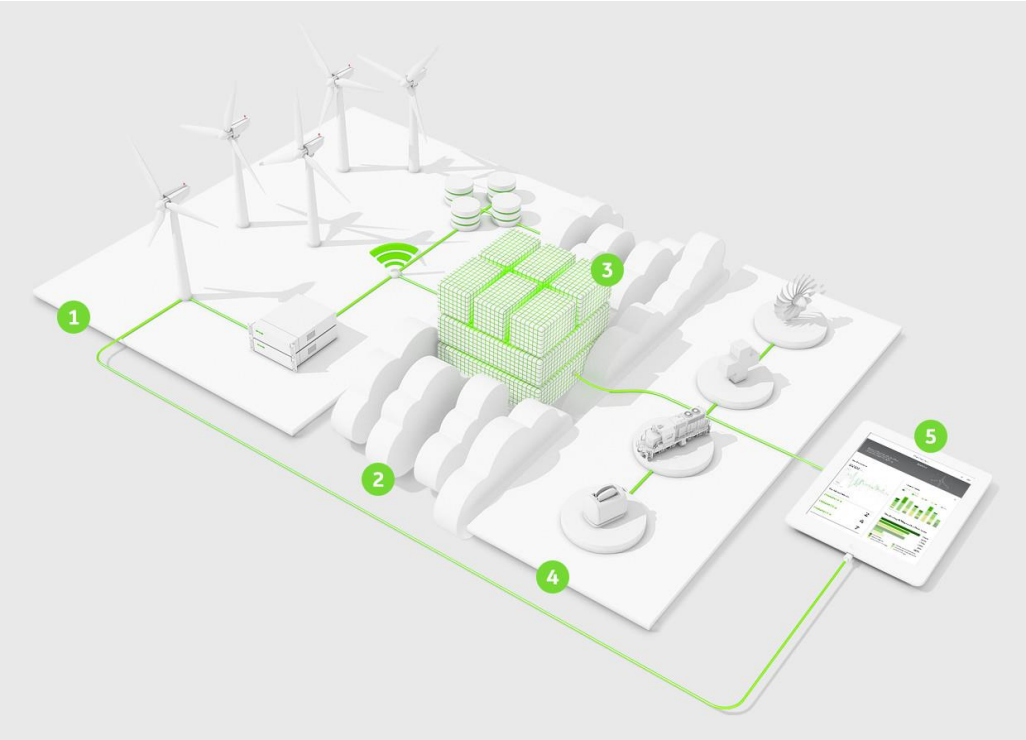
DATASTAX

Producers of Data

Consumers of Data

	Test Stands	Field Tests	Fleet / Consumer
Engineer	Raw Data in Local Files. Searchable meta-data. ASAM ODS	Small-scale streams of high-speed data.	Massive-Scale Streaming of low-speed data.
Test Labs		Raw log data (post-test) ASAM ODS	Real-time access to latest data points
Enterprise	Summarized. Large-scale comparisons.		Batch Analytics & Summarization

Relevant Examples From Successful IoT Companies



Source https://www.predix.com/how-it-works

Helpful: Willingness to Think Out-of-the-Box

- 10 TB data ingestion per day
- Doubling every 8 months
- Outgrew in-house solutions
- Switched to active-active Cassandra for 24*7 with 7 9s

Source: http://www.b2bnn.com/2015/09/how-microsoft-truly-uses-spark-and-cassandra-for-big-data-analytics/

Consumer IoT: Netflix

- > 1 Trillion ops per day
- > 50 Million customers
- Active-Active Cassandra for scalability & reliability
- Chaos Monkey kills active production servers to test reliability

Powered By Cassandra

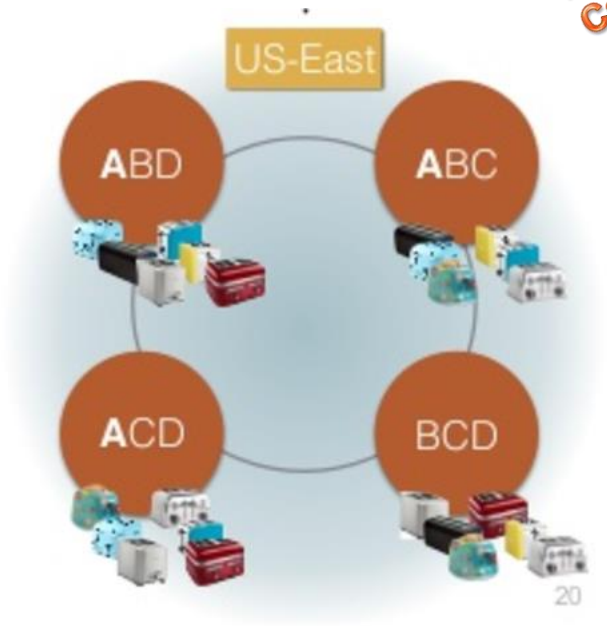
Source: http://www.slideshare.net/PhilipFisherOgden/netflix-viewing-data-architecture-evolution-qcon-2014
and: http://techblog.netflix.com/search/label/Cassandra

100% Availability + Massive Scalability

DATASTAX

NETFLIX

Powered By Cassandra



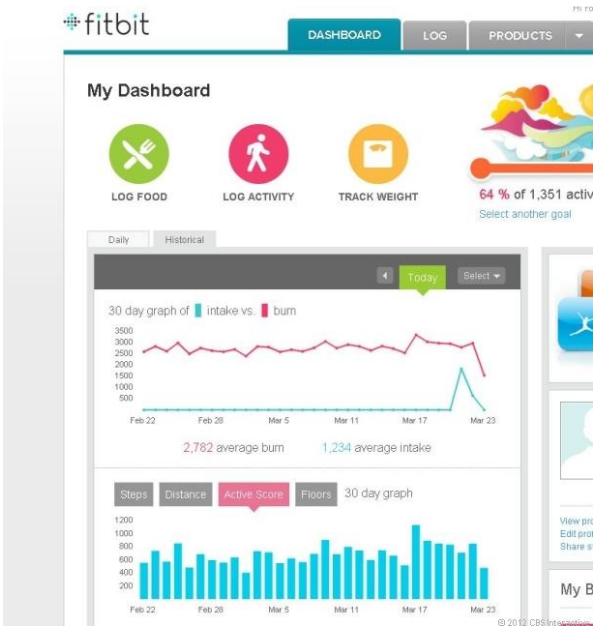
Source: http://www.slideshare.net/PhilipFisherOgden/netflix-viewing-data-architecture-evolution-qcon-2014
and: http://techblog.netflix.com/search/label/Cassandra

Personal IoT Devices – FitBit.com

DATASTAX

"Data is the core of Fitbit's business. We have millions of users' devices syncing with our site regularly throughout the day. We are looking to move towards an active-active multi-datacenter strategy... The majority of our data is currently in MySQL which doesn't meet a lot of these needs. We are looking for a highly skilled Cassandra engineer to lead us in that transition…."

Actual Job Posting: 20-Oct-2016



Powered By Cassandra

Source: https://www.glassdoor.com/job-listing/cassandra-engineer-fitbit-JV_IC1147401_KO0,18_KE19,25.htm/

Industrial Internet: GE - Predix

DATASTAX

Disclaimer: Presenter spent 5 months re-designing & testing the GE Predix Time-Series database architecture

Predix

Powered By Cassandra

- Global IoT-as-a-Service to industrial customers worldwide
- Scalable to billions of devices @ 1hz to 1khz
- Active-Active Cassandra for 100% uptime
- Graph Database for meta-data
- Global / Regional data replication
- > 4 Trillion Ops/day by 2020

Source: http://www.ge.com/digital/press-releases/GE-to-Open-Up-Predix-Industrial-Internet-Platform-to-All-Users

Industrial Internet: GE Aviation

GE Aviation

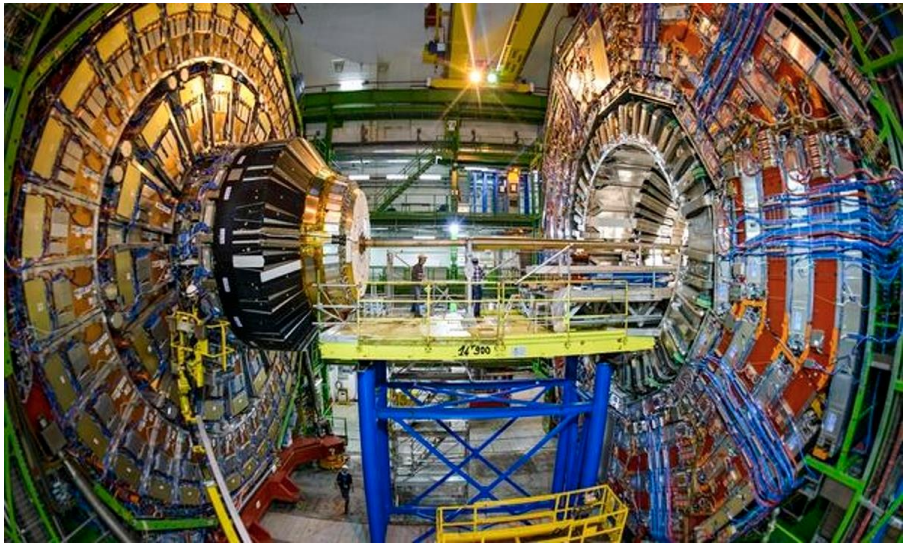
Powered By Cassandra

- Real-time and batch analysis of 30k jet engines
- Spark Streaming apps
- Spark Batch Analytics
- Active-Active Cassandra for 24/7 * 100%
- Graph Database for meta-data

Source: http://www.ge.com/digital/press-releases/GE-to-Open-Up-Predix-Industrial-Internet-Platform-to-All-Users

Test Data @ Massive Scale: Large Hadron Collider

DATASTAX

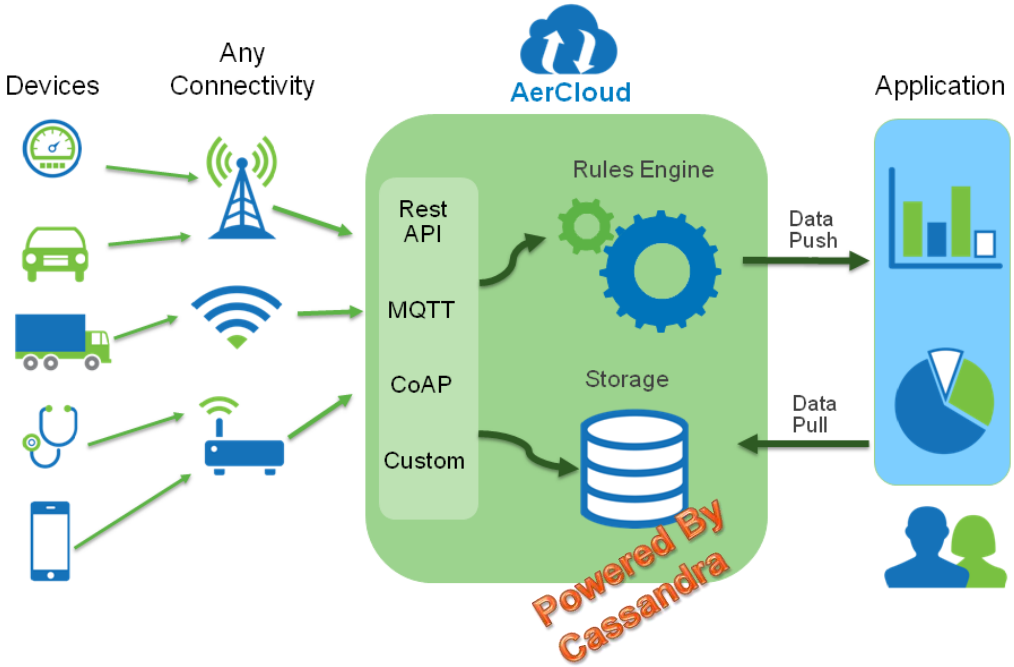


- Data acquisition system (TDAQ) supports 30k applications and 2k connected servers.
- Relies upon Cassandra's horizontal scalability for massive data ingestion rates without data loss.

Powered By Cassandra

Source: http://indico.cern.ch/event/93877/contributions/2118095/attachments/1104220/1575372/PBEAST_ACAT_Poster_v5.pdf

Telematics Example: Aeris



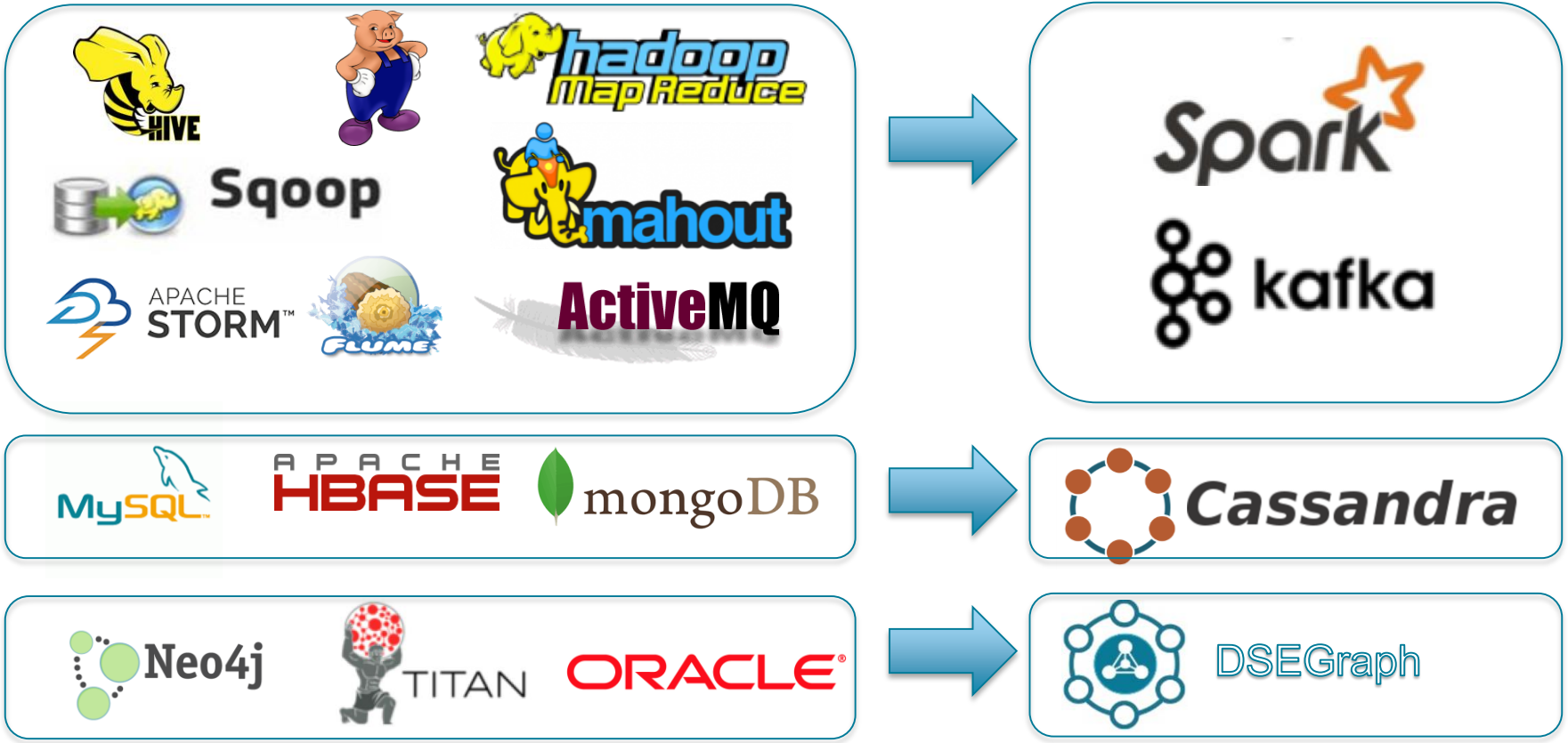
aeris

- Acura & Hyundai
- 500k Trucks on Aeris Network
- Horizontal Scalability for Millions of Vehicles & Billions of devices
- Active-Active Cassandra

Source: http://www.aeris.com/technology/aercloud/

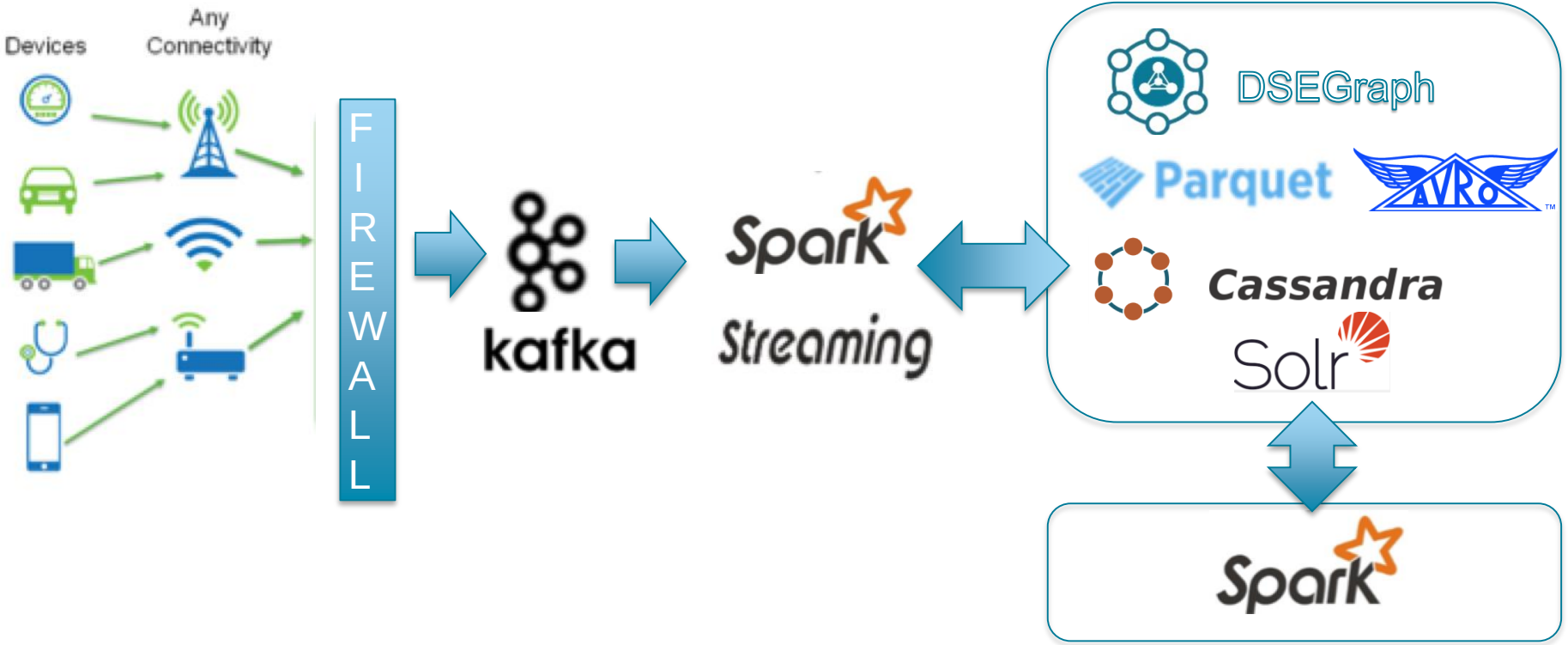
Evolution of Tool Sets in NoSQL 2.0

DATASTAX



Source https://www.predix.com/how-it-works

Common IoT / Time-Series Tech Stack



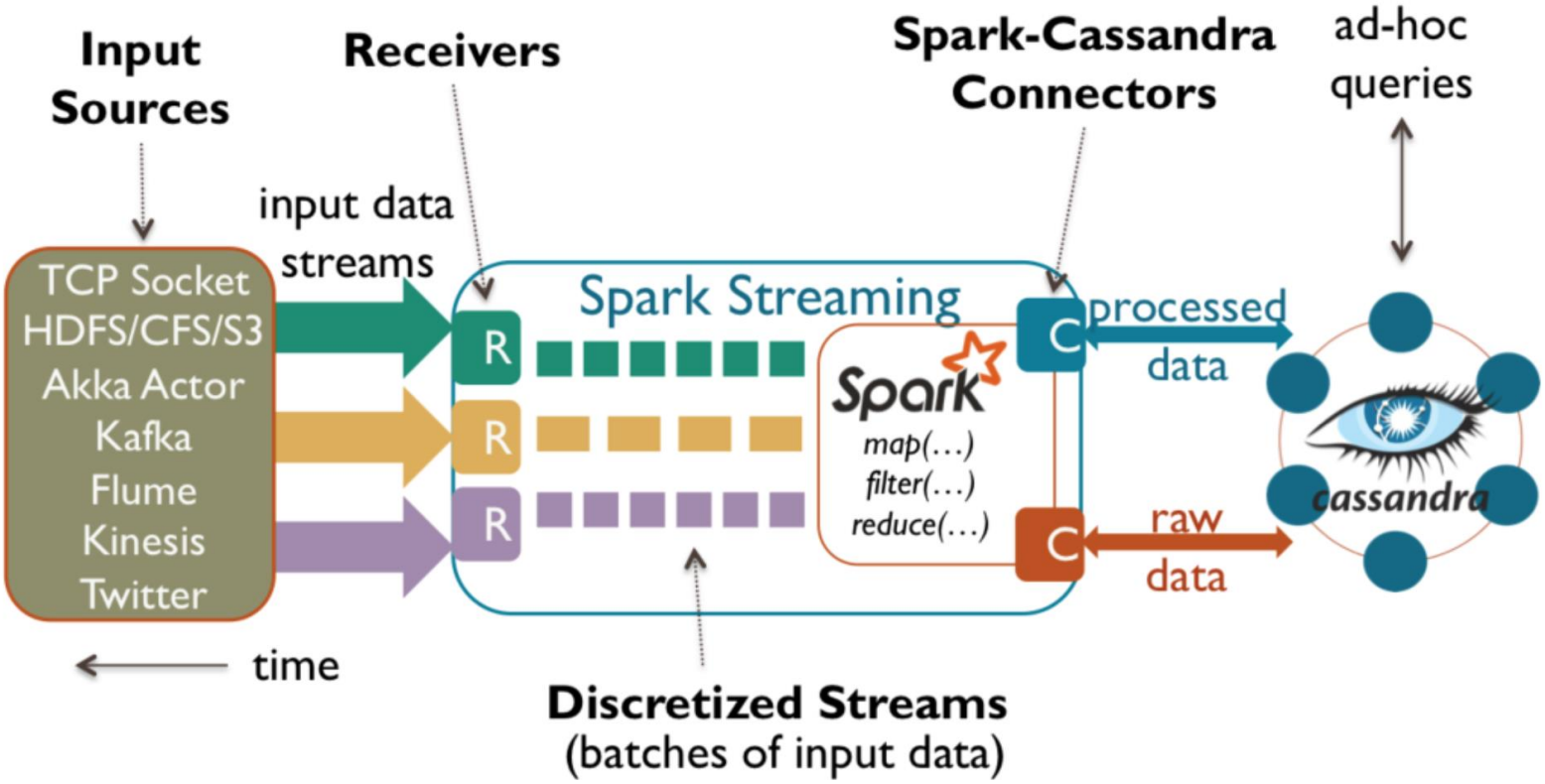
Source https://www.predix.com/how-it-works

Spark vs Spark Streaming

zillions of bytes

gigabytes per second

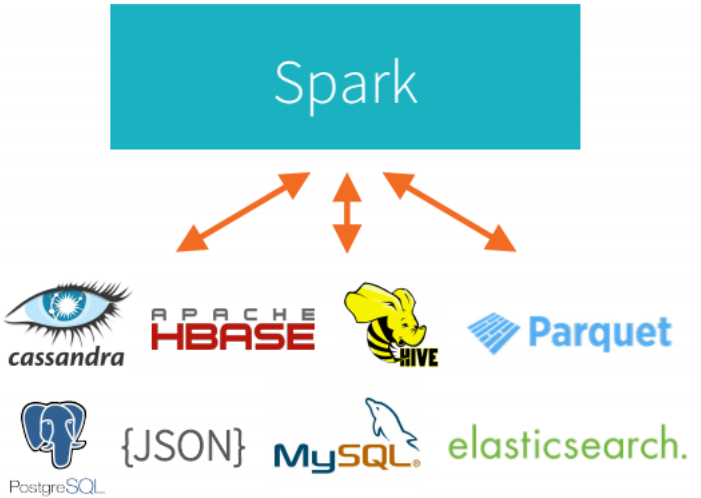
Spark Streaming



Source: http://www.slideshare.net/helenaedelson/streaming-bigdata-helenawebinarv3

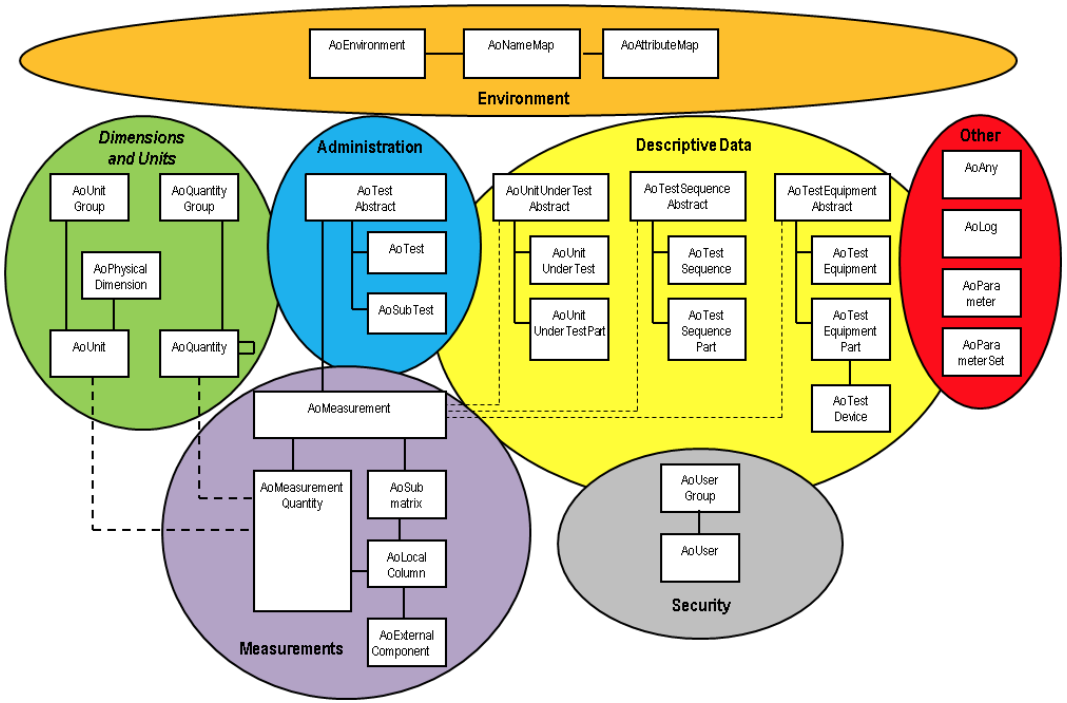
Spark External Data Access

- Data Plugins allow extensibility with
- Multiple sources can be accessed using SQL-92 syntax
- Simple API for Connectors
- Why not an ODS/Spark Connector?



Source: http://www.slideshare.net/helenaedelson/streaming-bigdata-helenawebinarv3

Implications for ASAM ODS?



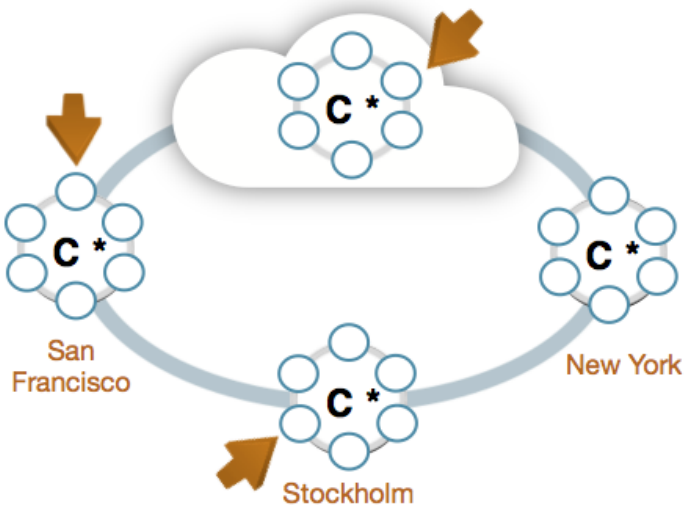
Source: http//asam.net

‘OPEN’ Data Services?

DATASTAX

- ODS – Idea of the 90’s based upon technology from the 70’s
- API should be Logical – not physical
- Base Model:
  - Relatively inflexible, and data types are limited (based on limitations of RDBMS) e.g. Maps, JSON, UDTs
  - Optimized for storage efficiency – not for retrieval or analytics

Cassandra – *The* Database for the Cloud



Cassandra is **DESIGNED** to solve…

- High Performance, High Scalability tasks

Cassandra is **DIFFERENT** because…

- Native fully distributed architecture
- "Primary" and "Master" DO NOT EXIST

Cassandra can then **DELIVER**…

- Global, always-on, linear-scale database
- Multi-Datacenter and Cloud exploitation
- Locality-aware for privacy & regulations

Thanks!

mark.quinsland@datastax.com