Edge Analytics: Use Cases for Industry 4.0

Theo Ruland
Data and Analytics Group, Cisco
April 19th 2016
Agenda

Data & Analytics Portfolio

Edge Analytics

Use Cases

Demo
The Digital Economy is Disrupting Everything
Cisco’s Data & Analytics Portfolio

Data Integration & Management
- Data Virtualization
- Data Preparation

Edge Analytics
- Connected Streaming Analytics
- ParStream

Analytics to Support Cisco Architectures
- Network Deployment Analytics
- Network Compliance Analytics & Management
- Collaboration Analytics
- Contact Center Analytics
- Data Center Analytics

Analytics to Support Vertical Industries
- Service Provider Analytics
- Site Analytics

Process Integration & Workflow Automation
- ServiceGrid
- Workload Automation

Energy & Asset Optimization
- Energy Analytics & Management
- Asset Management
Edge Analytics
A New Approach is Needed to Reach and Analyze That Data

Structured Data

Traditional Data Warehouse

Widely Distributed, Streaming, Short Shelf Life, Too Big to Move

Unstructured Data

Big Data Store

Data Streaming at the Edge

“Most data will be processed at the edge” (mobile devices, appliances, routers)

37%

SHORT-TERM  LONG-TERM  REAL-TIME

Three years from now, where will most data generated by IoT solutions be processed?
Cisco Streaming Analytics - CSA

CSA is an embeddable, horizontal and hierarchically scalable distributed in-memory streaming database designed and built for analytics at the edge:

- CSA runs in routers, switches, small compute and commercial servers
- The streaming engine enables multi-stream, concurrent processing of “raw streams” and the generation of multiple “derived streams” – from simple aggregates to advanced machine learning algorithms
- Interfaces allow ANSI SQL northbound for applications, pluggable framework analytics models and data source connectors

CSA applies predicates, aggregations, and joins with metadata tables, contextual data to identify anomalies and detect trends:

- Generate Events
- In Memory Processing, with an option to commit to integrated storage
- Integrated Querybase
  Seamless Query over Live and Stored Data
- Visualize
  Analyze, Export

Raw Streams

Generate Derived Streams
Cisco ParStream

Cisco ParStream is a distributed massively parallel processing columnar database designed and built for the Internet of Things.

Its small footprint enables implementation at network edges and inside the network, as well as in data centers. Its innovative compression and bit mapped indexing technologies enable Cisco ParStream to meet IoT historian and buffering requirements—such as providing sub-second response times on billions of data records and thousands of columns, while continuously importing new data.

- Continuously import high-velocity data with low latency
- Analyze and filter billions of records
- Query data structures with thousands of columns
- Get answers in milliseconds, without Cubes
- Execute thousands of concurrent queries
ParStream Responds to Distinct Requirements for IoT

**DATA GROWTH**
Data is growing faster and bigger because of number of sensors

10B+ rows 5TB+

**FAST DATA**
Data streamed from sensors requires fast ingestion

1M+ rows per sec

**EDGE ANALYTICS**
IoT data is mostly generated at the ‘Edges’ of the network

1000+ locations

**REAL-TIME INSIGHTS**
Use cases require near Real Time Analytics

Millisecond query response time
Choose Your Database Based on Your Use-case

- **High Response Time**
  - **Stream-Analytics**
    - Complex Event Processing
    - In-Memory DB
  - **Real-Time IoT Analytics**
    - Cisco ParStream
    - Massively Parallel (MPP) Real-Time

- **Low Response Time**
  - **Operations Analytics**
    - OLTP Reporting
  - **Batch-Analytics**
    - Cassandra
    - Hadoop

- **Gigabyte**
  - OLAP
- **Terabyte**
  - OLAP
- **Petabyte**
  - OLAP

**Big Data**

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Confidential
ParStream’s Patented Technology Provides a Competitive Advantage

1. **High Performance Compressed Indexes**
   Provide ultra-high query performance

2. **Massive Parallel Processing**
   Delivers linear scalability and high query throughput

3. **Lockless Architecture**
   Enables ultra-fast query and data import performance

4. **Small Footprint**
   Enables analytics at the edge with a low TCO

- SQL API / JDBC / ODBC
- C++ UDx API
- Immediate Queries
- In-Memory and Disk Technology
- Multi-Dimensional Partitioning
- Massively Parallel Processing (MPP)
- Shared Nothing Architecture
- 3rd Generation Columnar Storage
- High Performance Compressed Index (HPCI)
- High Speed Parallel Loader with Low Latency
Delivering actionable insights requires use-case specific applications & data integration

M2M pattern – siloed vertical solutions

Big Data pattern – horizontal integration

Network Analytics
POS Analytics

No Data Integration
Individual Solutions
Dedicated Ressources

POS
POS
POS

4 billion records
4 billion records

Client
Network Analytics
POS Analytics
Fraud Analytics
?

Cloud / Data Center

Network

Source

POS
POS
POS
POS
Delivering actionable insights from IoT data requires big + fast data, real-time analytics and decentralization.

**Big Data pattern – horizontal integration**

- **Client**: Network Analytics, POS Analytics, Fraud Analytics, ?
- **Cloud / Data Center**: Shared WAN
- **Network**: 50 billion devices

**IoT pattern – de-centralization at scale**

- **Client**: Network Analytics, POS Analytics, Fraud Analytics, ?
- **Cloud / Data Center**: Shared WAN
- **Fog**: Edge
- **Source**: POS, POS, POS

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Confidential
Fog Computing Delivers Real-time Insights by Minimizing Network Traffic

Overcoming Bandwidth Limitations and Reducing Report Delays Requires Analytics to be Pushed Closer to the Data Source

Central Analytics Intelligence

Application

Query

Search Results

40 Records Found

ParStream Geo-Distributed Server

Database

ParStream

Hybrid Edge / Centralized Intelligence

Application

Query

Search Results

40 Records Found

ParStream Geo-Distributed Server

<100 Records

4 Billion Records

4 Billion Records

4 Billion Records

4 Billion Records

4 Billion Records

7 Records

18 Records

5 Records

12 Records

8 Records
Analytics At the Edge Architecture

Streaming Data
- Infrastructure Events & Logs
- Network Events & Logs
- Application Events & Logs
- Wi-Fi/Mobile Events & Logs
- Sensors/MT Connect
- ESB’s
- Flume/Thrift/Kafka, etc.

Connected Streaming Analytics Trulink
- Streaming Connector Framework
- Handler

Real-time Enrichment & Sessionization
- Handler

Services Framework APIs
- DV

ParStream
- Fast inserts and Edge Storage for Raw Data & Historical Aggregates

Connected Streaming Analytics TruCQ
- Real Time Analytics
- ANSI SQL on Streams
- In-Memory Models
- Shared Stream Query Processor

Embedded Postgres

Analytics & Reporting Apps
- Real-Time Dashboards
- Report Builder
- Visualization Framework
- Real-Time Service API

- Cloud Services
- Web Apps
- Mobile Apps
- Monitoring & Alerts
- Action Engines

Raw Data

- Data Actuation & Enrichment
- Control & Action Engines
- Contextual & Semantic Engines

- Systems & Databases
- Command & Control
- Infrastructure Management
- Services Management
- Business Systems

APIs

Handler

Handler

Handler

Handler
Preparing for the “Hyper-distributed” World

Many times you want to store and analyze the data locally before sending northward.
Cisco’s Plans

Cisco’s Plans is to Dominate the IoT Market
• Include ParStream as a critical component in our IoT framework
• Use ParStream in several existing and new IoT customer projects
• Use ParStream to help forge new IoT partnerships

Cisco Does Not Plan to Enter the Database Market
• Continue to sell ParStream opportunistically outside IoT
• Do not promote ParStream as standalone database
Use Cases / Case Studies
IoT Analytics at the Edge
IoT Enterprise Use Case: Oil & Gas

Edge Location
- Sensors & Edge Processing

Processing Platforms
- Fog Processing

Regional
- Data Center

Headquarters
- IT Central

Edge Analytics with CSA
- Deployed on Cisco Devices

Data Platform and Applications
- Deployed in Customer Data Center or Regional Center
Customer Proof Point for IoT in Renewable Energy:
Real-time Analytics for Wind Turbines

Business Challenge
• Optimize wind turbine performance by quickly adjusting to changing environmental factors (e.g., wind direction, temperature, etc.)
• Minimize turbine downtime thru predictive maintenance.

Use Case
• Real-time and continuous monitoring of data from 20,000 wind turbines (each with 150 sensors), including analysis of 30TB of historical data

ParStream’s Technology Value Proposition
• Real-time monitoring of continuous data-flow for immediate insights/actions
• Historical analysis thru enabling storage and analytics in an integrated platform by immediately importing and storing readings from turbines.

Benefits/Results (estimated)
• 15% improvement in productivity
• Decreased downtime
• $158M of annual economic benefits
How it works with ParStream - Technical Setup

- The devices at the windfarms stream aggregated data into the cloud
- Aggregated data is stored in the cloud and feeds software programs
- Granular data is kept at the windfarm for inspection and advanced analytics
- Previously data was uploaded daily, with ParStream upload every minute
- Data about weather, locations, prices etc. adds context and meaning to turbine data
Mazak SmartBox
Industry-Shaping Collaboration

In October 2015, we announced an industry-shaping collaboration with Mazak Corporation and MEMEX on “SmartBox,” a technology that enables real-time data and analytics from manufacturing machines.

About SmartBox
• The SmartBox is a mini electrical cabinet about the size of a typical household medicine chest
• It can be mounted on the side of a machine enclosure; this enables the box to be connected to a machine tool in several ways

Cisco Solution
• Cisco’s real-time analytics capabilities are embedded into a Cisco Industrial Ethernet (IE) 4000 switch inside SmartBox; measuring vibrations and temperature
• It provides real-time visibility and insights into data right on the factory floor

Expected Business Results
• Access to real-time manufacturing data to improve overall productivity and responsiveness to customer / market changes
• Identify and easily fix downtime-related inefficiencies to improve overall equipment utilization
Process Manufacturing: Machine Protect

Use Case
Operator alert when machine parameters cross normal threshold
Real time machine stats with baseline recommendations
Trends that indicate potential machine failure
Mazak Edge Analytics Architecture

Connected Streaming Analytics
- Model Configuration API
- Streaming Model Scoring
- Anomaly Alert & Data Export

MTC Handler

MTConnect Agents
- MC1
- MC2
- MC3
- MC4

Cisco IE4k Switch

JDBC

ParStream

CSA
- Data Ingest
- Alert Stream Mediation
- Web UI

Fog Node

HCN8800

E1060

HCN8800

Orbitec

Cisco IE4k Switch

ParStream
Streaming Analytics on Mazak’s Smart Box
Business Results

- Identify and easily fix machine downtime
- Increase equipment utilization
- Improve productivity