Innovazione nell'analisi dei dati
8 - 15 maggio

Coordinatore: F. Stanco
DMI - Università degli Studi di Catania

M. Marroccia, G. Ursino, G. Scuderi, F. Milotta
STMicroelectronics
1. Intro → ST and Digital Transformation
2. Designing an IoT Reference Architecture → Focus on Data
3. Data Scientist’s Tasks → Focus on IoT
4. Designing an IoT Reference Architecture → Deepening
5. Demo Session
from past session...
VUCA

- Volatile
- Uncertain
- Complex
- Ambiguous

Do you know?

- Top 10 In-Demand Jobs in 2018
  - Did not exist in 2008

- 1984: 1000 @ devices
- 2020: 75,000,000,000 @ devices

- The « PRESENT » has never been SO SHORT

- 5,500,000,000 searches on
- On Google per day
- 400 hours of video per minute to YouTube
- 692,000,000 Tweets per day
- 54,000,000,000 What’s app per day

- 2,200,000,000 Facebook Users
- 1,100,000,000 Instagram Users
- 33 zettabytes of new data created in 2018

- DISRUPTION is new « Normal »
- BUT
- CONSTRUCTION is ESSENTIAL

We are preparing STUDENTS
- For jobs that do not yet exist...
- For technologies that have not been invented...

— Imagination is more important than knowledge.” — A. Einstein
Where will we be in 5 years?
Innovazione nell'analisi dei dati

IoT Architectures

venerdì 15 maggio - 09.00-12.00

Coordinatore: F. Stanco
DMI - Università degli Studi di Catania

M. Marroccia, G. Ursino, G. Scuderi, F. Milotta
STMicroelectronics
ST and DIGITAL TRANSFORMATION

the role of Information Technology
DIGITAL TRANSFORMATION @ ST
Leveraging Digital Technologies

SMAC²

- Social
- Mobile
- Analytics
- Cloud & Cybersecurity
DIGITAL TRANSFORMATION @ ST
Fostering a data-driven decision process
DIGITAL TRANSFORMATION @ ST
Cloud Computing: a Strategy for the Future
DIGITAL TRANSFORMATION @ ST
Technologies, architectures, jobs for the future

Cloud
Enterprise Architect

Cloud/Integration Architect

Business Intelligence

IOT

Power BI

Lambda Architecture

Advanced Analytics

Data Scientist

Artificial Intelligence

Machine Learning

SmAc²

Microsoft Azure

Amazon Web Services

HYBRID Architecture

Elasticity vs. Scalability

SMAC²

Data Scientist

Smart Driving

Smart Industry

Smart Things

Industry 4.0

Enterprise computing

Cloud

iPaaS

Multi cloud
DIGITAL FAB @ ST
Our AGILE Digital Incubator

- Opening: Dec 2018
- CEO Roadshow
- A collaborative workplace
- Electronica Fair Munich, 2018
- Predictive Maintenance
- Papers@International Conferences
Designing an IoT Reference Architecture ➔ Focus on Data

Giuseppe Ursino
STMicroelectronics
Internet of Things (IoT) adds value in three major areas:

• increasing efficiency,
• improving health/safety
• creating better experiences.

The Industrial Internet of Things deals with the first two areas, increasing efficiency and improving health/safety.

IIoT refers to a subcategory of the broader Internet of Things. IoT includes IIoT plus things like asset tracking, remote monitoring, wearables, and more.

IIoT focuses specifically on industrial applications such as manufacturing or agriculture.

In recent years, innovations in hardware, connectivity, big data analytics, and machine-learning thanks also to Cloud Computing (so Internet Technologies) have converged to generate huge opportunities for industries. Hardware innovations mean that sensors are cheaper, more powerful, and run longer on battery life. Connectivity innovations mean that it’s cheaper and easier to send the data from these sensors to the cloud.

Big data analytics and machine learning innovations mean that, once sensor data is collected, it’s possible to gain incredible insight into manufacturing processes.

These insights can lead to massive increases in productivity and drastic reductions in cost. Whatever is being manufactured, it can be done faster, with fewer resources, and at a lower cost.

ANY SUFFICIENTLY ADVANCED TECHNOLOGY IS INDISTINGUISHABLE FROM MAGIC.”

Arthur C. Clarke, Profiles of the Future
Every IoT project needs a **reference architecture** that defines what functionality is required, where that functionality will operate, and how data and control will flow.

- **Layers**, tiers and interfaces are the fundamental building blocks of an Internet of Things (IoT) architecture.

- **Layers define what** capabilities an IoT component, function or process must possess. The five key layers are device, communication, **Data**, function and process.

- **Tiers define where** to deploy a component, function or process. The three logical deployment tiers are edge, **platform** and enterprise.

- **Interfaces** define how data and control flow into, out of and through the system. The two types of interfaces are cross-tier and cross-layer.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Function Layer</th>
<th>Data Layer</th>
<th>Communication Layer</th>
<th>Device Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Layer</td>
<td>Governance</td>
<td>Data Architecture</td>
<td>Event-Driven Architectures</td>
<td>Sensors</td>
</tr>
<tr>
<td></td>
<td>Operations</td>
<td>Data Models</td>
<td>Network Technology</td>
<td>Things/ Devices</td>
</tr>
<tr>
<td></td>
<td>Management</td>
<td>Data Storage</td>
<td>Communication Service Providers</td>
<td>Aggregations/ Gateways</td>
</tr>
<tr>
<td></td>
<td>Business Applications</td>
<td>Metadata</td>
<td></td>
<td>Digital Twin</td>
</tr>
<tr>
<td></td>
<td>Layer Interfaces</td>
<td>Layer Interfaces</td>
<td></td>
<td>Layer Interfaces</td>
</tr>
<tr>
<td></td>
<td>Security</td>
<td>Security</td>
<td>Security</td>
<td>Security</td>
</tr>
</tbody>
</table>

**Diagram:**

- **Edge Tier**
- **Platform Tier**
- **Enterprise Tier**
Logical Flow of Data in an IoT Architecture

Figure shows Data originates at endpoints, and as it moves toward the enterprise IoT platform it is normalized and enriched, increasing its value. If prediction is the objective, data is the raw material that enables that outcome. Data is an important complement to prediction. As the cost of prediction falls, the value of quality data goes up. It is important to understand how the data flows first, before trying to figure out how to physically accomplish one’s IoT data journey.
Three Tiers of An IoT Architecture

Focus on Physical Data Flow

**Edge Tier**
- Sensors and Actuators
- IoT Device
- Local Network
- HW/Local Connection
- Long-Range Communication
- Events
- Commands

**Platform Tier**
- IoT Gateway
- Stream Processing
- Event Processing
- Dispatch and Orchestration
- Data Analytics
- Device Gateway Service
- Data Warehouse
- Edge Device Authentication

**Enterprise Tier**
- Application Services
- Data Analytics
- Business System
- Workflow Process(es)
- API/Integration Gateway
- Command
- Commands

**IoT Physical Data Flow**

**IoT Edge**
- Sense
- Sensor Data
- Control Commands
- Long-Range Communication
- Edge Gateway
- Control and Manage

**Central Platform**
- IoT Data and Events
- Control and Manage
- Enterprise Systems
- Enterprise Integration

**Enterprise Systems**
- Command
- Commands
The Edge tier is where data is sampled and collected from the environment by instrumented “things” or devices. These “things” may include consumer devices (such as thermostats), appliances or industrial systems (such as central air conditioning systems) that contain sensors to collect data, or configurable parts (such as actuators) to alter the operation of the device. The edge tier may also contain optional IoT gateways that can provide localized data analysis, event processing and storage, as well as help integrate legacy devices to IoT platforms that use protocols such as MQTT and WebSockets.

The Platform tier is where the IoT system aggregates systemwide data and events from many edge locations. It will often perform stream processing and event processing for many edge locations (see The Function Layer section). It will also orchestrate tasks or invoke enterprise applications. The IoT platform also contains device and platform management functionality. The platform may be either a single monolithic platform or a composite of complementary capabilities from one or more providers.

The Enterprise tier is where IoT integrates with the set of applications, processes and services required to accomplish a business objective. For example, applications such as inventory management, enterprise resource planning, product quality and reliability, outlier detection, action recommendation, and customer relationship management often reside in the enterprise tier. These applications may benefit from the data-driven insights provided by an IoT system. Many IoT platforms include APIs that enterprise applications use to extract data and events from the platform for their own purposes.
A Predictive Maintenance Architecture

- **Device**
  - Ethernet/IP
  - OPC UA
  - OPC UA Server
  - Industrial IoT Gateway
  - Manufacturing Digital Twin
  - To/From APIs

- **Comm.**
  - Plant data storage
  - Plant metadata
  - Plant sensor data
  - Data and event models
  - Data storage
  - Metadata
  - Data read/write models and metadata

- **Function**
  - Command/response
  - Stream processing
  - Event processing and policy
  - Predictive maintenance
  - APIs

- **Data**
  - Edge stream processing
  - Stream processing
  - Event processing and policy
  - Predictive maintenance
  - APIs

- **Process**
  - To/From user
  - Plant operations system
  - IoT management
  - To/From layers and tiers
  - RIMS = Repair Inventory Management System
  - MPLS = Multiprotocol Label Switching
  - CSP = Communications Service Provider

- **Edge Tier**
  - Ethernet/IP
  - Communication agent
  - MQTT Pub/Sub
  - MPLS CSP

- **Platform Tier**
  - Manufacturing
  - OPC UA server
  - X.509
  - To/From APIs
  - Dispatch code
  - Predictive maintenance algorithm

- **Enterprise Tier**
  - To/From function layer code
  - To/From function layer code
  - To/From function layer code
  - To/From function layer code

- **RIMS** = Repair Inventory Management System

- **Access Control**
IoT Platform typical Architecture
Key Applications

- Smart manufacturing
- Factory automation
- Functional safety and security
- Condition monitoring and predictive maintenance
- Smart motion/motor control
- 3D printing
- Power & energy management
- Industrial robots
- Industrial lighting
- Sensors for industrial, medical, aerospace & defense

Industry Dynamics

- Smart Industry initiatives (Industry 4.0, IIoT, …)
- Integrated distributed manufacturing
- Flexible, reconfigurable factories
- Optimization of factory infrastructure life cycle
- Cloud-based condition monitoring & predictive maintenance

Key Trends

- Next levels of automation with distributed control
- Safer working environments & new man-machine interaction models
- Higher energy efficiency for industrial machinery
- Capture & exploitation of manufacturing data
- Artificial Intelligence & machine learning
More than 30 years **experience** in developing products for industrial applications

Secure **supply** chain with in-house manufacturing

Highest **quality** standards thanks to automotive experience

Deep **application knowledge** enabling us to deliver products optimized for the application as well as system solutions

The broadest product offer based on industry-leading innovative **technologies**

Comprehensive tools, SW and support
What are the enablers for Smart Industry?

**More efficient**

- Higher efficiency at all points in power usage
  - Power conversion & energy harvesting
  - Power Management
  - Power storage
  - Motor Control

**More Intelligent & Aware**

- Sensors collect information about every machine and distributed local processing allows data to be turned into information
- Safe & Secure real-time processing
- Products contain the instructions for their manufacturing
- Machines are aware of the humans around them and provide easier and safer interactions

**More Connected**

- Machines are connected inside the factory, to the larger supply chain and to the cloud
- Real-time communication down to the lowest level (sensor & actuator)
- All communications must be secure
How does ST enable this?

- Power Discrete & Modules
- Smart power
- Microcontrollers & Secure solutions
- Connectivity & Communication
- Sensors & actuators
- Analog & Signal conditioning

- Smart manufacturing
- Factory automation
- Smart Motion control
- Industrial robots
- Industrial lighting
- Power Management
Enterprises Internal IoT Activities’ Benefits
Percentage of Respondents (Up to 3 Responses Allowed per Participant)

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process improvement — business processes</td>
<td>37%</td>
</tr>
<tr>
<td>Process improvement — manufacturing processes</td>
<td>36%</td>
</tr>
<tr>
<td>Supply chain visibility, control, coordination, integration</td>
<td>34%</td>
</tr>
<tr>
<td>Remote monitoring and control of operations</td>
<td>28%</td>
</tr>
<tr>
<td>Asset monitoring or optimization (e.g., utilization, maintenance, etc.)</td>
<td>27%</td>
</tr>
<tr>
<td>Workforce productivity enhancement</td>
<td>26%</td>
</tr>
<tr>
<td>Conserving resources (e.g., energy, water, fuel, raw materials)</td>
<td>25%</td>
</tr>
<tr>
<td>Enhanced worker or visitor safety/health</td>
<td>20%</td>
</tr>
<tr>
<td>Address regulatory or compliance control need</td>
<td>20%</td>
</tr>
<tr>
<td>Enhanced risk assessment for development of SLAs and warranties</td>
<td>18%</td>
</tr>
<tr>
<td>Other</td>
<td>0.1%</td>
</tr>
<tr>
<td>No benefits generated by internal IoT activities</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

n = 1,290

Base: Exclusively invested/primarily invested/equally invested in both internal and external initiatives (CL2), excluding not sure.
CL3. What do you think would be the most significant benefits generated by your organization’s internal IoT activities?
Note: Multiple responses allowed (up to three responses).
Data Scientist’s Tasks → Focus on IoT

Filippo Milotta
STMicroelectronics
Advanced Analytics

Data Science Lifecycle

“In the Information Era our data become a treasure, but they are like a rough diamond: we need to master how to process them in the right way.”

DATA SCIENTIST
Modeling & Deployment

DATA OWNER
Data Acquisition & Understanding

MANAGEMENT
Business Understanding

Industry 1.0
Mechanization and the introduction of steam and water power

Industry 2.0
Mass production assembly lines using electrical power

Industry 3.0
Automated production, computers, IT systems and robotics

Industry 4.0
The Smart Factory, Autonomous systems, IoT, machine learning
### DATA SCIENTIST
- Scientific Approach
- Problem Solving Mindset
- Support Domain Experts

### DATA DOMAIN EXPERT
- Business Understanding
- Domain Knowledge
- Company Expertise

#### Collaborations
- Academy
- Participations to conferences
- Internships
- Dissemination
- Spreading the Digital Transformation culture

#### Collaborations
- Manufacturing
- Product Quality
- Marketing
- Research & Development
- Training program
Some definitions: AI, ML and DL

- **Artificial Intelligence (AI)**
  - The broader concept of machines being able to carry out tasks in a way that we would consider smart.

- **Machine Learning (ML)**
  - An application of AI, based around the idea that we should really just be able to give machines access to data and let them learn from themselves.

- **Deep Learning (DL)**
  - A subset of ML, based on artificial Neural Networks.
Examples of Standards for Industrial Automation

ISO ICS 25.040.01 INDUSTRIAL AUTOMATION SYSTEMS IN GENERAL - https://www.iso.org/ics/25.040.01/x/
ISO ICS 35.240.50 IT APPLICATIONS IN INDUSTRY - https://www.iso.org/ics/35.240.50/x/


Who will make the decision?
Data Analysis Approach
Predictive Maintenance

- **Monitoring** of industrial systems
- **IoT**: network of sensors connected in cloud
- **IA / ML → Predictive Maintenance**
• **Data Analysis:**
  • Starting from the **root-cause analysis**, highlight important cause-effect relationships

• **Training** of a classification/regression model for predicting incoming faults

• **Continuous system monitoring**
AI Techniques

- **Increase the number of features** that can be processed
- Integrate predictive maintenance with **automated industrial processes**
- Deploy **fault prediction model** directly on sensors
• IoT: **More accessible user-interface** for the operator
• IoT: **Centralized monitoring** of the system
• **Early detection** of incoming faults
• Reduce time for maintenance → Costs and risk are reduced
• Improvement of the fab **safety**
Critical points

• Industrial facilities must be handled as a network (general responsibility), not as independent nodes (single responsibility)

• Acquisition and process of unstructured data

• Training of unsupervised models
Designing an IoT Reference Architecture → Deepening

Giordano Scuderi
STMicroelectronics
Internet Of Things applications

• IoT applications can be seen as:
  • **things** (devices) sending data that generates **insights**
  • insights generate **actions** to improve business processes, city life and our home living
IoT General Design Principles
By decoupling the ingestion from the processing the IoT application can make several decisions without impacting devices

- This is achieved through the use of queues or related messaging services
- Get data into a queue as soon as possible
Principle: Be ready for the data tsunami on Day 1

- The ingestion layer shall be a highly scalable platform that can handle a high rate of streaming device data
  - Be ready for the data tsunami on Day 1
Principle: design for offline behavior

- Design your software solution to handle situations in which there is no connectivity for an extended period of time.
- Track devices that are not communicating with the cloud application on a regular timeframe.
There is no point in adding static data which could be determined from within the cloud application rather than doing it on the IoT device.

As data is ingested in the application, data should be enriched and compressed.

data deserialized

Principle: enrich data at the cloud
Principle: ensure least privilege permissions

- Each IoT device should be given permission to dedicated channels within the application
- By restricting access, one compromised device will have fewer opportunities to impact any other devices
Principle: Cost optimization

- Select resources which are cost-effective
- Expenditure awareness: know where are you spending
- Optimize over time!
IoT Architecture
Demo Session

Filippo Milotta
STMicroelectronics
Demo tools: How and What

**HOW**
- Data Architecture

**WHAT: Azure**
- DataBricks
- MLflow
- RStudio
- MLStudio

**Typical Layers:**
1. Data Ingestion
2. Data Preparation
3. Data Warehouse
4. Data Model
5. Data Visualization

**USE-CASES**
- Outlier Detection
- Anomalies Classification
- Root Cause Analysis
- Action Recommendation
Overview on Use-Cases shown in this demo

USE-CASES

Outlier Detection

Anomalies Classification

- Scratch
- Ring

Action Recommendation

Root Cause Analysis
**What is it?**

- DataBricks provides a **Unified Analytics and Analysis Platform** powered by Apache Spark for data scientist teams to collaborate with data engineering and lines of business to build data products
- **Ref:** [https://databricks.com/company/about-us](https://databricks.com/company/about-us)

**How does it work?**

- Coding into notebooks, like **Jupyter Notebook** (Python)
- A Jupyter Notebook document is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots, …
- Several languages available: **Python, R, Scala, SQL, bash, html**

**Features**

Notebooks, MLFlow for tracking ML experiments, RStudio on DataBricks, **Scalable configuration** of Driver Nodes (like a tunable VM), Data ingestion from ADLS, Apache Spark environment (parallel computation, redundancy), MLLib to optimize ML code

**Voice of the user**

“A complete suit for coding in many languages, scaling computational power on the need, pay-as-you-go, and leveraging the power of Azure”
What is it?

- MLflow is an open source platform for managing the end-to-end machine learning lifecycle. It supports Java, Python, R, and REST APIs
- Ref: https://docs.databricks.com/applications/mlflow/index.html

How does it work?

There are two key elements:

- **Experiments**: They are the primary unit of organization in MLflow; all MLflow runs belong to an experiment.
- **Run**: It is a collection of parameters, metrics, tags, and artifacts associated with a machine learning model training process.
  - Each Run records the following information: Source, version, Start & End time, Parameters, Tags, Metrics, Artifacts.

Features

Each experiment lets you visualize, search, and compare runs, as well as download run artifacts or metadata for analysis in other tools. The experiment UI performs the following key tasks:

- List and compare runs
- Search for runs by parameter or metric value
- Visualize run metrics
- Download run results

Voice of the user

“An useful tool for reporting and keeping track during the Data Science flow. You can easily keep track of any fine-tune setting together with the outcomes”
What is it?
• RStudio is an IDE for R programming. We leverage DataBricks to run an instance of RStudio Server.
• Ref: https://docs.azuredatabricks.net/spark/latest/sparkr/rstudio.html

How does it work?
DataBricks allows to run RStudio Server in either Open Source or Pro Edition version (Limitations applied)
The environment is accessible from any user who can also access DataBricks

Features
• **Scalable configuration** of Driver Nodes (like a tunable VM), Data ingestion from ADLS, Versioning with GIT (just checked), Environment Configuration (needed libraries are automatically installed when DataBricks is started).

• Consolidate versioning with GIT, improve management of user sessions (this can be done in the Pro Edition through admin tools, while it’s more tricky with the Open Source one), leverage Apache Spark for performance improvement (scalability, efficiency)

Voice of the user
“ This solution finally enabled us to move from a local to a cloud environment, in a seamless and smooth way”
**What is it?**

- A node (graph) based development environment, in which you can drag-and-drop tool you can use to source, transform, and analyze data through various manipulation and statistical functions, and generate a set of results
- Ref: https://docs.microsoft.com/en-us/azure/machine-learning/studio/what-is-ml-studio

**How does it work?**

- A set of premade nodes (blocks of the graph) is available for development
- No specific coding skills are required
- Blocks for free coding in Python and R are available, if needed

**Features**

- We leverage the free workspace tier (limitations applied) for double checking workflows also defined in DataBricks
- **Fast prototyping** with meaningful report (visualization of charts and dataset, also in intermediate steps
- Model deployment for web services purposes (model consumption). Not free-tier allows high computational power.

**Voice of the user**

“Fast prototyping and no coding skill required. An easy and sharp way to design a preliminary solution, before of moving into a complete environment like DataBricks”
Thank you by…

Giuseppe Ursino
Digital Transformation
Enterprise Architect

Filippo Milotta
Digital Transformation
Data Scientist

Giordano Scuderi
Digital Transformation
Solution Architect

Emanuela Ali'
Ufficio Affari Generali
Catania Site

Mario Marroccia
IT Director
Head of ERP Factory