NREL Transforming ENERGY

NREL Stratus - Enabling Workflows to Fuse Data Streams, Modeling, Simulation, and Machine Learning

Aaron Andersen
David Rager

Smoky Mountains Computational Science & Engineering Conference
October 18-20, 2021

Transforming **ENERGY** through computational excellence

Contents


Advanced Computing

HPC Environment

ESIF Data Center

- Warm water cooling
- Focus on energy efficiency

Eagle

- 2168 Node (Intel Skylake)
- 50 Node (Intel Skylake, Nvidia V100 GPU)
- Infiniband EDR
- 17 PB Lustre Filesystem

Stratus Environment

Big Data Analytics

- data warehousing
- data management tools

Containerized Applications

- multiple scheduling systems
- Docker containers at the edge
- Docker serverless functions

Growth In

- IoT support for field experiments
- grid management studies

Ongoing Support for

- data processing workflows
- public web applications
- publishing of large open data sets

HPC, Cloud & Edge

High Performance Computing (HPC)

- Highly optimized compute for simulation
  - Numerically intensive
  - Parallelized
  - Tightly coupled
- Scheduled & Allocated resource
- Optimized for long running resulting in low cost per flop
- Periodic incremental investment
  - Systems ”on the floor” 4~5 years
  - Fixed cost

Cloud

- Dynamic cost
  - Spot
  - Reserved
  - On-demand
- Big Data
  - Analysis
  - Distribution
- High Availability
  - Data streams
  - Multiple cloud sites
  - Dynamic scaling
  - Scheduling for ML pipelines

Edge

- Deploying computational resources “close” to the IoT device
- Driven by
  - Real or Near-realtime
  - Data reduction
- Highly dynamic
- Challenges around fault tolerance
  - Multiple single points of failure expected

Reference Architecture

Edge

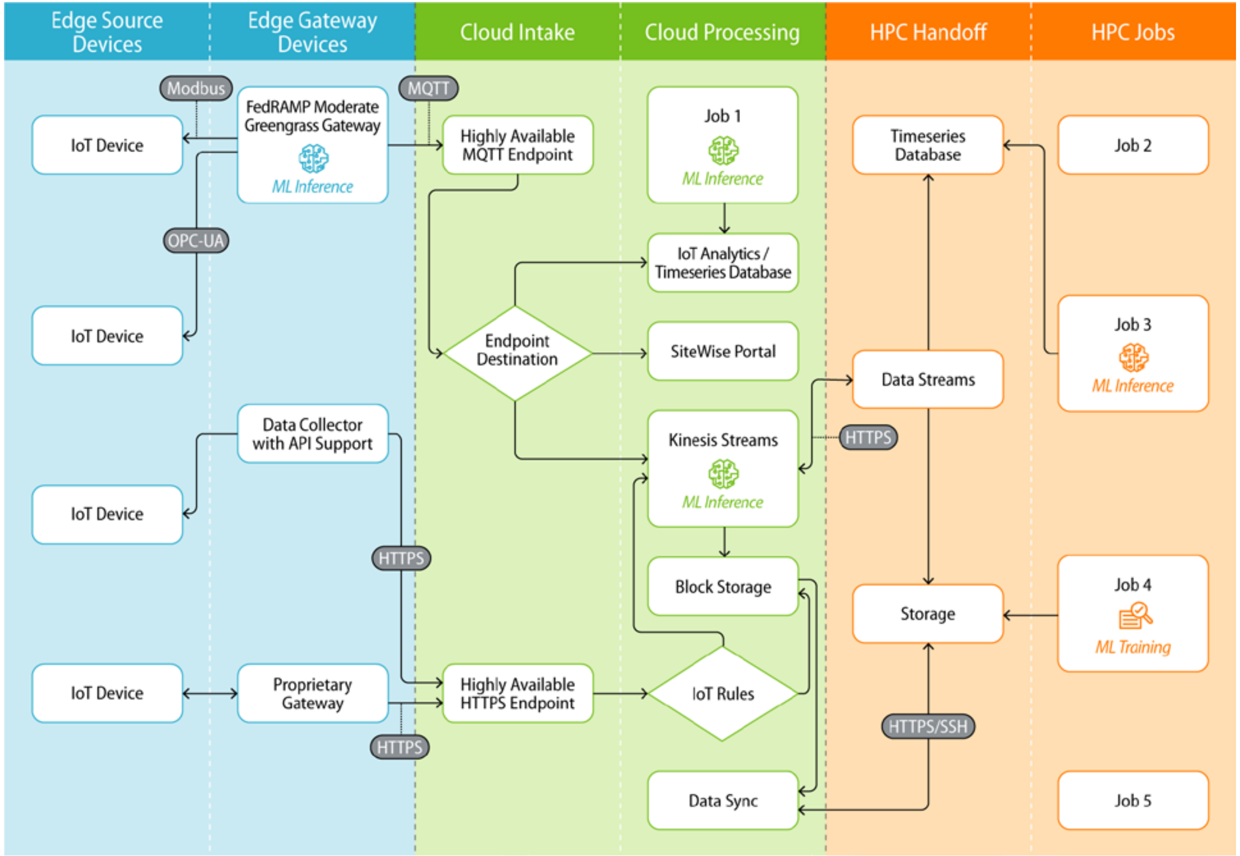
- Low Latencies
- Smart aggregation

Cloud

- High availability
- Manage to the edge
- Security
- Dynamic Scale

HPC

- Large scale modsim
- ML training & Inference
- Process very large datasets



Transforming **ENERGY** through computational excellence

Areas of Concern Reference Architecture

- Edge
  - Highly constrained compute capabilities
  - Data separated by location or devices
  - Intermittent connectivity and potential signal points of failure
- Cloud
  - Complexity
  - Variable costs
  - Complex deployments
  - Data transfer costs
- HPC
  - Scheduling and scheduling contention
  - Wall time limits and policies

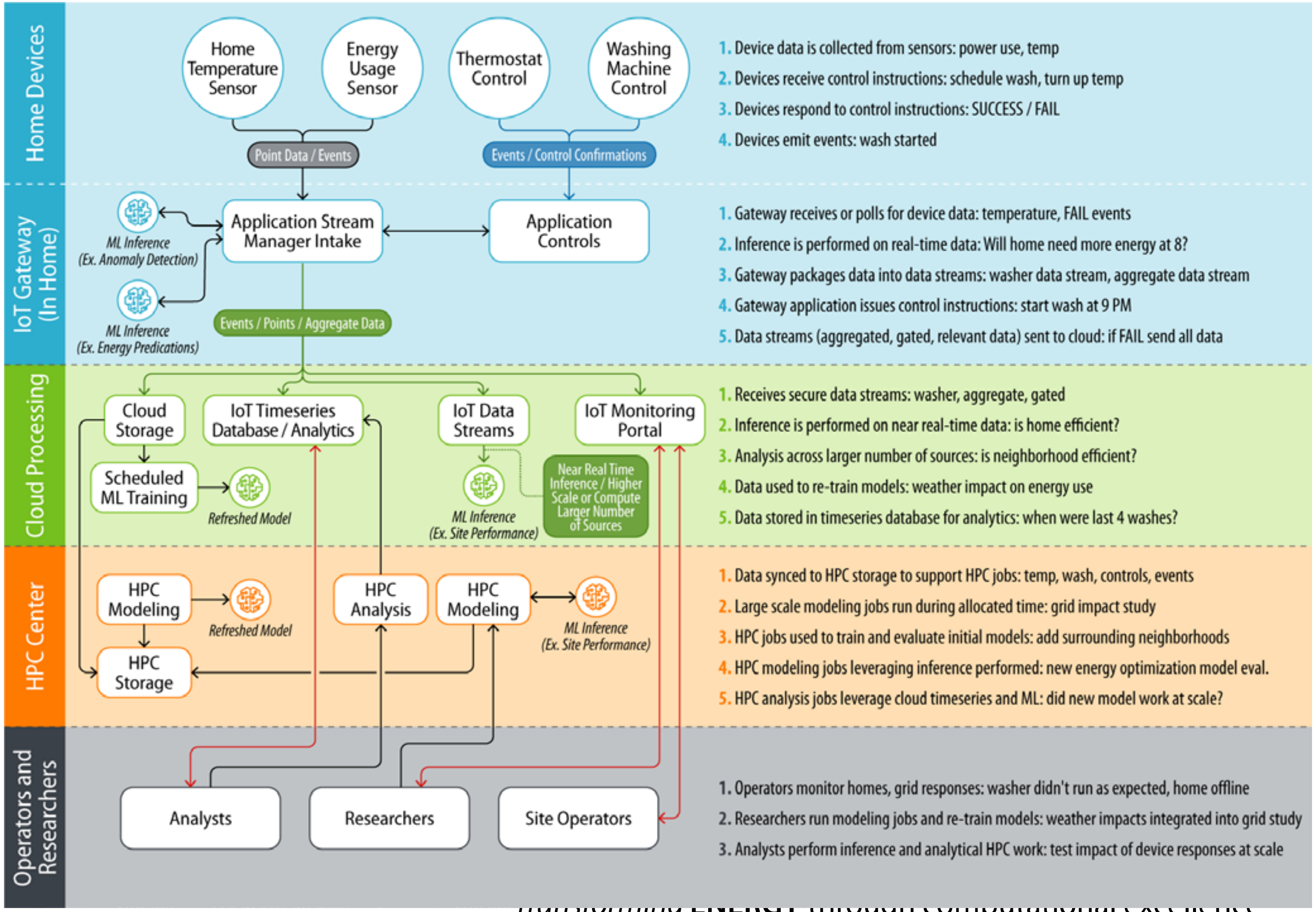
Smart Neighborhood Example

Overview

- Improve smart controls
- Measure satisfaction
- Measure changing energy usage
- Simulation of effects at large scale

Goals

- Optimize energy usage for current and future infrastructure
- Improve the energy usage profile for adoption of renewable energy resources

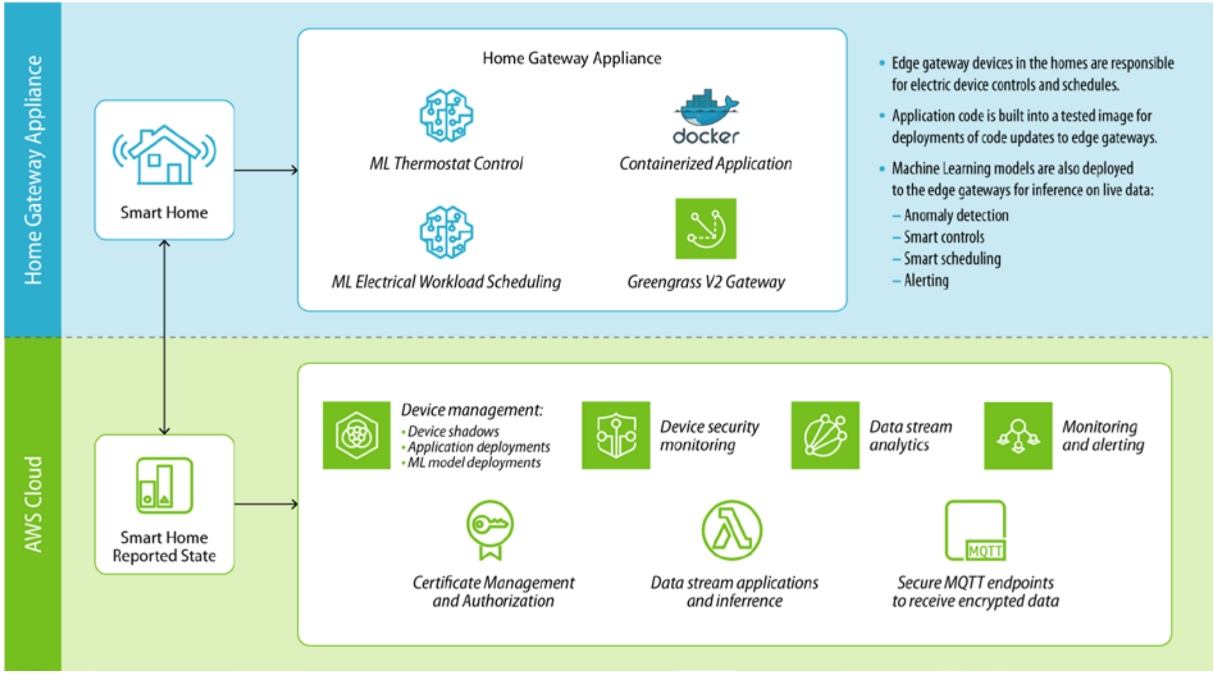


Workflow Components

- Data Generation and Intake
  - Live access to data streams
- Data Cleaning/Transformation
  - Multiple device integration
  - Data quality checks
  - Aggregation
- Synthetic Data/Modeling
  - Tuning strategic model
  - Forecasting or policy simulation
  - Data often provided up to cloud
- Machine Learning Inference
  - Assist future behavior (prediction)
  - Anomaly detection
  - Create simulation data
- Machine Learning Training
  - Regular scheduling
- Edge Device Management
  - AWS GreenGrass

Edge Device Management

- Deployment of ML to the Edge
- Management of Large footprint
- Monitor streams
- Live dashboards



Challenges

Challenges

Complexity

- Incrementally built upon highly available and scalable services
- Development of patterns and practices
- AWS costs and limits

Data flow and Data management

- Published data streams with APIs
- Compressed files
- Query access via DBMS systems

Transforming **ENERGY** through computational excellence

Questions

www.nrel.gov

NREL/PR-2C00-81102

Transforming **ENERGY** through computational excellence