



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

- 1 **Advanced Computing at NREL**

- 2 **HPC, Cloud & Edge**

- 3 **Reference Architecture**

- 4 **Smart Neighborhoods Example**

- 5 **Challenges & Future**



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



A landscape photograph featuring three white wind turbines on a rolling green hill. The sky is bright blue with scattered white cumulus clouds. In the foreground, there is a field of tall, green grass with some yellow wildflowers. A dark, semi-transparent banner is overlaid at the bottom of the image, containing the text 'HPC, Cloud & Edge' in white.

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

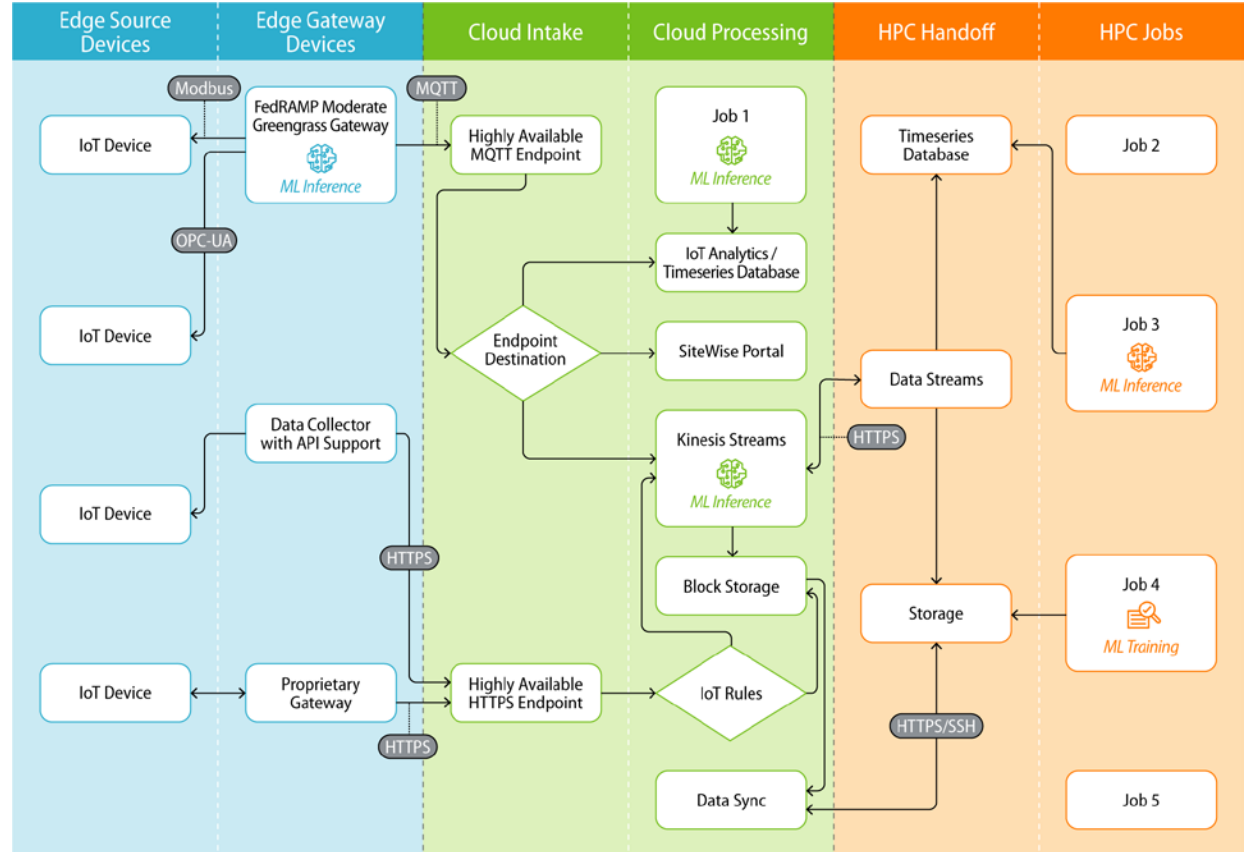
- Low Latencies
- Smart aggregation

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



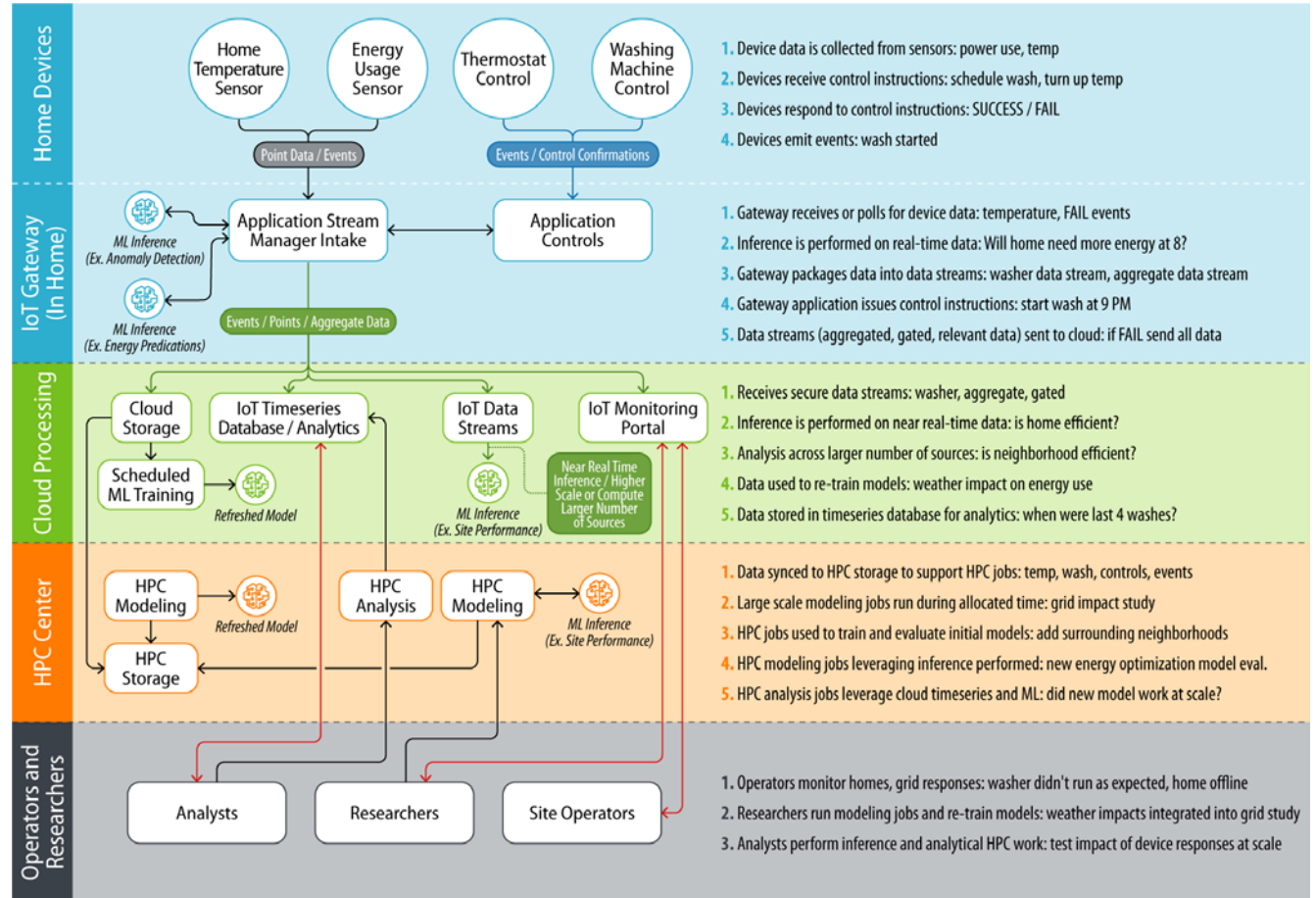
Smart Neighborhood Example

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

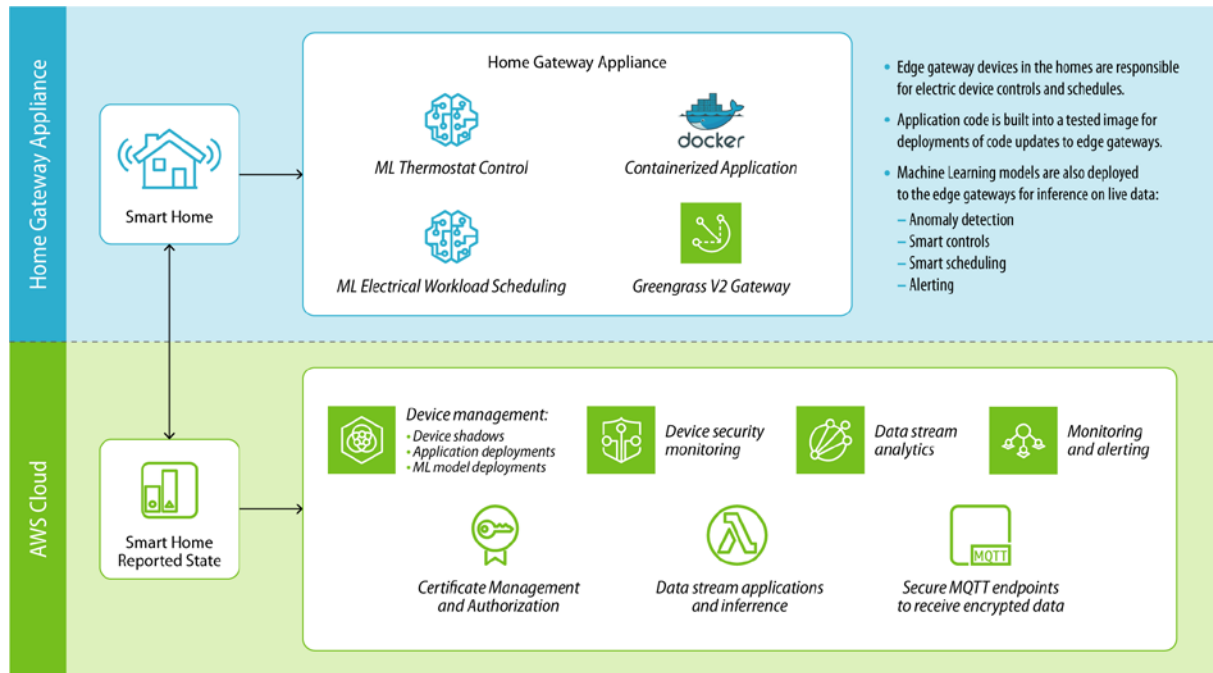


Workflow Components

- 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



Transforming **ENERGY** through computational excellence

A photograph of a server room. In the background, there are several rows of black server racks with glowing green lights. The floor is a light-colored, polished surface. In the foreground, a raised floor is visible, supported by a metal frame. Underneath the floor, there is a complex network of pipes, conduits, and cables. A prominent blue pipe runs vertically through the center. The walls are a bright yellow color. The overall scene suggests a data center environment with intricate infrastructure.

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



Questions

www.nrel.gov

NREL/PR-2C00-81102

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Transforming **ENERGY** through computational excellence