

www.inl.gov



Advanced Vehicle Testing Activity (AVTA) Data Warehouse

Patti McGuire

**U.S. / China Workshop On PEV and Charging
Infrastructure Data Collection and Reporting
Idaho Falls, Idaho
August 2013**

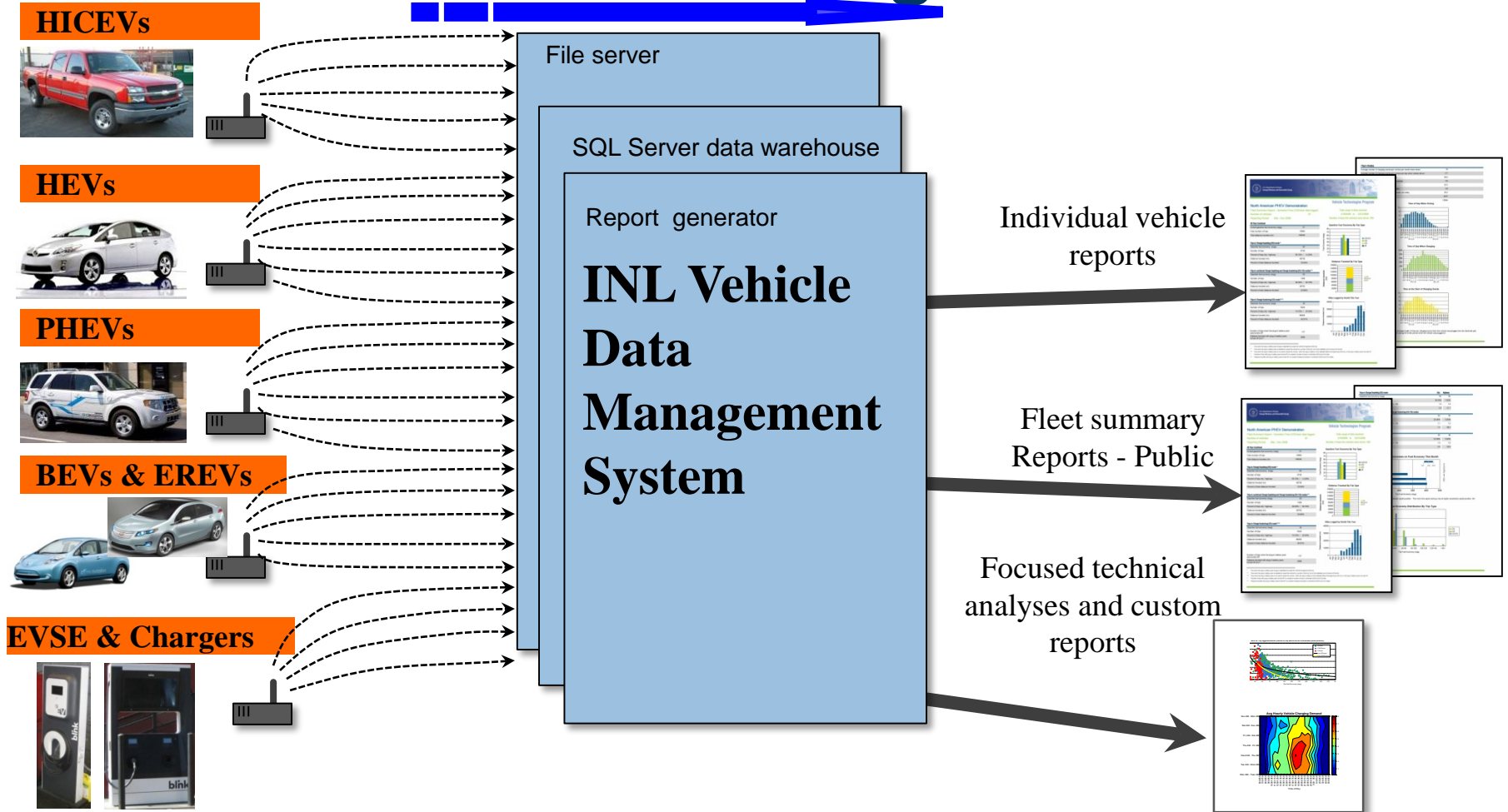
INL/MIS-13-29806



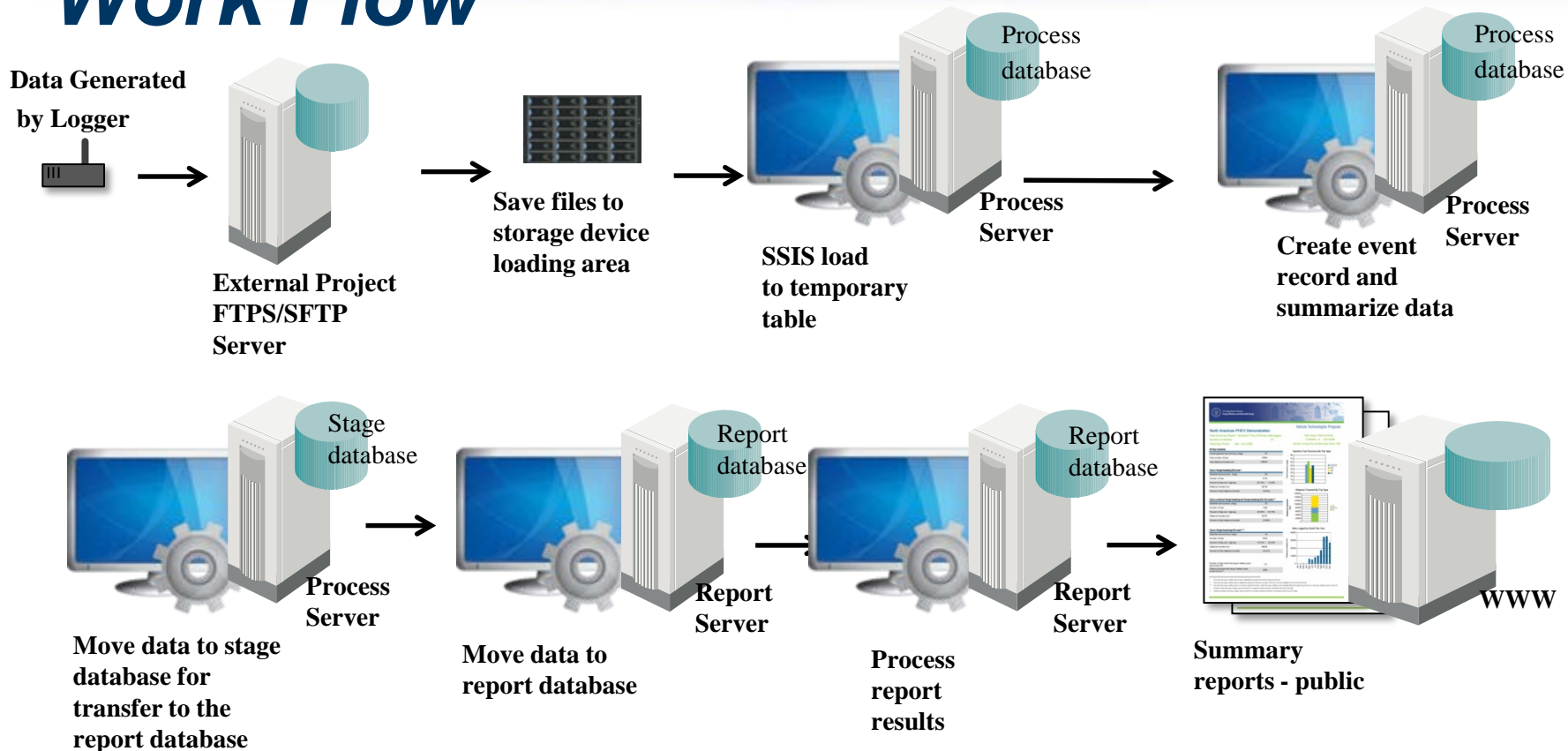
Handling Data

- Data warehouse
- Work Flow
- Receiving files
- File types
- Servers
- Processing files
- Basic database practices
- Quality Assurance (QA)
- Standard report calculations
- Security

Data Warehouse Management



Work Flow



Receiving Files

- **Data Files pushed to INL SFTP server**
- **Data Files are pulled from data provider servers**
- **Data Files are emailed in some cases or arrive on hard drives**

- **Files are moved to INL file share. All data files are archived**

- **No direct links from INL databases to data providers**

File Types

- **Comma delimited files (CSVs)**
- **Column delimited files (txt)**

- **Vehicle Data (internal and external logging)**
- **EVSE data**
- **Maintenance data**

- **No Excel files allowed (Excel is not available on INL database servers)**

Servers

- **Secure FTP Server**
- **Processing Server: Data are loaded from the raw files into a process database.**
 - **After process is complete: data are copied to a stage database.**
 - **Stage is then backed up and copied to the Report Server**
- **Report Server: Stage database is recovered and data are copied to the report database for reporting**
- **WWW: Reports are published on WWW server**

Processing files

- Data are loaded using SQL Server Integrated Services (SSIS)
- Preprocessing is some times required. Perl or in SSIS.
- Load data into a process database into a temporary table.
- Track the filename with a load identifier
- Create a unique identifier (numeric) for each record
- Process data into standard data tables: raw data and event data (summarization of raw data)

Basic database practices

- **Relational database**
- **Numeric keys in the database are always faster**
- **Standard database, table and field names (Process, Stage, Report)**
- **Optimize indexing**
- **Processing – loads the data into the temporary tables then does summations into the stage database. Stage is backed up and copied to the Report server.**
- **Re-usable stored procedures for calculations**
 - **Finding the first and last value, maximum and minimum value, average value.**

Quality Assurance (QA)

- **Use tables to store the database name, table name, field name, low and upper bounds.**
- **Use tables to store rules for a result**
- **Report data that is out of bounds to data reviewer**
- **Mark data as unusable in reports base on QA check**

Calculation standards

- **Data in standard units (Miles per hour or Kilowatts).**
- **Use single stored procedure to make similar calculations**
- **Use tables to drive report results.**
- **Prepare report results into a result tables at report run time**
- **Use SQL Server Reporting Services (SSRS) for reports**

Security

