尊敬的各位领导，各位来宾，大家早上好！我是上海市新能源汽车公共数据采集与监测中心的王成名，今天我将会对上海数据中心1+X和全业务能力发展模式做一个交流。
 我们中心是2014年由经信委指导成立的一家民非组织，按照国家标准，目前接入了的车企数量超过36万辆，数据采集频率为3-10秒。目前，我们的采用频率从原来的3-10秒提升到1秒。在内容项目增加了三大类八大项。修订后的标准更有利与安全监管、事故鉴定等场景应用和数据分析。新地标已经通过了地标委和审计局组织的评审会，目前已经通过了，预计2021年会执行。
 今年年初我们做了“十四五”规划，总结过去走过的7年，我们认为我们走过了三个阶段。前两年是我们的基础阶段，我们实现了基础数据的采集和存储。第二阶段是我们的能级提升阶段，我们做了比较简单的数据仓库，为什么我说是一个比较简单的数据仓库？因为我们的数据仓库只有两层，这是一个不完整的数据仓库。2019年，我们数据中心发展进入了第三个阶段，这是我们迈向智能化的阶段。我们主要实现从网关接入、数据质量、数据标签标定、数据分析方面实现智能化。这对我们数据的基础场景带来了一定的挑战。

我们数据中心是一个多元异构的大数据平台。我们有新能源汽车平台、动力电池溯源平台、加氢车展一体化数据平台，这样一个多元异构的数据平台，无论是从数据层面还是从数据体量来说都有一定的壁垒。我们服务两个层面：为政府提供安全监管和政策评估方面的服务，为企业提供数据分析及数据产品相关服务。
 我们建立这样一个多元异构的大数据平台，需要解决五大重要问题，首先是如何实现分布式的数据接入能力。目前主流的技术是不支持分布式的。同时我们要做一个多层架构的数据存储，因为面对PB级别的数据存储，如果说采用单一的数据源存储，会面临很大问题，一个是性能问题，一个是经济成本问题。如是PB级别，采用热存储会达到300万以上的开支。因此，我们需要一套很高效的数据分层和压缩机制，形成一个平滑的技术体系。

由于时间的关系，我简单介绍一下数据分层方面我们上海做的一些探索和研究。我们的数据存储架构是基于业务模型进行设计的，比如说一辆车发生了事故，这个时候我们拉出时间做数据分析，就是基于位置和时间的数据分析，我们可以采用列式数据库进行存储，我们称之为热存储。如果说分析近年来新能源汽车情况，这种数据存储我们一般会采用文件存储来实现，我们称之为温存储。再比如说三五年之前的数据，我们不会把它用来分析，或者说分析的频率很低，这部分的数据怎么办？我们肯定是不能删的，我们要对其进行冷存储或者称之为数据归档，数据归档技术我们称之为OSS归档，这是一个很复杂的技术方案，叫做数据的冻结。再次使用数据我们要对数据进行解冻。我们为什么要采用数据分层架构，主要是出于经济成本的考虑。

还有数据的冷热分离。比如说HBase数据，它不适合做数据分析，如何去解决这个问题呢？我们必须要开发一款或者是借助其他的中间件来解决底层异步解耦问题，把数据同步到大数据平台，这样我们就实现了数据冷热分离。同时数据分析之后，数据归档要回流，分析结果要在前端展示。目前，我们的数据都是非结构化的物联网数据，现有的大数据技术不支持直接分析，那么我们怎么做的呢？我们要把非结构化的物联网数据转化为结构化数据，转化为行列的表格。基于这种结构化的数据可以满足绝大多数的数据分析，当然这个过程也是比较复杂的。
 总结一下，我们无非就是构建了一个数据终台，一套可持续、让数据用起来的机制；一套持续不断的把数据资产变成数据服务的机制。我们看到，目前构建的数据终台包括了：数据汇集、数据提取加工、数据服务、可视化数据产品。

我们做了这么多事情，归结起来，我们数据中心技术层面用了五年时间，进入了智能化、中台化、开放化的阶段，形成了系统到分析的业务模式和发展阶段，我们的服务领域将面向长三角，希望在长三角和其他地方平台推广和输出我们的上海模式和上海的解决方案。时间的关系，我今天的分享就到这里，谢谢大家。