



数据仓库大作业汇报

——对电动汽车数据的分析

Electric Vehicle Title and Registration Activity

—— 小组成员：黄勖、石宇昊、曾鸿勇、邹慧 ——

目录



- 1 背景需求分析
- 2 结构设计
- 3 环境部署
- 4 ETL过程（视频演示）
- 5 可视化分析



背景与需求分析

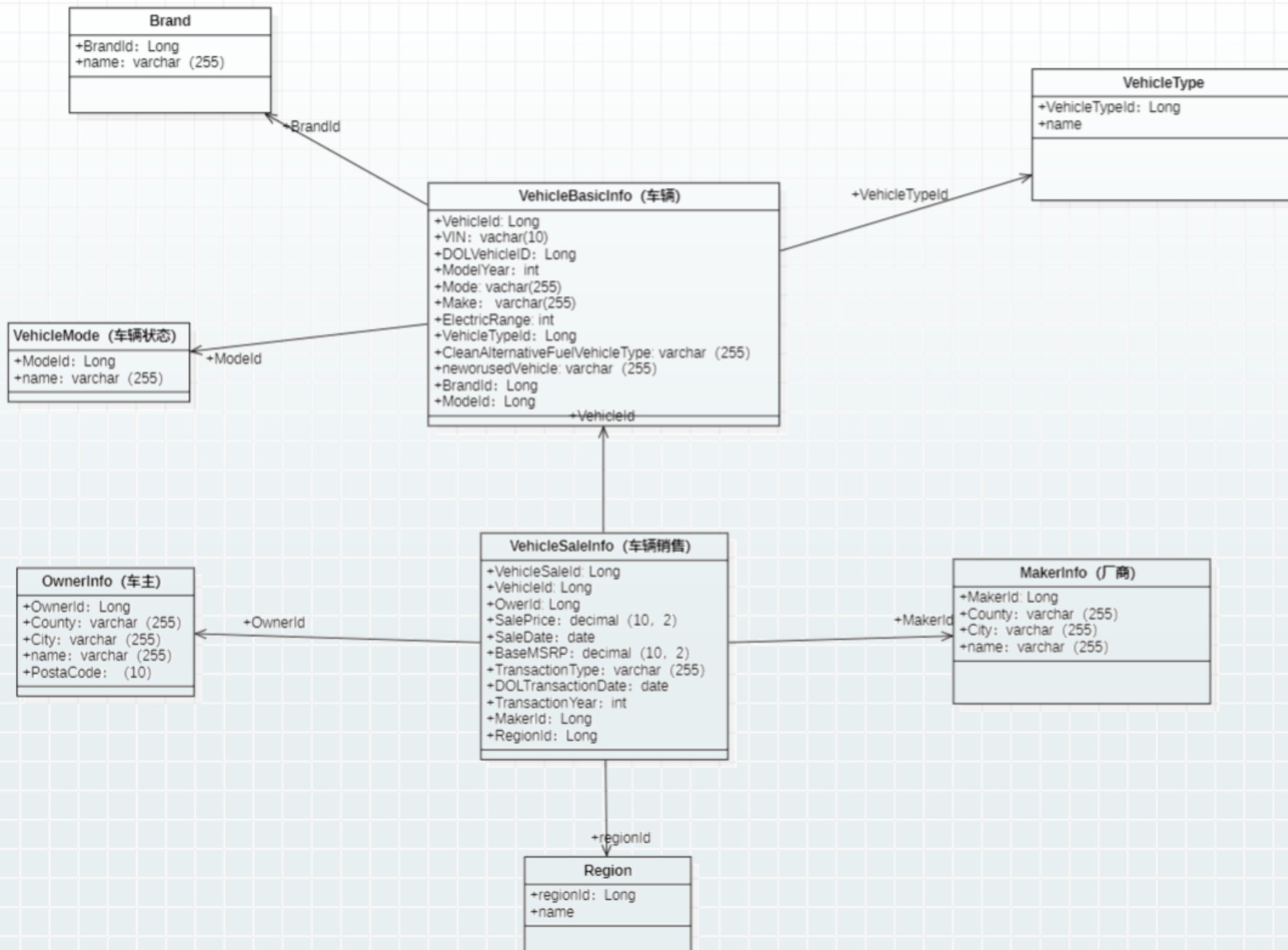


我们搜集了2010-2023电动汽车车辆销售与所有权变更活动数据集，这个数据集记录了华盛顿州电动汽车约85万所有权变更和车辆注册交易。该数据集详细记录产权的变更情况，以及授权车辆在华盛顿州公共市场上进行的交易。这些记录提供了有关电动汽车所有权和使用模式的关键信息，有助于了解该州电动汽车市场的趋势和普及率。这些数据对于政策制定者和行业利益相关者具有重要价值，希望我们的分析能够为电动汽车市场的发展动态提供帮助，评估市场变化的影响，并制定支持电动出行在华盛顿州可持续发展的战略。可以帮助新能源汽车投资人考虑如何投资使利益最大化



结 构 设 计

事实表和维度表



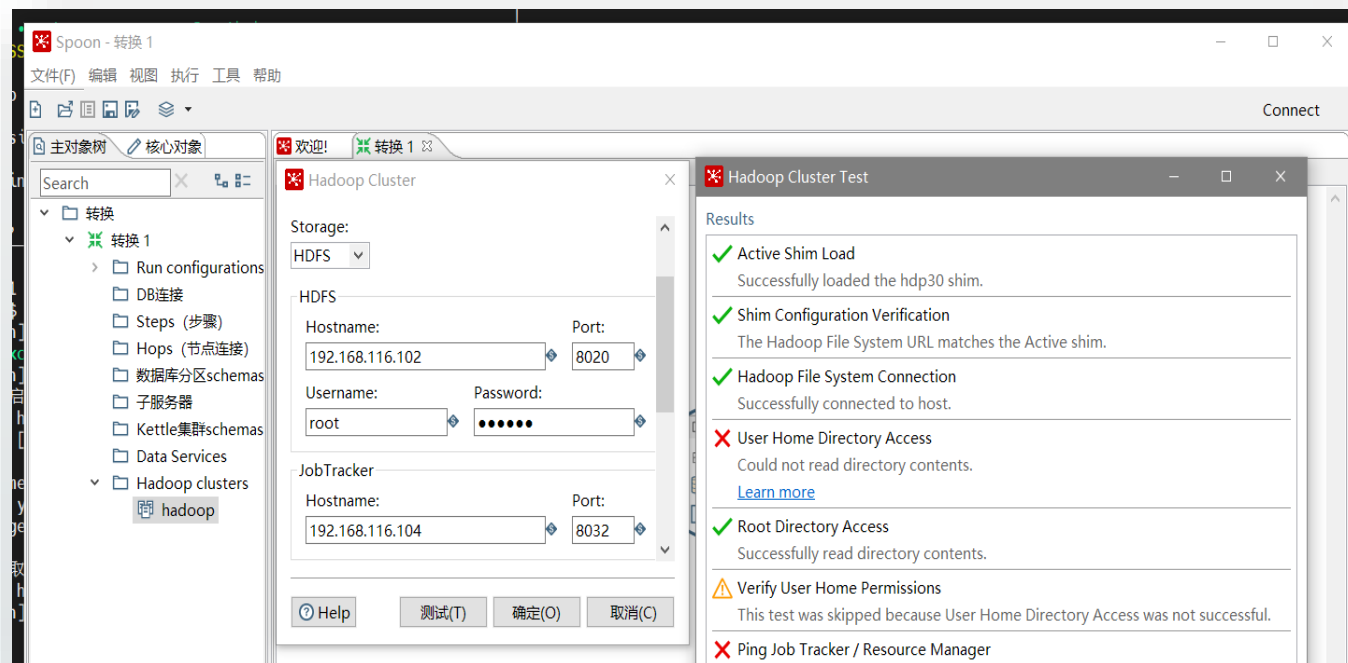


环 境 部 署

Hadoop

Hive

Presto



我们的数据仓库物理结构的解决方案基于Hadoop、Hive和Presto，形成了一个强大的大数据解决方案。Hadoop作为底层分布式存储和计算框架，提供高可靠性和可扩展性，存储大规模数据。Hive作为数据仓库的元数据存储和查询引擎，通过HQL将结构化查询转化为MapReduce任务，实现高层次的SQL-like查询。Presto则作为交互式查询引擎，支持快速查询和分析，


```
5YJYGDEEXM 154520421 2021 TEST A Model V D MIII I Kira KTDVI AMD WA E2022022002 15
5YJYGDEEXM Last login: Last login: Sun Dec 31 23:44:38 2023 from 192.168.116.1 5
5YJYGDEEXM [atguigu@had [atguigu@hadoop102 ~]$ ./start_presto.sh 03
5YJYGDEEXM [1] 6720 Started as 6938 >&1 & 5
1N4BZ0CP2G [atguigu@had [atguigu@hadoop102 ~]$
Time taken:
hive (car)>
```

硬件配置

操作系统（OS）：选择了CentOS作为操作系统，它是一种广泛用于企业环境的Linux发行版，提供了稳定性和安全性。

硬件规格：每台虚拟机配置了4GB的内存，共有3台虚拟机。此外，每台虚拟机还有50GB的硬盘空间，用于存储系统和应用程序的数据。

&

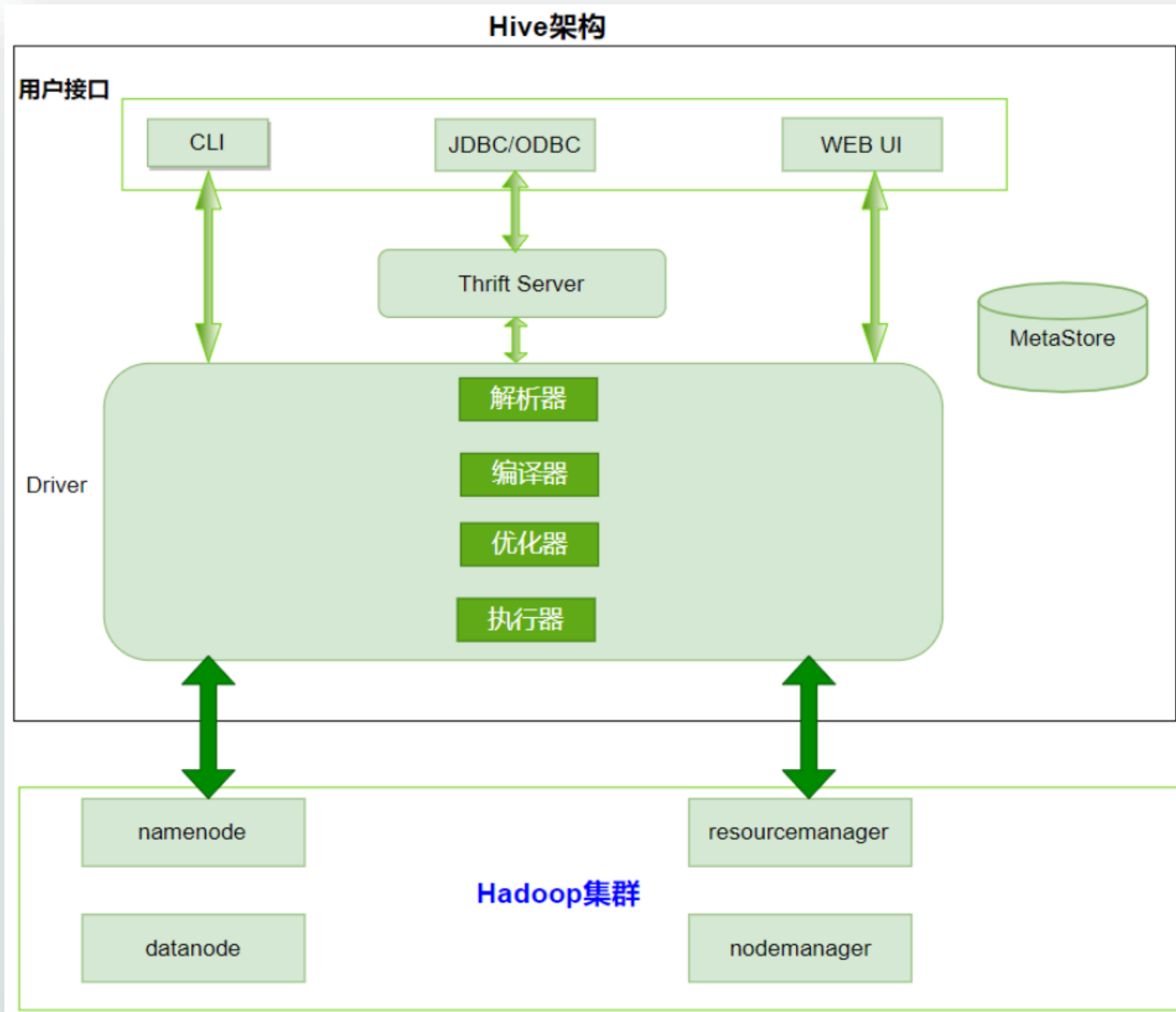
软件配置

数仓平台：采用Hadoop和Hive作为数仓平台，Hadoop用于分布式存储和处理大规模数据，而Hive提供了SQL接口，使用户能够通过SQL查询分布式存储的数据。

ETL工具：选择了Kettle作为ETL（抽取、转换、加载）工具，它能够方便地处理数据的提取、清洗和加载，是数据仓库中常用的工具之一。

数据可视化工具：使用Superset作为数据可视化工具，Superset能够通过简单的配置和交互性地创建丰富的数据可视化报表，支持多种数据源。

Hive架构



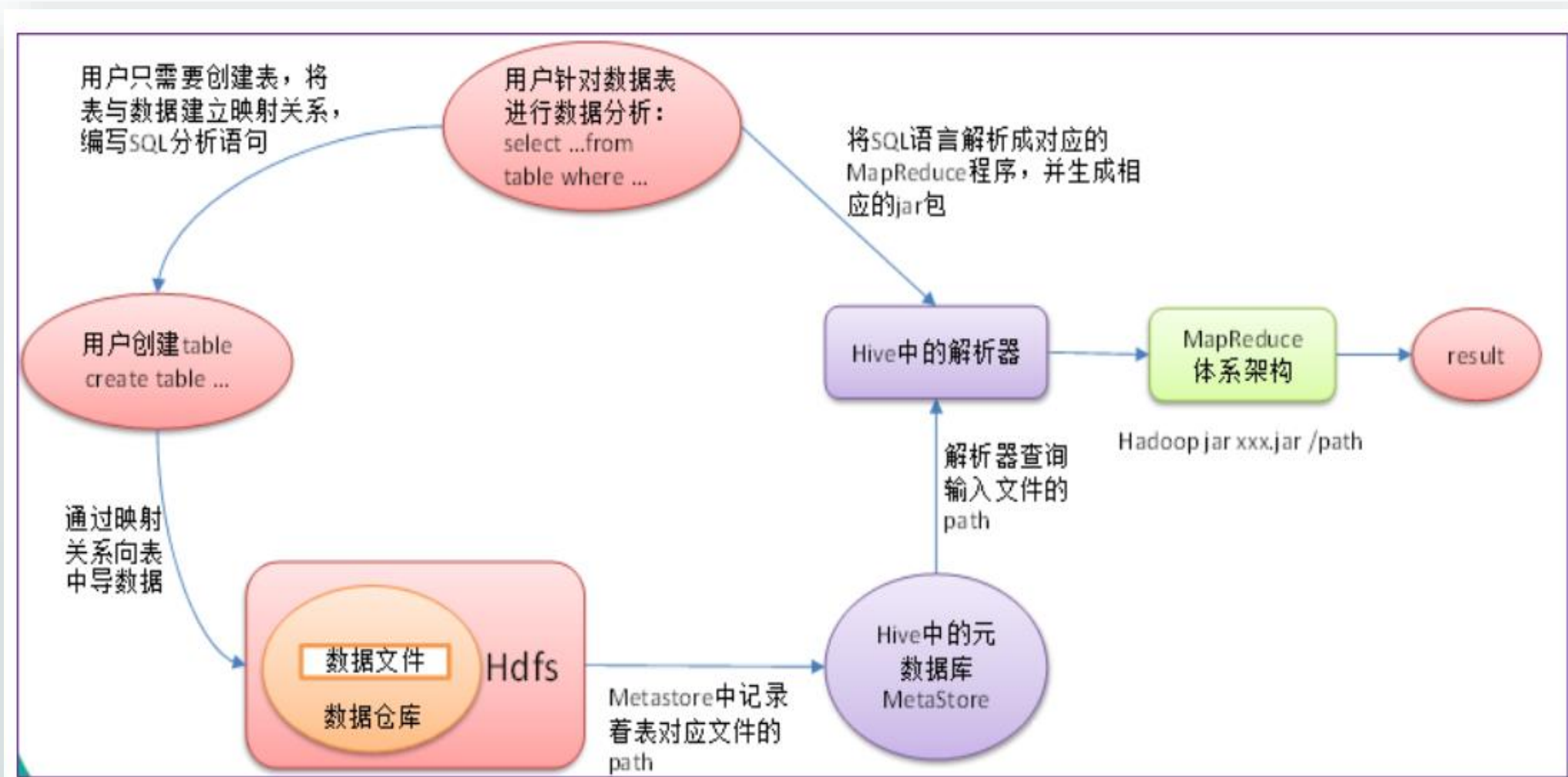
Hadoop:

Hadoop是一个开源的分布式存储和计算框架，它提供了处理大规模数据的能力。Hadoop采用分布式存储模型，将数据划分为多个块并存储在不同的节点上，以实现高度的可扩展性和容错性。

Hive:

Hive是一个基于Hadoop的数据仓库工具，它提供了类似SQL的查询语言（HiveQL），允许用户通过SQL方式查询和分析存储在Hadoop中的数据。在物理设计上，Hive将查询转化为一系列的MapReduce作业，这些作业在Hadoop集群上并行执行。

Hive架构



Presto 由一个 Coordinator 和 多个 Worker 组成

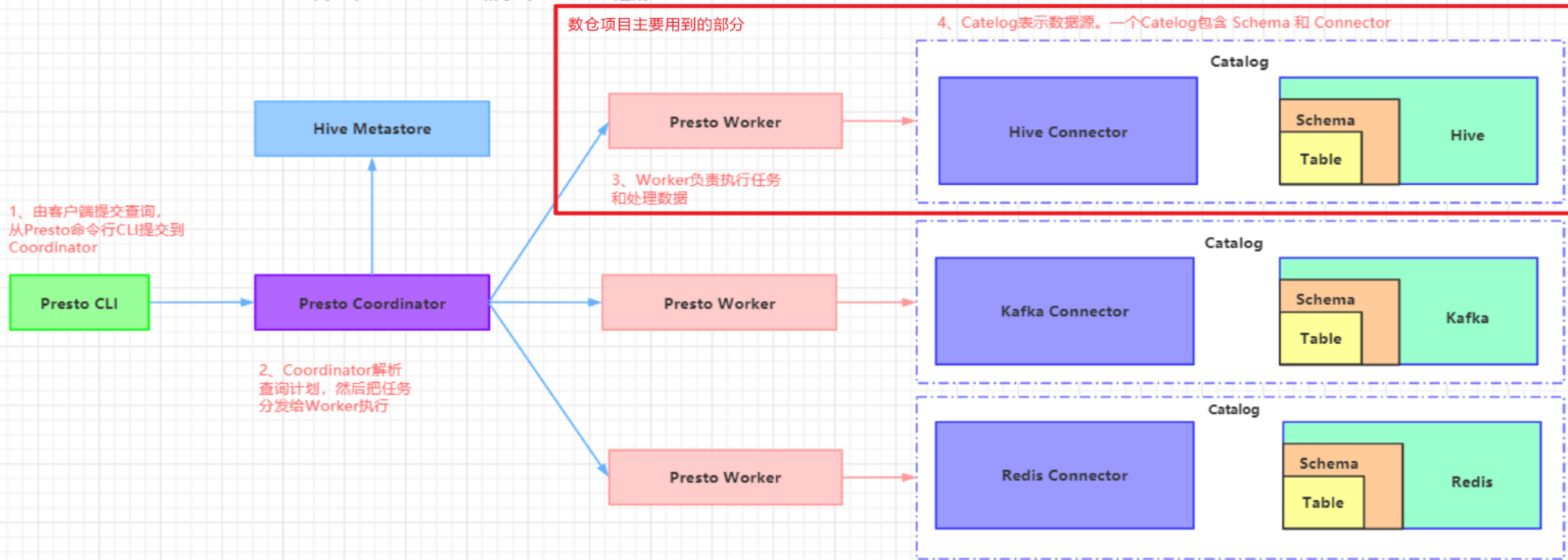
数仓项目主要用到的部分

3、Worker负责执行任务和
处理数据

4、Catalog表示数据源。一个Catalog包含 Schema 和 Connector

1、由客户端提交查询，
从Presto命令行CLI提交到
Coordinator

2、Coordinator解析
查询计划，然后把任务
分发给Worker执行



Presto是一个高性能的分布式SQL查询引擎，用于查询大规模分布式数据。与Hive不同，Presto不使用MapReduce，而是通过使用自己的执行引擎来直接查询数据。

视频演示

(简洁版, 详细过程请参考ftp上传文件)

① ETL操作: 数据转换

WPS Office

找稻壳模板

presto操作文档.docx

文件

开始

插入

页面

数据

视图

分享

格式刷

粘贴

宋体

11

A+

A-

B

I

U

A

田

羊

%

000

0.00

0.00

工作表

条件格式

求和

筛选

查找

T17

fx

WA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1	Clean AlteVIN (1-10)	DOL Vehicle	Model Year	Make	Model	Vehicle	Pr	Electric	F	Odometer	F	Odometer (New or Use	Sale Price	Sale Date	Base MSRP	Transactio	DOL Transe	Transactio	County	City	State of
2	Battery	E15YJ3E1EC6I	1981598	2020	TESLA	Model 3	Passenger	308	0	Odometer	1	New	0	0	Original	1	January 07	2020	King	KIRKLAND	WA
3	Battery	E17SAYGDEE8F	251929492	2023	TESLA	Model Y	Passenger	0	15	Actual Mil	New	51630	August 11	0	Original	1	September	2023	King	KIRKLAND	WA
4	Battery	E17SAYGDEE8F	251929492	2023	TESLA	Model Y	Passenger	0	0	Odometer	1	New	0	0	Original	1	September	2023	King	KIRKLAND	WA
5	Battery	E15YJ3E1EC6I	1981598	2020	TESLA	Model 3	Passenger	308	0	Odometer	1	Used	0	0	Registrati	October 11	2021	King	KIRKLAND	WA	
6	Battery	E15YJ3E1EC6I	1981598	2020	TESLA	Model 3	Passenger	308	0	Odometer	1	Used	0	0	Registrati	October 30	2020	King	KIRKLAND	WA	
7	Battery	E15YJYGDEEX	154530421	2021	TESLA	Model Y	Passenger	0	0	Odometer	1	Used	0	0	Registrati	April 14	2023	King	KIRKLAND	WA	
8	Battery	E15YJYGDEEX	154530421	2021	TESLA	Model Y	Passenger	0	0	Odometer	1	Used	0	0	Registrati	April 05	2022	King	KIRKLAND	WA	
9	Battery	E15YJYGDEEX	154530421	2021	TESLA	Model Y	Passenger	0	15	Actual Mil	New	63290	May 07 202	0	Original	1	May 26 202	2021	King	KIRKLAND	WA
10	Battery	E15YJYGDEEX	154530421	2021	TESLA	Model Y	Passenger	0	0	Odometer	1	New	0	0	Original	1	May 26 202	2021	King	KIRKLAND	WA
11	Battery	E11N4BZ0CP2C	348645972	2016	NISSAN	Leaf	Passenger	84	0	Odometer	1	Used	0	0	Registrati	February 2	2021	King	SEATTLE	WA	
12	Battery	E11N4BZ0CP2C	348645972	2016	NISSAN	Leaf	Passenger	84	0	Odometer	1	Used	0	0	Registrati	April 03	2019	King	SEATTLE	WA	
13	Battery	E11N4AZ0CP8I	185343021	2013	NISSAN	Leaf	Passenger	75	31440	Actual Mil	Used	31820	November 2	0	Transfer	1	December 1	2016	King	SEATTLE	WA
14	Battery	E11N4AZ0CP8I	185343021	2013	NISSAN	Leaf	Passenger	75	0	Odometer	1	Used	0	0	Registrati	November 0	2017	King	SEATTLE	WA	
15	Battery	E11N4BZ0CP2C	348645972	2016	NISSAN	Leaf	Passenger	84	0	Odometer	1	Used	0	0	Registrati	February 2	2023	King	SEATTLE	WA	
16	Battery	E11N4AZ0CP8I	185343021	2013	NISSAN	Leaf	Passenger	75	0	Odometer	1	Used	0	0	Registrati	December 1	2016	King	SEATTLE	WA	
17	Battery	E11N4AZ0CP8I	185343021	2013	NISSAN	Leaf	Passenger	75	0	Odometer	1	Used	0	0	Registrati	October 06	2018	King	SEATTLE	WA	
18	Battery	E11N4BZ0CP2C	348645972	2016	NISSAN	Leaf	Passenger	84	0	Odometer	1	Used	0	0	Registrati	March 07	2022	King	SEATTLE	WA	
19	Battery	E11N4BZ0CP2C	348645972	2016	NISSAN	Leaf	Passenger	84	0	Odometer	1	Used	0	0	Registrati	March 09	2020	King	SEATTLE	WA	
20	Battery	E11N4BZ0CP2C	348645972	2016	NISSAN	Leaf	Passenger	84	28238	Actual Mil	Used	14699	March 26	0	Transfer	1	April 03	2019	King	SEATTLE	WA
21	Battery	E17SAYGAE5F	255086793	2023	TESLA	Model Y	Passenger	0	0	Odometer	1	New	0	0	Original	1	October 26	2023	King	BELLEVUE	WA
22	Battery	E17SAYGAE5F	255086793	2023	TESLA	Model Y	Passenger	0	15	Actual Mil	New	57490	September	0	Original	1	October 26	2023	King	BELLEVUE	WA
23	Battery	E11G1FZ6S0X	144901522	2019	CHEVROLET	Bolt EV	Passenger	238	0	Odometer	1	Used	0	0	Registrati	October 12	2021	Clark	BATTLE GRC	WA	
24	Battery	E11G1FZ6S0X	144901522	2019	CHEVROLET	Bolt EV	Passenger	238	2897	Actual Mil	New	42130	October 17	0	Original	1	October 17	2019	Clark	BATTLE GRC	WA
25	Battery	E11G1FZ6S0X	144901522	2019	CHEVROLET	Bolt EV	Passenger	238	0	Odometer	1	Used	0	0	Registrati	October 26	2020	Clark	BATTLE GRC	WA	
26	Battery	E11G1FZ6S0X	144901522	2019	CHEVROLET	Bolt EV	Passenger	238	0	Odometer	1	New	0	0	Original	1	October 17	2019	Clark	BATTLE GRC	WA
27	Battery	E11G1FZ6S0X	144901522	2019	CHEVROLET	Bolt EV	Passenger	238	0	Odometer	1	Used	0	0	Registrati	October 06	2023	Clark	BATTLE GRC	WA	
28	Battery	E11G1FZ6S0X	144901522	2019	CHEVROLET	Bolt EV	Passenger	238	0	Odometer	1	Used	0	0	Registrati	October 06	2022	Clark	BATTLE GRC	WA	
29	Battery	E11V2VMP87F	252241594	2023	VOLKSWAGEN	ID. 4	Passenger	0	32	Actual Mil	New	49101	August 24	0	Original	1	September	2023	Clark	VANCOUVER	WA
30	Battery	E11V2VMP87F	252241594	2023	VOLKSWAGEN	ID. 4	Passenger	0	0	Odometer	1	New	0	0	Original	1	September	2023	Clark	VANCOUVER	WA
31	Battery	E17SAXCAE51F	224190844	2023	TESLA	Model X	Passenger	0	0	Odometer	1	New	0	0	Original	1	November 0	2022	Clark	VANCOUVER	WA
32	Battery	E17SAXCAE51F	224190844	2023	TESLA	Model X	Passenger	0	0	Odometer	1	New	0	0	Original	1	November 0	2022	Clark	VANCOUVER	WA

① ETL操作: 数据转换

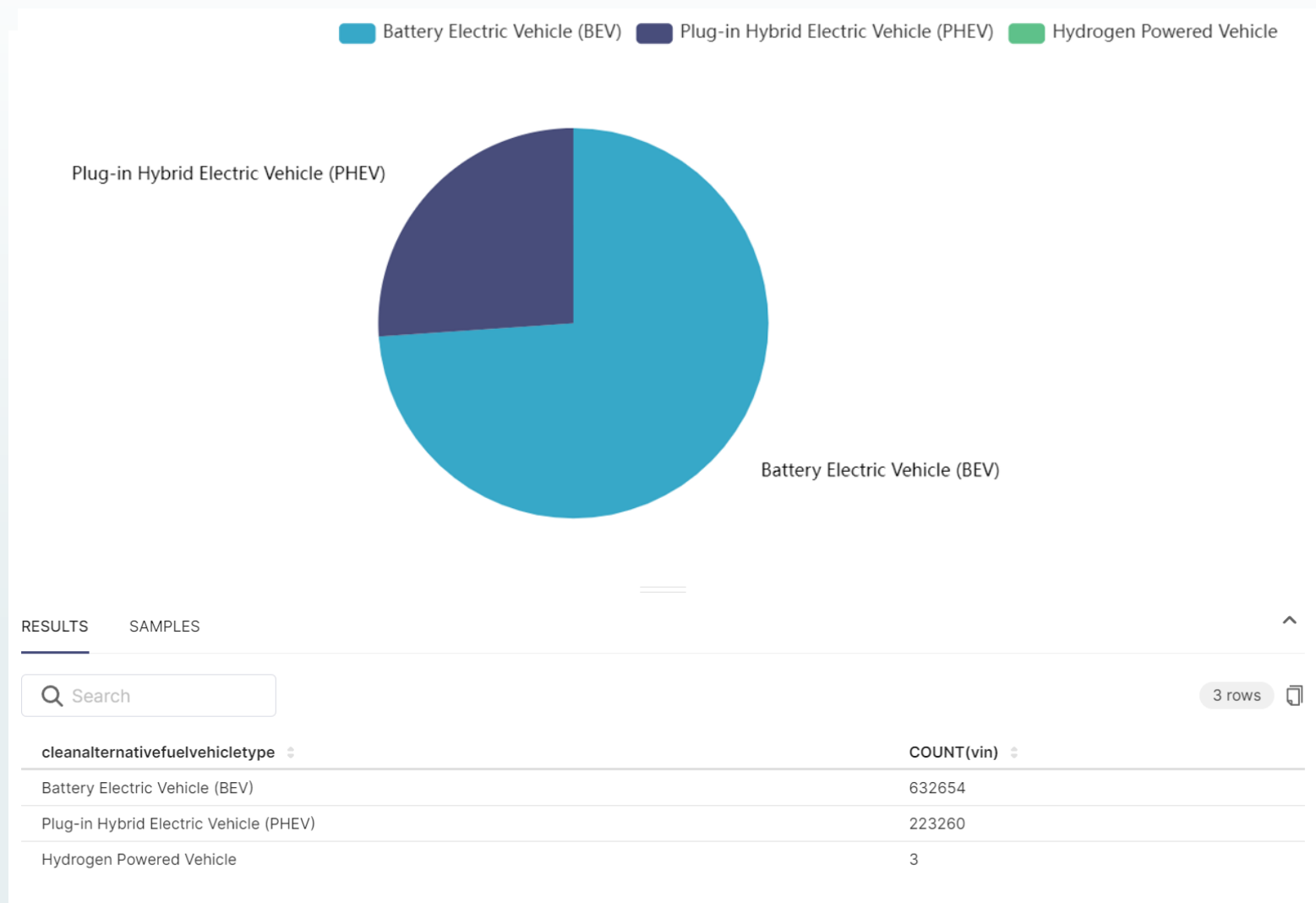
视频主要介绍的是数据清洗

视频主要介绍的是数据清洗



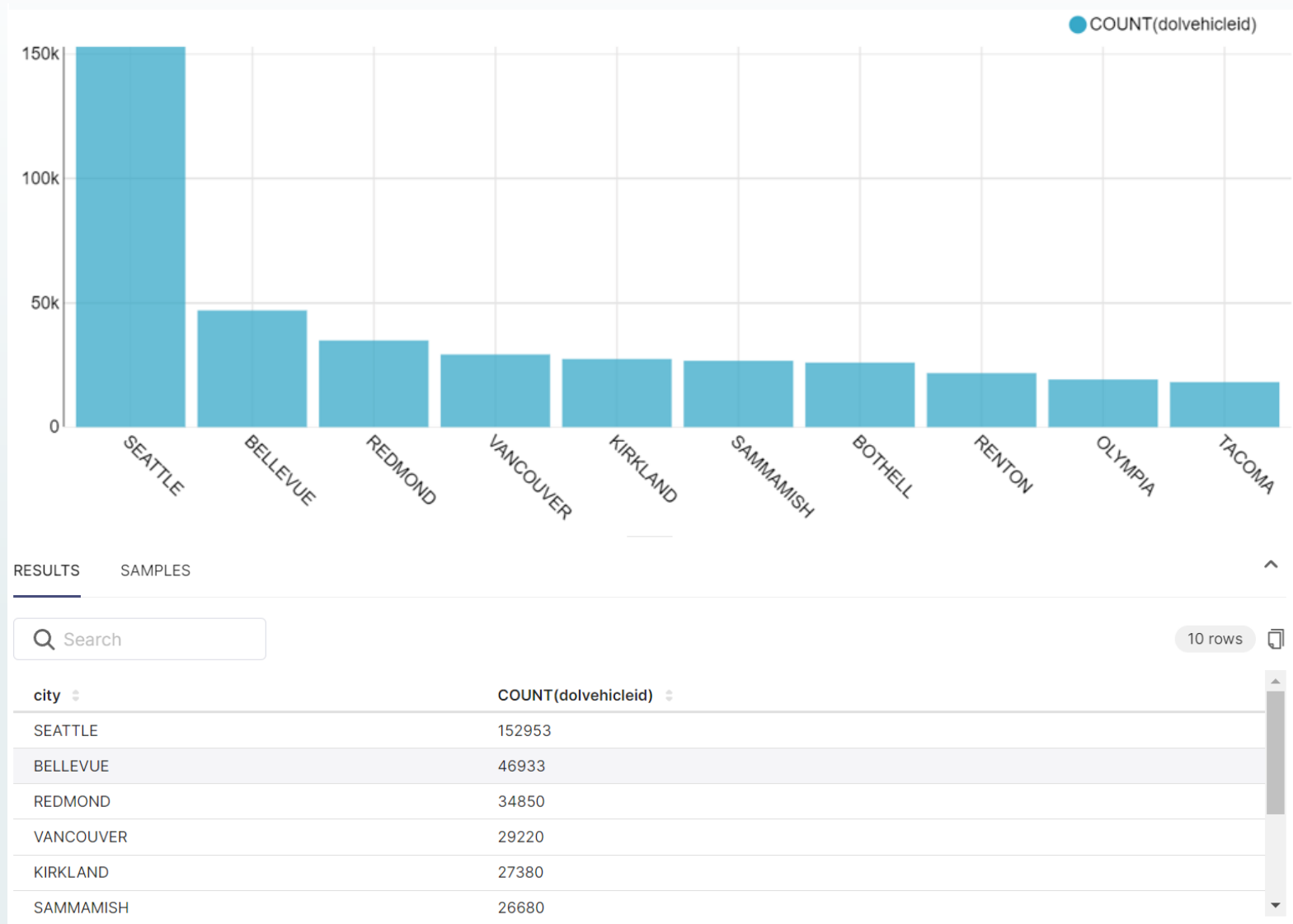
可 视 化 分 析

1. 新能源各类车型占比



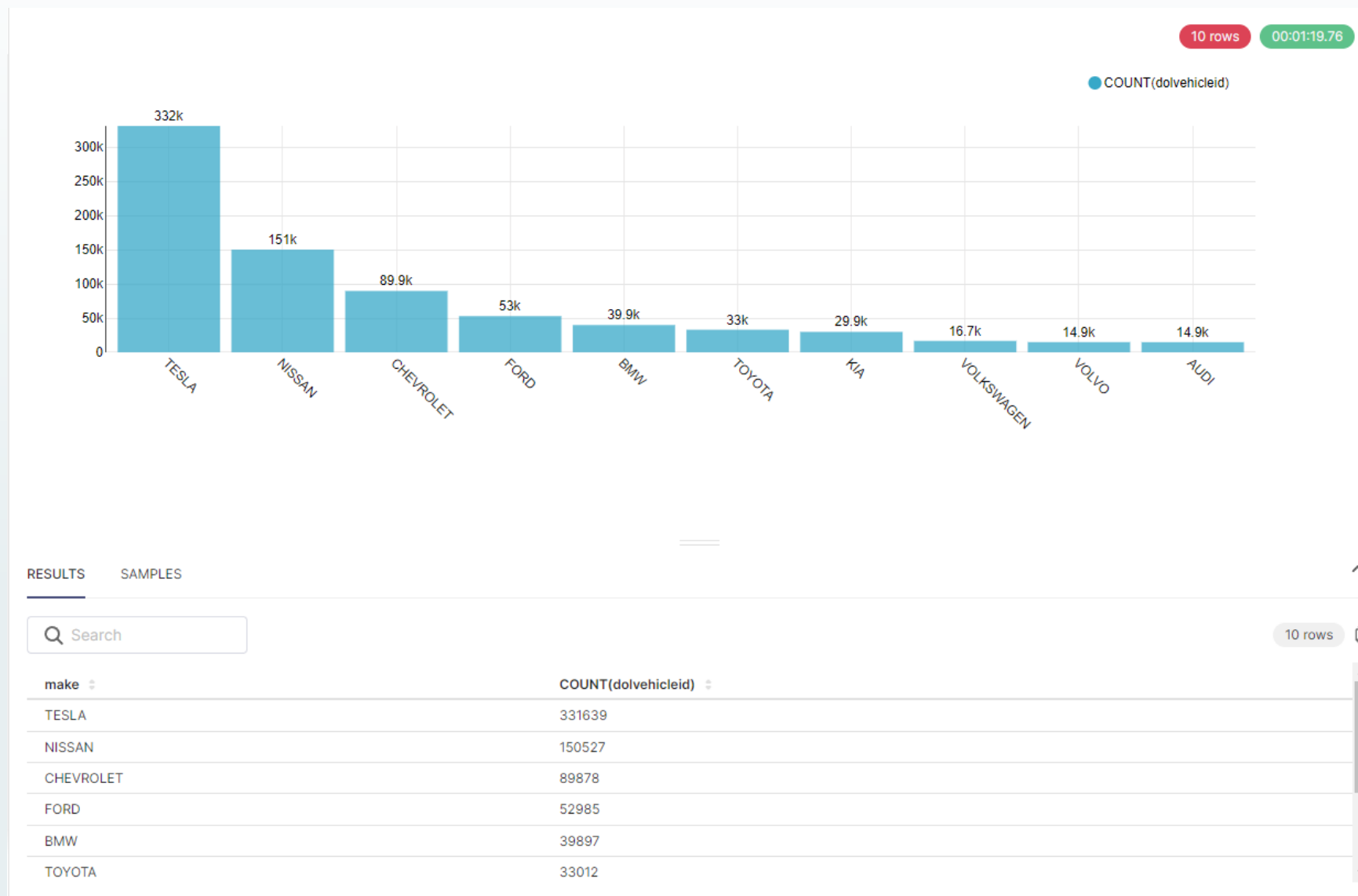
- 市场主导趋势：纯电动汽车在新能源车型市场中占据主导地位，超过插电式混合动力汽车和氢能源车
- 潜在增长机会：插电式混合汽车仍然有市场份额，显示出消费者对同时拥有电动和燃油驱动系统的混合动力车型的需求。
- 氢能源车型的挑战：氢能源汽车在市场中份额很小，可能由于消费者对其认知度和接受度较低，以及基础设施建设和成本问题，甚至会收到国家战略的影响。

2. 销量前十城市



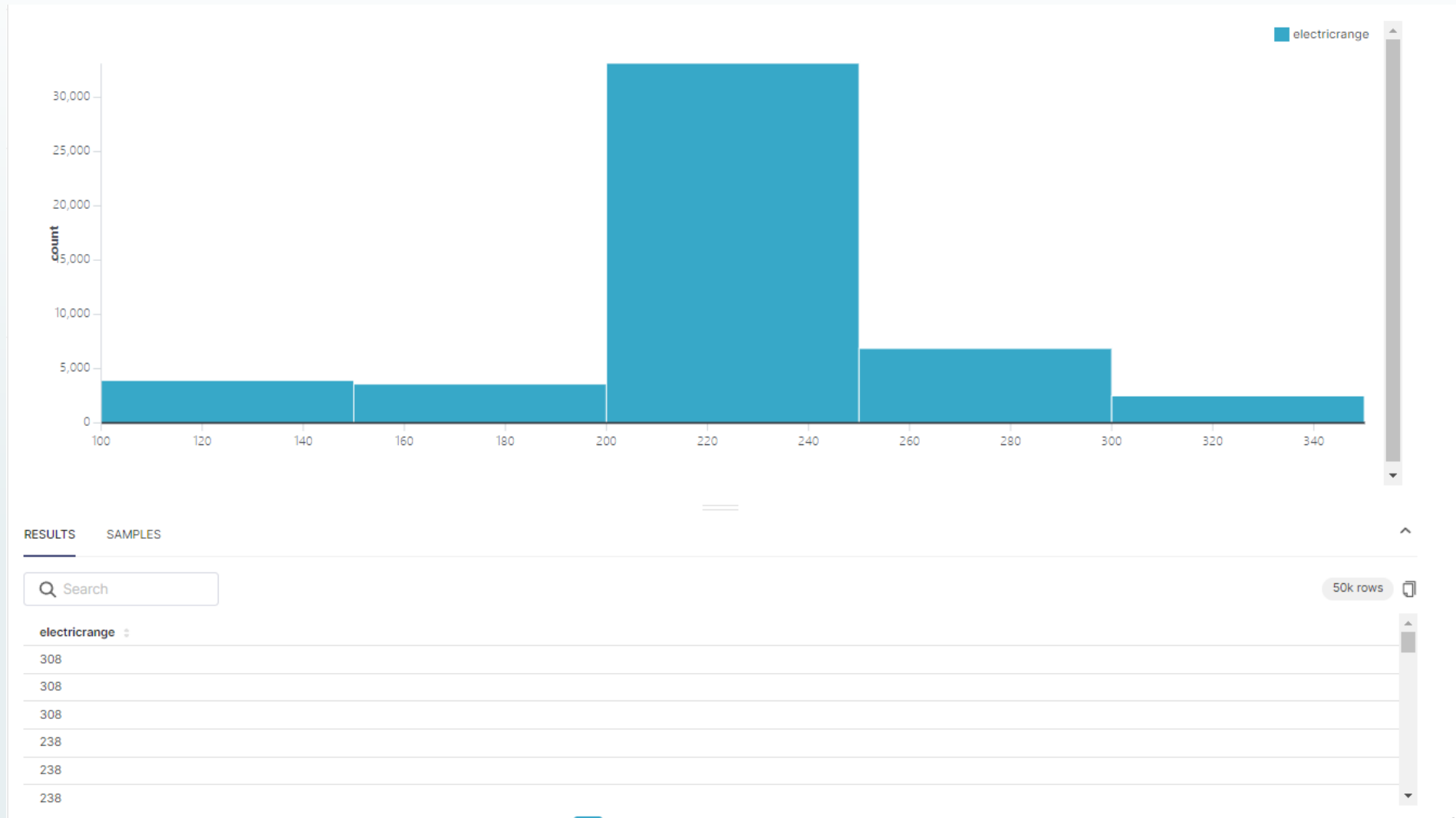
- 市场热点区域：SEATTLE（西雅图）是新能源汽车购买量最高的城市，具有巨大的市场潜力。
- 城市特定需求：不同城市的购买量与其特定需求和特点相关。例如，除SEATTLE外，其他华盛顿州城市的购买量相差不大，这可能与该地区的经济状况、居民收入水平以及对环境可持续性的关注有关。
- 市场潜力分析：购买量前十的城市数据可帮助制定适应当地市场的销售策略和营销活动。
- 充电基础设施规划：购买量较高的城市需要更多充电基础设施来满足新能源汽车的需求。

3. 新能源销售前十厂商



- 市场领导者：TESLA是新能源汽车销量最高的厂商。
- 品牌竞争：传统汽车厂商与TESLA等新兴厂商竞争激烈。
- 品牌认知度：TESLA的品牌认知度更高，其他厂商可以借鉴其经验。
- 市场多样性：消费者在新能源汽车市场有多种选择。

4. 续航里程分布直方图



- 用户需求：200到250的中等续航里程电动汽车有较高需求。
- 市场机会：满足这一续航里程范围的产品可能存在市场机会。
- 我们可以看出在100-200续航里程范围，通过增加成本、优化技术来提高里程所获得的需求增幅大、性价比高，并且200-250续航历程范围已经可以满足大部分用户的需求，250以上里程范围再提高续航里程，性价比较低

5. 不同清洁能源车型与车辆状态的相关关系



- 纯电动车型（BEV）在新车和二手车市场上销量高，是一个有前景的商业领域。
- 氢燃料车型在市场上需求低，商业机会有限，需要更多市场研究和推广工作。
- 插电式混合动力车型在二手车市场上有销量，但新车市场需求相对较低。
- 二手清洁能源车型销量高，商家可加大对二手清洁能源车型的销售和推广力度。

6. 销售前五的品牌销售趋势图



- 销售趋势：TESLA的销售量显著增长，其他品牌变化较小。
- 品牌竞争力：TESLA展现出强大竞争力，其他品牌需提高竞争力。
- 市场趋势：TESLA的增长反映了市场快速增长和消费者认可。

通过集成Hadoop、Hive和Presto等大数据技术，我们创建了一个强大的数据仓库，用于存储和分析华盛顿州电动汽车车辆登记与所有权变更活动数据集。在我们的总体设计中，Hadoop作为分布式存储和计算框架，为大规模数据提供高可靠性和可扩展性。Hive作为元数据存储和查询引擎，通过HQL转换为MapReduce任务，支持复杂的SQL-like查询。Presto则作为交互式查询引擎，连接Hive等数据源，实现低延迟的数据分析。总体而言，我们完成了一个完整的数仓的workflow，最后取得了很多可视化的结论分析并挖掘出了很多有用的信息，小组的每个人都收获良多。



感 谢 倾 听

T H A N K F O R L I S T E N I N G

—— **小组成员：黄勖、石宇昊、曾鸿勇、邹慧** ——